PROGNOSE UND PRÄVENTION VON SCHWIERIGKEITEN BEIM SCHRIFTSPRACHERWERB

Theoretische Aufarbeitung der vorhandenen Literatur und längsschnittliche Entwicklung und Evaluation des curriculumbasierten Messverfahrens „Phonolino“

Dissertation

zur Erlangung des akademischen Grades
eines Doktors der Philosophie (Dr. phil.)

vorgelegt von Lea Schulz,
geb. am 19.03.1984 in Flensburg

am
Institut für Sonderpädagogik
der Europa-Universität Flensburg

Erstgutachter: Prof. Dr. Armin Castello, Europa-Universität Flensburg
Zweitgutachter: Prof. Dr. Horst Biedermann, Pädagogische Hochschule St. Gallen

Datum der Einreichung: 26.06.2017
Datum der Disputation: 11.12.2017

Für meine Familie

Danksagung

Mein Dank gebührt all jenen Menschen, die mich während der Promotion unterstützt und gefördert haben.

An erster Stelle möchte ich meinen beiden Betreuern Prof. Dr. Armin Castello und Prof. Dr. Horst Biedermann für ihre Unterstützung, ihren Rat und ihre Begleitung, für fachliche Expertise und vor allem aufmunternde Worte danken. Sie standen mir mit hilfreichen Ideen sowie Kritik als meine wichtigsten Ratgeber zur Seite. Sie ermutigten mich „hartnäckig“ zu bleiben, den Winden, die wehten Stand zu halten. Ohne eure helfenden Hände, läge diese Arbeit nicht vor. Ein Dank gebührt ebenfalls Prof. Dr. Volker Müller-Benedikt. Er eröffnete mir den Zugang zu Methodik und Statistik und führte mich in die Tiefen der klassischen Testtheorie.

Ein liebevoller Dank gebührt insbesondere meinen drei Kindern Niklas, Miko und Benne, die mich über die Jahre hinweg in Höhen und Tiefen erlebt und geliebt und immer wieder auch in schwierigen Zeiten zum Lachen gebracht haben. Besonders meinem Ehemann Björn möchte ich danken, der mich in guten wie in schlechten Tagen begleitet und mich aufgemuntert hat sowie mir den Rücken freihalten konnte. Er ist mein Fels in der Brandung.

Ein besonderer Dank gilt meinem Vater Wolfgang Schulz, der mir bedingungslos in allen Lebenslagen mit Rat und Tat zur Seite stand, mein wissenschaftliches wie privates Tun mit mir reflektierte, mich unterstützte und mir dazu verhalf meine Tätigkeiten und auch den Weg meiner wissenschaftlichen Arbeit zu fokussieren - und mich zugleich bei jeglichen emotionalen Höhen und Tiefen mit offenen Armen empfangen hat. Hierfür möchte ich einen seiner liebsten Autoren zitieren:

> *„Krise kann ein produktiver Zustand sein. Man muß ihr nur den Beigeschmack der Katastrophe nehmen.“ (Max Frisch)*

Du hast mich gelehrt, in Krisenzeiten das Positive zu sehen und die Katastrophenprophezeiungen anderen zu überlassen. Danke, Papi.

Ein Dank auch an meine Mutter Evelyne Hartmann sowie an meine zahlreichen Geschwister für Anteilnahme, Unterstützung, Interesse sowie die Liebe, die sie meinen Kindern entgegengebracht haben, während diese den Großteil ihrer Ferien im fernen Flensburg verbrachten. Damit hatte ich die Möglichkeit, die Gratwanderung zwischen einer Vollzeitberufstätigkeit, wissenschaftlicher Qualifizierung und der Koordination meiner eigenen kleinen Familie zu bewältigen.

Ich möchte auch meinen Freunden für Interesse, Geduld und emotionaler Unterstützung danken, insbesondere Sarah Dwinger und Svenja Affeldt, die für mich, da sie sich in der gleichen Lage befinden, große Empathie und tiefes Verständnis aufbrachten. Sarah, ohne dein Angebot zum gemeinsamen Öffnen der schon verloren geglaubten Arbeit, hätte ich den Ordner wohl in der Ecke verstauben lassen. Für Anteilnahme und Ermutigung, sowie für ein vorübergehendes „Schreib-Asyl“ danke ich ebenfalls der Familie Czimmek, insbesondere Angelika Czimmek für tiefgehende und aufbauende Gespräche.

Für die Durchführung der Studie an den Kindertagesstätten und Schulen in Schleswig-Holstein danke ich den Erziehern, Eltern und insbesondere den 258 Kindern, die mir mit ihrer Aufrichtigkeit und ihrem kreativen Umgang mit dem Messverfahren Phonolino unter Beweis gestellt haben, dass der Mensch und insbesondere das Kind bei jedwedem Messverfahren im Vordergrund steht.

Für die anfängliche finanzielle Unterstützung danke ich der Graduiertenförderung des Landes Schleswig-Holstein.

Abstract

Replikationsstudien bekannter Screening-Verfahren zur Prognose schriftsprachlicher Fähigkeiten zeigen auf, dass eine sichere Klassifikation von Risiko- und Nicht-Risikokindern nicht zu gewährleisten ist (vgl. bspw. Marx & Weber 2006). Mögliche Ursachen der Klassifikationsproblematik könnten eine nicht angemessene Methodik oder unzureichende Prädiktoren der Screening-Verfahren sein.

In der vorliegenden Längsschnittuntersuchung wird eine Lernverlaufsdiagnostik als alternatives Messverfahren evaluiert: die curriculumbasierte Messung (CBM) „Phonolino“ zur Prognose und Prävention von Schwierigkeiten beim Schriftspracherwerb im Kindergartenalter. Ziel der Untersuchung mit N = 214 Kindern ist die Überprüfung der Objektivität, Reliabilität, Validität und Änderungssensibilität des Messinstruments. Es wird untersucht, ob sich die aus dem US-amerikanischen Bereich stammende Messmethode in Kombination mit den wissenschaftlichen Erkenntnissen zur phonologischen Informationsverarbeitung im deutschen Sprachraum als Indikator für die Lese- und Schreibleistung beweist und bessere Vorhersagen für Risiko- und Nicht-Risikokinder im ersten Schuljahr treffen kann als bisher bekannte Früherkennungs-verfahren.

Es konnte für je zwei nebeneinander liegende Messzeitpunkte der CBM Phonolino eine hohe Paralleltestreliabilität nachgewiesen werden ($.64 \leq r \leq .82$). Für die prognostische Validität ergaben sich für das Lesen ($.28 \leq r \leq .43$) und die Rechtschreibung ($.34 \leq r \leq .56$) mittlere bis hohe Korrelationskoeffizienten. In Kombination mit der zusätzlichen Variablen der Intelligenz und Phonolino konnten sehr gute bis gute prognostische Vorhersagen getroffen werden. Die Änderungssensibilität von Phonolino konnte ebenfalls nachgewiesen werden.

Replication studies of known screening instruments to predict written language skills show that a reliable classification of risk and non-risk children cannot be guaranteed (see, for example, Marx & Weber 2006). An inadequate screening methodology or insufficient predictors could be a possible cause of this problem.

In this panel study, a progress-monitoring procedure is evaluated as an alternative measurement method: the curriculum-based measurement (CBM) “Phonolino” for the prediction and prevention of difficulties in reading acquisition at preschool age. This study of N = 214 children examines the reliability, validity and sensitivity of growth of the measurement tool. It will be examined whether the US measurement method combined with scientific conclusions for phonological processing in German-speaking countries proves to be a indicator for the

performance in literacy, making better predictions of risk and non-risk children in first grade than other already available popular early diagnosis methods.

The results show that two adjacent measurement points of CBM Phonolino have a high level of test-retest stability ($.64 \leq r \leq .82$). The prognostic validity resulted in reading ($.28 \leq r \leq .43$) and spelling ($.34 \leq r \leq .56$) as a moderate to high coefficient of correlations. In combination with the additional variable of intelligence and Phonolino, very good or good prognosis for first grade achievement were achieved. Sensitivity could although be proved.

Inhaltsverzeichnis

ABSTRACT **6**

ABBILDUNGSVERZEICHNIS **13**

TABELLENVERZEICHNIS **16**

FORMELN **20**

ABKÜRZUNGSVERZEICHNIS **21**

EINLEITUNG **22**

I. PROGNOSE SCHRIFTSPRACHLICHER FÄHIGKEITEN IM VORSCHULALTER **34**

1 Notwendigkeit früher Diagnostik und Prävention **36**
1.1 Veränderungen im Bildungssystem 38
1.2 Potenzial der Prävention 41
1.3 Mögliche Folgen ohne frühzeitige Prognose 44
1.3.1 Prävalenz und Komorbidität der Lese-Rechtschreibschwäche 44
1.3.2 Verlauf und Stabilität der Lese-Rechtschreibschwäche 50

2 Testinstrumentarien zur Prognose schriftsprachlicher Fähigkeiten im Vorschulalter **53**
2.1 EXKURS: Gütekriterien für Screening-Verfahren 54
2.1.1 Objektivität 55
2.1.2 Reliabilität 56
2.1.3 Validität 57
2.1.3.1 Inhaltliche Validität, Konstruktvalidität und kriterienorientierte Validität 58
2.1.3.2 Prognostische Validität 59
2.1.4 Weitere Gütekriterien 69
2.2 Ausgewählte Testverfahren zur Vorhersage von Lese- und Rechtschreibschwierigkeiten im Vorschulalter 70
2.2.1 Bielefelder Screening (BISC) 71
2.2.2 Die Differenzierungsprobe 79
2.2.3 Phonologische Bewusstheit bei Kindergartenkindern und Schulanfängern (PB-LRS) 80
2.2.4 Heidelberger Auditives Screening in der Einschulungsuntersuchung (HASE) 81
2.2.5 Der Rundgang durch Hörhausen 84
2.3 Vergleich der Verfahren 87

3 Probleme der Klassifikation von Risikokindern durch bisherige Testinstrumentarien **89**

II. URSACHEN DER KLASSIFIKATIONSPROBLEMATIK **96**

1 Inhalte der Screening-Verfahren **96**
1.1 Zur Übertragbarkeit englischer Studien auf den deutschen Sprachraum 103
1.2 Phonologische Informationsverarbeitung 106
1.2.1 Phonologische Bewusstheit 108
1.2.1.1 Korrelationsstudien zur phonologischen Bewusstheit 112

1.2.1.2 Trainingsstudien zur phonologischen Bewusstheit ... 118
1.2.1.3 Gruppenvergleich ... 127
1.2.1.4 Entwicklung der phonologischen Bewusstheit: Ursache oder Wirkung? ... 129
1.2.2 Phonologisches Rekodieren beim Zugriff auf das semantische Lexikon ... 137
1.2.3 Phonetisches Rekodieren im Arbeitsgedächtnis ... 142
1.2.4 Wechselseitige Abhängigkeiten der Komponenten der phonologischen Informationsverarbeitung ... 150
1.3 Buchstabenkenntnis ... 154
1.4 Sprache ... 156

2 Endogene und exogene Einflüsse auf die Schülerleistungen ... 165
2.1 Schülerspezifische Variablen ... 170
2.1.1 Genetische Prädisposition ... 170
2.1.2 Kognitive Voraussetzungen ... 173
2.1.2.1 Auditive Informationsverarbeitung ... 177
2.1.2.2 Visuelle Informationsverarbeitung ... 179
2.1.2.3 Intelligenz ... 182
2.1.3 Weitere Persönlichkeitsvariablen ... 187
2.2 Variablen der familiären Umwelt ... 188
2.3 Schulische und unterrichtliche Variablen ... 190

3 Methodik der Screening-Verfahren ... 193
3.1 Pädagogische Diagnostik ... 194
3.2 Klassifikation herkömmlicher Testverfahren ... 196
3.3 Kritische Betrachtung der klassischen Testtheorie ... 198

III. CURRICULUMBASIERTE MESSUNG (CBM) - EINE ALTERNATIVE DIAGNOSTIKMETHODE ZUR UNTERSTÜTZUNG DER PRÄVENTION ... 202

1 Geschichte und Entwicklung ... 204

2 Beschreibung und Aufbau ... 208
2.1 Definition ... 208
2.2 Durchführung und Dokumentation ... 211
2.3 Einsatzbereich ... 212
2.4 Aufgabenstichproben ... 215

3 Aktuelle Studien ... 217
3.1 Darstellung der Untersuchungsergebnisse aus dem US-amerikanischen Bereich ... 218
3.2 Darstellung der Untersuchungsergebnisse aus dem deutschsprachigen Bereich ... 220
3.3 Curriculumbasierte Messung im Vorschulalter ... 225

4 Interventionsparadigma der CBM im Vergleich zum traditionellen Ansatz ... 230
4.1 Der ATI-Ansatz („Aptitude-Treatment-Interaction“) ... 231
4.2 Der RTI-Ansatz („Response-To-Intervention“) ... 232

5 Kritische Betrachtung der CBM unter dem RTI-Paradigma ... 243

6 Erkenntnisse zur Konzeption und Evaluation von CBM ... 249

IV. CURRICULUMBASIERTE MESSUNG „PHONOLINO“ FÜR DIE ÜBERPRÜFUNG DER PHONOLOGISCHEN INFORMATIONSVERARBEITUNG IM LETZTEN KINDERGARTENJAHR 255

1 Allgemeine Hinweise zur Konzeption und Evaluation von CBM für die Messung phonologischer Informationsverarbeitung 255

2 Erhebungsmaterial „Phonolino“ 256
2.1 Zielsetzung von Phonolino 256
2.2 Vortest zu Phonolino 258
2.3 Aufbau von Phonolino 260
2.4 Motivationale Komponenten 264
2.5 Konzeption der Items von Phonolino 265
2.5.1 Begründung zur Auswahl der Items 268
2.5.2 Zur Komplexität der Aufgaben 270
2.5.3 Zweidimensionales Konstrukt zur Erstellung der Aufgaben 272
2.5.4 Erstellung der Parallelmessungen 276

V. FORSCHUNGSFRAGEN DER EMPIRISCHEN UNTERSUCHUNG 282

1 Ableitung der Forschungsfragen zur Überprüfung des Messverfahrens nach den Gütekriterien der klassischen Testtheorie 286

2 Ableitung der Forschungsfragen zur Änderungssensibilität 289

3 Ableitung der Forschungsfragen zur Optimierung des Prognosemodells 291

VI. METHODE UND UNTERSUCHUNGSAUFBAU 293

1 Untersuchungsdesign und zeitlich-organisatorischer Ablauf 293
1.1 Zeitplan für die Durchführung von Phonolino 294
1.2 Zeitplan für weitere Testungen 295

2 Beschreibung der Stichprobe 296
2.1 Beschreibung der Grundgesamtheit 296
2.2 Beschreibung der Bildungsstätten 299
2.3 Verluste von Probanden 300
2.4 Deskriptive Analyse der Untersuchungsgruppe 304

3 Durchführung der Untersuchung 306
3.1 Durchführung und Auswertung von Phonolino 306
3.2 Durchführung und Auswertung der standardisierten Tests und Fragebögen 307

4 Untersuchungsinstrumentarien 308
4.1 Eingesetzte Testverfahren im Vorschulalter 308
4.1.1 Bielefelder Screening (BISC) 309
4.1.2 CFT 1 – Grundintelligenztest Skala 1 309
4.2 Eingesetzte Testverfahren im 1. Schuljahr 311
4.2.1 Deutscher Rechtschreibtest für 1. Klassen (DRT 1) 311
4.2.2 Knuspels Leseaufgaben (Knuspel-L) 311

5 Statistische Verfahren 313
5.1 EXKURS: Grundbegriffe verwendeter statistischer Methoden 316
5.1.1 Korrelationen 316
5.1.2 Signifikanz 318

5.2 Analyse der klassischen Testgütekriterien 319
5.2.1 Itemanalyse 319
5.2.1.1 Analyse der Rohwertverteilung 319
5.2.1.2 Itemschwierigkeit 320
5.2.1.3 Trennschärfe 321
5.2.1.4 Dimensionalität 322
5.2.2 Reliabilitätsanalyse 324
5.2.2.1 Paralleltest-Reliabilität 324
5.2.2.2 Split-Half-Reliabilität 325
5.2.2.3 Interne Konsistenz 327
5.2.3 Validitätsanalyse 328
5.2.3.1 Inhaltsvalidität 329
5.2.3.2 Konstruktvalidität 329
5.2.3.3 Das Vierfelderschema 330
5.2.3.4 Kennwerte der klassifikatorischen Validität 335
5.3 Methoden zur Bestimmung der Änderungssensibilität 337
5.4 Methoden zur Erweiterung des Prognosemodells 342
5.5 Datenauswertung 348

VII. DARSTELLUNG DER ERGEBNISSE 349

1 Klassische Testgütekriterien 349
1.1 Objektivität 349
1.2 Itemanalyse 350
1.2.1 Analyse der Rohwertverteilung 350
1.2.2 Itemschwierigkeit 351
1.2.3 Trennschärfe 352
1.2.4 Dimensionalität 353
1.3 Reliabilität 355
1.3.1 Paralleltestreliabilität 355
1.3.2 Split-Half-Reliabilität 359
1.3.3 Interne Konsistenz 361
1.4 Validität 361
1.4.1 Übereinstimmungsvalidität 362
1.4.2 Prognostische Validität 364

2 Darstellung der Ergebnisse zur Änderungssensibilität 372

3 Darstellung der Ergebnisse zur Erweiterung des Prognosemodells 382

VIII. DISKUSSION DER ERGEBNISSE 393

1 Interpretation der Ergebnisse zu den klassischen Testgütekriterien 393
1.1 Objektivität 393
1.2 Itemanalyse 393
1.3 Reliabilität 394
1.4 Validität 397

2 Interpretation der Ergebnisse zur Änderungssensibilität 403

3 Interpretation der Ergebnisse zur Erweiterung des Prognosemodells 406

4 Zusammenfassende Beantwortung der Fragestellungen 408

IX. FAZIT 414

1 Kritische Diskussion der Untersuchung 414

2 Erkenntnisse für Diagnostikverfahren zur Prädiktion von Lese- und Rechtschreibfähigkeiten 425

3 Ausblick 435

X. LITERATUR 447

XI. ANHANG 492

EIDESSTATTLICHE ERKLÄRUNG 493

Abbildungsverzeichnis

ABB. 1: FILTERFUNKTION DES SCREENINGS (NACH TRÖSTER 2009, S. 69) 34
ABB. 2: FRÜHERKENNUNGSSCHEMA (NACH TRÖSTER 2009, S. 62)....42
ABB. 3: GÜTEKRITERIEN: OBJEKTIVITÄT, RELIABILITÄT UND VALIDITÄT 55
ABB. 4: ÜBERPRÜFUNGSMÖGLICHKEITEN DER RELIABILITÄT (NACH LUKESCH 1998, S. 45) 57
ABB. 5: RATZ-INDIZES UND SIGNIFIKANZNIVEAUS FÜR DIE VORHERSAGE VON SPEZIFISCHEN UND UNSPEZIFISCHEN KRITERIUMSLEISTUNGEN AM ENDE DES 2. SCHULJAHRES MIT SPEZIFISCHEN, UNSPEZIFISCHEN UND KONFUNDIERTEN PRÄDIKTOREN (VERÄNDERT NACH MARX, JANSEN & SKOWRONEK 2000, S. 25)....75
ABB. 6: ÜBERSICHT ZUM NÜRNBERGER DIAGNOSE- UND FÖRDERPROGRAMM (VERÄNDERT NACH EINSIEDLER 2003, S. 54)....84
ABB. 7: GEMEINSAME KOGNITIVE PRÄDIKTOREN VON SCHRIFTSPRACHKOMPETENZ UND RECHENLEISTUNG (VERÄNDERT NACH PETERMANN 2003, S. 195) 100
ABB. 8: KONNEKTIONISTISCHES MODELL DER PHONOLOGISCHEN BEWUSSTHEIT VON STACKHOUSE UND WELLS 1997 (NACH SCHNITZLER 2008, S. 15)....111
ABB. 9: ANZAHL RICHTIG GELESENER WÖRTER IN DER ERSTEN UND ZWEITEN KLASSE IN EXPERIMENTAL- UND KONTROLLGRUPPE (ÜBERSETZT AUS DEM ENGLISCHEN, NACH LUNDBERG, FROST & PETERSON, S. 281) 120
ABB. 10: ANZAHL RICHTIG BUCHSTABIERTER WÖRTER IN DER ERSTEN UND ZWEITEN KLASSE IN EXPERIMENTAL- UND KONTROLLGRUPPE (ÜBERSETZT AUS DEM ENGLISCHEN, NACH LUNDBERG, FROST & PETERSON, S. 281) 121
ABB. 11: ANTEIL VON KINDERN MIT LESE-RECHTSCHREIBSCHWÄCHE AM ENDE DER ZWEITEN KLASSE IN DREI VERSCHIEDENEN GRUPPEN, (ANGELEHNT AN MANNHAUPT 2003, S. 48) 123
ABB. 12: ZUSAMMENHÄNGE ZWISCHEN DER PHONOLOGISCHEN BEWUSSTHEIT UND SCHRIFTSPRACHLICHEN FÄHIGKEITEN (NACH SCHNITZLER 2008, S. 75) 136
ABB. 13: ZWEI-WEGE-MODELL (NACH SCHEERER-NEUMANN 1989B, S. 18)....138
ABB. 14: STRUKTUR DES ARBEITSGEDÄCHTNISSES NACH BADDELEY (NACH WELLENREUTHER 2005, S. 81)....144
ABB. 15: DAS MULTIKOMPONENTENMODELL VON BADDELEY 2003 (ÜBERSETZT AUS DEM ENGLISCHEN, VERÄNDERT NACH HIPPMANN 2008, S. 45) 145
ABB. 16: ERWEITERTES ZWEI-KOMPONENTEN-MODELL DES PHONOLOGISCHEN ARBEITSGEDÄCHTNISSES VON BADDELEY (NACH HASSELHORN, GRUBE & MÄHLER 2000, S. 174) 146
ABB. 17: STRUKTURGLEICHUNGSMODELL, KRITERIUM: LESEVERSTÄNDNIS (NACH SCHNEIDER UND NÄSLUND 1992, S. 268) 152
ABB. 18: STRUKTURGLEICHUNGSMODELL, KRITERIUM: RECHTSCHREIBUNG (NACH SCHNEIDER UND NÄSLUND 1992, S. 270) 152
ABB. 19: THEORETISCHES MODELL ZUM ZUSAMMENHANG VON SPRACHENTWICKLUNG UND SCHRIFTSPRACHERWERB NACH GRIMM 1995 (VERÄNDERT NACH SCHNEIDER, MARX & WEBER 2002, S. 192) 159
ABB. 20: STRUKTURGLEICHUNGSMODELL ZUR VORHERSAGE VON LESEGESCHWINDIGKEIT UND TEXTVERSTÄNDNIS (NACH ENNEMOSER, MARX, WEBER & SCHNEIDER 2012, S. 63) 161
ABB. 21: STRUKTURGLEICHUNGSMODELL ZUR VORHERSAGE VON DER RECHTSCHREIBKOMPETENZ (NACH ENNEMOSER, MARX, WEBER & SCHNEIDER 2012, S. 64)....161
ABB. 22: ERKLÄRUNGSANSÄTZE FÜR LESE-RECHTSCHREIBSCHWIERIGKEITEN (NACH WALTER 2005A, O. S.) 166
ABB. 23: HAUPTKOMPONENTEN DER LEGASTHENIE AUS DER SICHT DES PHONOLOGISCH ORIENTIERTEN ANSATZES, VERÄNDERT NACH LUNDBERG 1995, S. 91 (NACH WALTER 2007A, S. 481) 167
ABB. 24: DAS AUFTRETEN VON FAMILIENMITGLIEDERN MIT LEGASTHENIE ÜBER VIER GENERATIONEN NACH GRIMM, NOETHEN & SCHULTE-KÖRNE 1998 (NACH WARNKE, HEMMINGER, ROTH & SCHNECK 2002, S. 37) 171

ABB. 25: VEREINFACHTES URSACHENMODELL DER LESE-RECHTSCHREIBSTÖRUNG AUS NEUROLOGISCHER SICHTWEISE (NACH WARNKE, HEMMINGER & PLUME 2004, S. 16) 175
ABB. 26: URSACHENMODELL BEI EINER LESE- UND RECHTSCHREIBSCHWÄCHE (AUS DEM ENGLISCHEN ÜBERSETZT, NACH CASTLES, MCLEAN & MCARTHUR 2010, S. 429).....176
ABB. 27: DARSTELLUNG DER BEGRIFFLICHKEIT „CURRICULUM-BASED MEASUREMENT“ (CBM).....208
ABB. 28: SUMMATIVE IM VERGLEICH ZU FORMATIVEN MESSVERFAHREN.....209
ABB. 29: SCHEMATISCHE DARSTELLUNG DES ZUSAMMENHANGS VON CBA, CBM, SBM, GOM, DIBS UND MM IN ANLEHNUNG AN TUCKER (1985), DENO (1985), HOSP ET AL. (2007), FUCHS UND DENO (1994), SHINN (1995) SOWIE VANDERHEYDEN UND BURNS (2005), ZIT. NACH VOß & HARTKE 2014, S. 86 210
ABB. 30: CBM: GRAPHISCHE DARSTELLUNG DER LERNENTWICKLUNG (ÜBERSETZT AUS DEM ENGLISCHEN, VERÄNDERT NACH DENO 2003, S. 186) 211
ABB. 31: BEISPIEL FÜR EINE CBM FÜR DIE MESSUNG VON LESEFLÜSSIGKEIT (AUSSCHNITT AUS WRIGHT O. J., S. 6) 215
ABB. 32: BEISPIEL FÜR EINE CBM FÜR DIE MESSUNG MATHEMATISCHER LEISTUNGEN (ANGELEHNT AN WRIGHT O. J., S. 13).....216
ABB. 33: MESSUNGSSYSTEM DIBELS VOM KINDERGARTEN BIS ZUM SECHSTEN SCHULJAHR (VERÄNDERT NACH DIBELS 2002, S. 1).....229
ABB. 34: BEGRIFFSHERKUNFT DES TRADITIONELLEN ATI-ANSATZES.....231
ABB. 35: BEGRIFFSHERKUNFT DES NEUEREN RTI-ANSATZES 233
ABB. 36: PRÄVENTION NACH DEM RTI-MODELL.....234
ABB. 37: RTI-MODELL (ZIT. NACH MÜLLER & HARTMANN 2009, S. 27).....234
ABB. 38: 4-PHASEN-MODELL DARGESTELLT AN EINER EXEMPLARISCHEN DATENERHEBUNG DER CBM „RICHTIG GELESENE WÖRTER PRO MINUTE“ (AUS DEM ENGLISCHEN ÜBERSETZT, ANGELEHNT AN FUCHS & FUCHS 1998, S.206) 236
ABB. 39: BEISPIEL EINER DOKUMENTATION ZUR CBM „RICHTIG GELESENE WÖRTER PRO MINUTE“ ZUR BERECHNUNG EINER ZIELLINIE (AUS DEM ENGLISCHEN ÜBERSETZT, ANGELEHNT AN WRIGHT O. J., S. 4).....237
ABB. 40: BEISPIEL EINER CBM ZUM LESEN (RICHTIGE WÖRTER PRO MINUTE) FÜR EINE NEGATIVE UND EINE POSITIVE TRENDLINIE INNERHALB ZWEIER INTERVENTIONSMAßNAHMEN (NACH WRIGHT O. J., S. 10) 238
ABB. 41: BEISPIEL FÜR EINE ÄNDERUNG DER INTERVENTION NACH DER DREI-PUNKTE-REGEL VON WRIGHT (ANGELEHNT AN WALTER 2007C) 239
ABB. 42: DREI-PUNKTE-REGELUNG (ÜBERSETZT AUS DEM ENGLISCHEN NACH WRIGHT O. J., S. 6) 239
ABB. 43: KERNELEMENTE DES RTI IM KINDERGARTEN (AUS DEM ENGLISCHEN ÜBERSETZT, ANGELEHNT AN VANDERHEYDEN & SNYDER 2006, S. 526) 242
ABB. 44: EINORDNUNG VON PHONOLINO IN DAS RTI-MODELL 256
ABB. 45: BEISPIEL DER DURCHFÜHRUNGSANWEISUNG FÜR ITEM 3, DIE ERZIEHERIN SPRICHT DABEI MIT DER PHONOLINO-HANDPUPPE (VGL. PHONOLINO-HANDBUCH, ANHANG 1) 262
ABB. 46: SEITE 1 EINES CBM-PROTOKOLLS AUS DEM PROTOKOLLHEFT AM BEISPIEL DER MESSUNG PHONO3.....263
ABB. 47: SEITE 2 EINES CBM-PROTOKOLLS AUS DEM PROTOKOLLHEFT AM BEISPIEL DER MESSUNG PHONO3.....264
ABB. 48: ZUSAMMENHANG DER PHONOLOGISCHEN BEWUSSTHEIT IM ENGEREN UND IM WEITEREN SINNE (NACH SCHNITZLER 2008, S. 20) 271
ABB. 49: ZWEIDIMENSIONALES KONSTRUKT DER PHONOLOGISCHEN BEWUSSTHEIT (NACH SCHNITZLER 2008, S. 29). 273
ABB. 50: EINORDNUNG DER MESSUNGEN VON PHONOLINO AM ZWEIDIMENSIONALEN KONSTRUKT DER PHONOLOGISCHEN BEWUSSTHEIT (VERÄNDERT NACH SCHNITZLER 2008, S. 29) 276
ABB. 51: ANZAHL DER GESCHWISTER DER PROBANDEN 297
ABB. 52: ANZAHL DER KINDERBÜCHER IM HAUSHALT DER PROBANDEN 298
ABB. 53: PROTOKOLL FÜR DAS INTERVIEW ZUR PRÄDISPOSITION 299
ABB. 54: BALKENDIAGRAMME ZUM GESCHLECHT UND ZUM MIGRATIONSHINTERGRUND DER KINDER DER VORLIEGENDEN STICHPROBE VON N= 214 305
ABB. 55: MODELL DES „OPTIMALEN SCREENINGS“ (ANGELEHNT AN CATTS 2006, S. 11) 333
ABB. 56: MODELL ZWEIER TYPISCHER SCREENINGS MIT UNTERSCHIEDLICHEM CUT-OFF-WERT (ANGELEHNT AN CATTS 2006, S. 12) 334

ABB. 57: GRAFISCHE DARSTELLUNG EINER LOGISTISCHEN FUNKTION.....343
ABB. 58: VERÄNDERUNG DER MITTELWERTE DER LÄNGSSCHNITTLICH ERMITTELTEN ROHWERTE ALLER ITEMS VON PHONOLINO ÜBER DIE ACHT CURRICULUMBASIERTEN MESSUNGEN IM LETZTEN KINDERGARTENJAHR (2008/2009) HINWEG, N= 91 374
ABB. 59: VERÄNDERUNG DER MITTELWERTE DER LÄNGSSCHNITTLICH ERMITTELTEN ROHWERTE DER ITEMS „PSEUDOWÖRTER NACHSPRECHEN“ (PSW, N= 94), „SILBEN KLOPFEN“ (SK, N= 91) UND „LAUTISOLATION“ (LI, N=91) VON PHONOLINO ÜBER DIE ACHT CURRICULUMBASIERTEN MESSUNGEN IM LETZTEN KINDERGARTENJAHR (2008/2009) HINWEG.....374
ABB. 60: VERÄNDERUNG DER KORRELATIONEN ZWISCHEN PHONO 2 UND DEN NACHFOLGENDEN MESSZEITPUNKTEN.....380
ABB. 61: VERÄNDERUNG DER KORRELATIONEN ZWISCHEN PHONO 3 UND DEN NACHFOLGENDEN MESSZEITPUNKTEN.....380
ABB. 62: VERÄNDERUNG DER KORRELATIONEN ZWISCHEN PHONO 4 UND DEN NACHFOLGENDEN MESSZEITPUNKTEN.....381
ABB. 63: VERÄNDERUNG DER KORRELATIONEN ZWISCHEN PHONO 5 UND DEN NACHFOLGENDEN MESSZEITPUNKTEN.....381
ABB. 64: VERÄNDERUNG DER KORRELATIONEN ZWISCHEN PHONO 6 UND DEN NACHFOLGENDEN MESSZEITPUNKTEN.....382
ABB. 65: VERÄNDERUNG DER KORRELATIONEN ZWISCHEN PHONO 7 UND DEN NACHFOLGENDEN MESSZEITPUNKTEN.....382
ABB. 66: URSACHENFAKTOREN FÜR DIE KLASSIFIKATIONSPROBLEMATIK DER SCREENING-VERFAHREN ZUR PROGNOSE VON SCHRIFTSPRACHLICHEN LEISTUNGEN IN DER SCHULE 436

Tabellenverzeichnis

TAB. 1: VIERFELDERSCHEMA ZUR KLASSIFIKATION (MARX 1992A, S. 255) 60
TAB. 2: BEDEUTUNG UND BERECHNUNG DER GÜTEINDIZES (ANGELEHNT AN MARX, JANSEN & SKOWRONEK 2000, S. 46FF. UND TRÖSTER 2009, S. 105) 61
TAB. 3: INTERPRETATION DER TESTEFFIZENZ DES LIKELIHOOD RATIO (NACH BENDER 2001, S. 118) 64
TAB. 4: INTERPRETATION DES ODDS RATIO (NACH TRÖSTER 2009, S. 103) 64
TAB. 5: INTERPRETATION DER RATZ-WERTE NACH JANSEN (1992, IN JANSEN, MANNHAUPT, MARX & SKOWRONEK 2002, S. 47), ANGABEN IN PROZENT 69
TAB. 6: EINTEILUNG VON PRÄDIKTOREN UND KRITERIEN IN DER BIELEFELDER LÄNGSSCHNITTSTUDIE (NACH MARX, JANSEN & SKOWRONEK 2000, S. 17) 72
TAB. 7: AUFGABEN DES BIELEFELDER SCREENINGS VON JANSEN, MANNHAUPT, MARX & SKOWRONEK (2002, S. 16FF.) 73
TAB. 8: RATZ-INDIZES FÜR DIE BISC-MESSUNG 4 MONATE VOR DER EINSCHULUNG UND DEN KRITERIEN LESEN UND RECHTSCHREIBEN (BISC-WERTE AUS MARX & WEBER 2004, S. 199). 77
TAB. 9: RATZ-INDIZES ZUR VORHERSAGEGÜTE DES BISC, VIER MONATE VOR EINSCHULUNG (NACH MARX & WEBER 2006, S. 256) 78
TAB. 10: AUFGABEN DER DIFFERENZIERUNGSPROBE (NACH MARTSCHINKE, KIRSCHHOCK & FRANK 2005, S. 14) 80
TAB. 11: KENNWERTE DER KLASSIFIKATORISCHEN VALIDITÄT DES PB-LRS 81
TAB. 12: GÜTEINDIZES (IN PROZENT) FÜR DIE KLASSIFIKATORISCHE VALIDITÄT VON HASE (NACH TAB. 14 AUS TREUTLEIN, ROOS & SCHÖLER 2007, S. 18) 83
TAB. 13: RATZ-INDIZES VOM „RUNDGANG DURCH HÖRHAUSEN“ ZU ZWEI MESSZEITPUNKTEN (ZUSAMMENGESTELLT AUS MARTSCHINKE, KIRSCHHOCK & FRANK 2005, S. 38-39) 86
TAB. 14: VERFAHREN ZUR FRÜHERKENNUNG VON SCHULISCHEN LESE- UND RECHTSCHREIBSCHWIERIGKEITEN (ERWEITERT NACH MARX & LENHARD 2010, S. 290) 88
TAB. 15: VERGLEICH DER RATZ-WERTE ANDERER STUDIEN ZUR FRÜHERKENNUNG VON LESE- UND RECHTSCHREIBSCHWIERIGKEITEN AM ENDE DER ERSTEN KLASSE 89
TAB. 16: POTENZIELLE SCHÄDEN DURCH FALSCHE SCREENING-BEFUNDE IN DER FRÜHERKENNUNG (NACH TRÖSTER 2009, S. 86) 90
TAB. 17: VORHERSAGE VON SCHULLEISTUNGEN NACH WEINERT & SCHNEIDER 1987 (ANGELEHNT AN SCHNEIDER 1989, S. 161) 101
TAB. 18: KOMPONENTEN PHONOLOGISCHER INFORMATIONSVERARBEITUNG UND DEREN MÖGLICHE ERFASSUNGSMETHODEN (NACH WALTER 2007A, S. 479, SCHNEIDER 1989, S. 162) 108
TAB. 19: EINFACHE UND PARTIALKORRELATIONEN ZWISCHEN PRÄDIKTORMAßEN ZUR PHONOLOGISCHEN SPRACHENTWICKLUNG UND LESELEISTUNGEN GEGEN ENDE DER ERSTEN UND ZWEITEN KLASSENSTUFE NACH VELLUTINO & SCANLON 1987 (ANGELEHNT AN SCHNEIDER 1989, S. 164) 113
TAB. 20 KORRELATIONEN ZWISCHEN SCHRIFTSPRACHEFÄHIGKEITEN UND PHONOLOGISCHER BEWUSSTHEIT FÜR SCHÜLER DER ERSTEN KLASSE (ALPHABETISCHE STRATEGIE), N = 42 (NACH SCHNITZLER 2008, S. 70) 117
TAB. 21 KORRELATIONEN ZWISCHEN SCHRIFTSPRACHEFÄHIGKEITEN UND PHONOLOGISCHER BEWUSSTHEIT FÜR SCHÜLER DER DRITTEN/VIERTEN KLASSE (ORTHOGRAFISCHE STRATEGIE), N = 57 (NACH SCHNITZLER 2008, S. 70) 117
TAB. 22: KORRELATIONEN ZWISCHEN PHONOLOGISCHER BEWUSSTHEIT UND AUFMERKSAMKEIT UND GEDÄCHTNIS ZUR SCHREIBFÄHIGKEIT (AUS MARX 1992B, S. 225, TABELLE 20) 132
TAB. 23: DIE SUBGRUPPEN DER „DOUBLE-DEFICIT“-HYPOTHESE (NACH A. BERGLEZ 2002, S. 41) 140
TAB. 24: KOMPONENTEN DER PHONOLOGISCHEN SCHLEIFE, HYPOTHETISCH ZU DEN FUNKTIONALEN MERKMALEN UND EMPIRISCHEN INDIKATOREN ZUGEORDNET (NACH HASSELHORN, GRUBE & MÄHLER 2000, S. 176) 146
TAB. 25: MÖGLICHE PROBLEME DES LESEVERSTÄNDNISSES (ABGEWANDELT NACH HARRISON-PALMER 2012) 163

TAB. 26: IDEALTYPISCHE HANDLUNGS- UND ENTSCHEIDUNGSSTRATEGIEN (NACH WILD & KRAPP 2001, S. 523) 197
TAB. 27: ITEMS DER VORVERSION VON PHONOLINO, JULI 2008 258
TAB. 28: VERMINDERUNG DER ITEMS ALS RESULTAT DES VORTESTS 259
TAB. 29: ITEMS BEI PHONOLINO 261
TAB. 30: „DECISION-MAKING MATRIX“ (NACH DENO UND FUCHS 1987, S. 4) 266
TAB. 31: FRAGESTELLUNGEN DER „DECISION-MAKING MATRIX“ (NACH DENO UND FUCHS 1987, S. 4) 267
TAB. 32: PHONOLOGISCHE INFORMATIONSVERARBEITUNG BEI DEN ITEMS VON PHONOLINO (EINTEILUNG NACH SCHNITZLER 2008, S. 29F.) 272
TAB. 33: EINHEITEN DER DIMENSION DER OPERATION EINES ZWEIDIMENSIONALEN MODELLS (ANGELEHNT AN SCHNITZLER 2008, S. 21FF.) 274
TAB. 34: BEISPIEL FÜR DAS KV-MUSTER AUS DEN 278
TAB. 35: BEISPIEL AUS DEN PARALLELMESSUNGEN PHONO5 UND PHONO6 279
TAB. 36: BEISPIEL AUS DEN PARALLELMESSUNGEN CBM5 UND CBM6 280
TAB. 37: ZEITPLAN FÜR DIE PHONOLINO-MESSUNGEN IM KINDERGARTEN 294
TAB. 38: ZEITPLAN DER ERHEBUNG DER KRITERIEN IN DEN JAHREN 2008-2010 295
TAB. 39: ÜBERBLICK DER VERTEILUNG DER KINDER ÜBER DIE 15 KINDERTAGESSTÄTTEN IN SCHLESWIG-HOLSTEIN 296
TAB. 40: BESCHREIBUNG DER STICHPROBE ZU DEN TESTZEITPUNKTEN T0, T1, T2 IN KINDERGARTEN UND SCHULE 300
TAB. 41: BESCHREIBUNG DER STICHPROBE ZU DEN NEUN MESSZEITPUNKTEN DER CURRICULUMBASIERTEN MESSUNG IM KINDERGARTEN 301
TAB. 42: ERGEBNISSE DES KRUKAL-WALLIS-TESTS ZUR ÜBERPRÜFUNG AUF GRUPPENUNTERSCHIEDE ZWISCHEN DEN ELIMINIERTEN FÄLLEN UND DEN IN DEN BERECHNUNGEN VERBLEIBENDEN 214 FÄLLEN 303
TAB. 43: PRÄVALENZRATE DER LESE-RECHTSCHREIBLEISTUNGEN DER PROBANDEN AM ENDE DER ERSTEN KLASSE (GEMESSEN DURCH DEN DRT 1 UND KNUSPEL-L) 305
TAB. 44: STATISTISCHE VERFAHREN DER VORLIEGENDEN UNTERSUCHUNG 315
TAB. 45: INTERPRETATION DES KORRELATIONSKOEFFIZIENTEN (NACH COHEN 1988) 316
TAB. 46: INTERPRETATION DES SIGNIFIKANZNIVEAUS 318
TAB. 47: INTERPRETATION DER KENNWERTS DES KMO-KRITERIUMS 323
TAB. 48. ZUSAMMENHANG ZWISCHEN STICHPROBENGRÖßE, KOMMUNALITÄT UND DURCHFÜHRBARKEIT EINER FAKTORENANALYSE (NACH KLOPP 2013, S.4 ANGELEHNT AN BÜHNER 2006, S. 193) 324
TAB. 49: EINTEILUNG DER SCHÜLER IN GRUPPEN IM VIERFELDERSCHEMA (ABGEÄNDERT NACH MARX 1992A, S. 255) 332
TAB. 50: EINTEILUNG DER SCHÜLER IN GRUPPEN IM VIERFELDERSCHEMA MIT DER CBM PHONOLINO ALS PRÄDIKTOR UND DEN STANDARDISIERTEN TESTINSTRUMENTEN KNUSPEL-L (LESEVERSTÄNDNIS) UND DRT 1 (RECHTSCHREIBUNG) 333
TAB. 51: INTERPRETATION DER WERTE DER SENSITIVITÄT, ANGABEN IN PROZENT 336
TAB. 52: INTERPRETATION DER TESTEFFIZENZ DES LIKELIHOOD RATIO NACH BENDER 2001 (AUS TRÖSTER 2009, S. 96) 336
TAB. 53: INTERPRETATION DER RATZ-WERTE NACH JANSEN (1992, IN JANSEN, MANNHAUPT, MARX & SKOWRONEK 2002, S. 47), ANGABEN IN PROZENT 337
TAB. 54: DESKRIPTIVE DATEN 350
TAB. 55: ITEMSCHWIERIGKEIT FÜR PHONO 2 BIS PHONO 9, K = MAXIMALE PUNKTZAHL DES JEWEILIGEN ITEMS ÜBER ALLE ACHT MESSEZEITPUNKTE HINWEG 352
TAB. 56: TRENNSCHÄRFEKOEFFIZIENTEN VON PHONO 2 BIS PHONO 9 FÜR ALLE DREI ITEMS 353
TAB. 57: KOMMUNALITÄTEN ANHAND DER EXTRAKTIONSMETHODE (ANALYSE DER HAUPTKOMPONENTE) 354
TAB. 58: ROTIERTE KOMPONENTENMATRIX (ROTATIONSMETHODE: VARIMAX) 354
TAB. 59: KORRELATIONEN ZWISCHEN DER GESAMTSUMME ZU DEN MESSZEITPUNKTEN PHONO 2 BIS PHONO 9 356
TAB. 60: KORRELATIONEN ZWISCHEN DEN ITEMS „PSEUDOWÖRTER NACHSPRECHEN“ ZU DEN MESSZEITPUNKTEN PHONO 2 BIS PHONO 9 357
TAB. 61: KORRELATIONEN ZWISCHEN DEN ITEMS „SILBEN KLOPFEN“ ZU DEN MESSZEITPUNKTEN PHONO 2 BIS PHONO 9 358

TAB. 62: KORRELATIONEN ZWISCHEN DEN ITEMS „LAUTISOLATION“ ZU DEN MESSZEITPUNKTEN PHONO 2 BIS PHONO 9 359
TAB. 63: AUFTEILUNG DER MESSUNGEN FÜR DIE BERECHNUNG DER SPLIT-HALF-RELIABILITÄT 360
TAB. 64: ERGEBNISSE ZUR SPLIT-HALF-RELIABILITÄT FÜR DREI MODELLE....361
TAB. 65: KORRELATIONEN ZWISCHEN DEN MESSUNGEN PHONO 2 UND BISC1 SOWIE PHONO 6 UND PHONO 7 UND BISC 2 364
TAB. 66: PRODUKTE-MOMENT-KORRELATIONEN DER CBM „PHONOLINO“, PHONO 2 BIS PHONO 9 (PRÄDIKTOR) UND DEN AUßENKRITERIEN LESEN (KNUSPEL-L) UND RECHTSCHREIBEN (DRT1) 366
TAB. 67: EINTEILUNG DER ROHWERTE VON PHONOLINO IN QUARTIL ZU DEN MESSZEITPUNKTEN PHONO 2 BIS PHONO 9....367
TAB. 68: GÜTEINDIZES (GERUNDET AUF VOLLE PROZENT) ALS KLASSIFIKATORISCHE ANALYSE DER CBM PHONOLINO ZUR VORHERSAGE VON RISIKOKINDERN (PR ≤ 25) UND NICHT-RISIKOKINDERN (PR > 25) IM BEREICH DER LESEFÄHIGKEIT ANHAND DES KRITERIUMS KNUSPEL-L (T-WERT DER KLASSENNORM, CUT-OFF: T-WERT = 37)....368
TAB. 69: GÜTEINDIZES (GERUNDET AUF VOLLE PROZENT) ALS KLASSIFIKATORISCHE ANALYSE DER CBM PHONOLINO ZUR VORHERSAGE VON RISIKOKINDERN (PR ≤ 25) UND NICHT-RISIKOKINDERN (PR > 25) IM BEREICH DER RECHTSCHREIBFÄHIGKEIT ANHAND DES KRITERIUMS DRT (PROZENTRANG, CUT-OFF: PR = 10) 369
TAB. 70: GÜTEINDIZES (GERUNDET AUF VOLLE PROZENT) ALS KLASSIFIKATORISCHE ANALYSE DES BIELEFELDER SCREENINGS ZUR VORHERSAGE VON RISIKOKINDERN UND NICHT-RISIKOKINDERN IM BEREICH DER LESE- UND RECHTSCHREIBEFÄHIGKEIT ANHAND DES KRITERIUMS KNUSPEL-L (T-WERT DER KLASSENNORM, CUT-OFF: T-WERT = 37) UND DRT (PROZENTRANG, CUT-OFF: PR = 10) 371
TAB. 71: DESKRIPTIVE STATISTIKEN DER MITTELWERTE DES GESAMTWERTS VON PHONOLINO, SOWIE DER ITEMS "PSEUDOWÖRTER NACHSPRECHEN", "SILBEN KLOPFEN" UND "LAUTISOLATION" ZU DEN ACHT MESSZEITPUNKTEN 373
TAB. 72: DIFFERENZWERTE DER ITEMSCHWIERIGKEIT NEBENEINANDERLIEGENDER MESSUNGEN, K = MAXIMALE PUNKTZAHL DES JEWEILIGEN ITEMS ÜBER ALLE ACHT MESSEZEITPUNKTE HINWEG .375
TAB. 73: MAUCHLY-TEST AUF SPHÄRIZITÄT FÜR DIE GESAMTMESSUNG SOWIE DIE ITEMS "PSEUDOWÖRTER NACHSPRECHEN", "SILBEN KLOPFEN" UND "LAUTISOLATION" 376
TAB. 74: T-TESTS ZU JEDER ZWEITEN AUFEINANDERFOLGENDEN PHONOLINO-MESSUNG....378
TAB. 75: GLEICHUNGEN DER LOGISTISCHEN REGRESSION FÜR MODELL 1 (PHONOLINO-MESSUNGEN) UND MODELL 2 (PHONOLINO-MESSUNGEN UND CFT) ZUR PROGNOSE DER ABHÄNGIGEN VARIABLEN „SCHWIERIGKEITEN IM LESEN“ (DICHOTOMISIERT ANHAND DES CUT-OFFS T-WERT = 40)....384
TAB. 76: GLEICHUNGEN DER LOGISTISCHEN REGRESSION FÜR MODELL 1 (PHONOLINO-MESSUNGEN) UND MODELL 2 (PHONOLINO-MESSUNGEN UND CFT) ZUR PROGNOSE DER ABHÄNGIGEN VARIABLEN „SCHWIERIGKEITEN IM RECHTSCHREIBEN“ (DICHOTOMISIERT ANHAND DES CUT-OFFS PROZENTRANG PR = 10) 386
TAB. 77: OMNIBUS-TEST DER MODELLKOEFFIZIENTEN, ÜBERPRÜFUNG DES MODELLS AUF SIGNIFIKANZ ZUR PROGNOSE VON KINDERN MIT SCHWIERIGKEITEN IM LESEN IM ERSTEN SCHULJAHR SOWIE VERGLEICH DER BEIDEN AUFGESTELLTEN MODELLE 1 UND 2 387
TAB. 78: OMNIBUS-TEST DER MODELLKOEFFIZIENTEN, ÜBERPRÜFUNG DES MODELLS AUF SIGNIFIKANZ ZUR PROGNOSE VON KINDERN MIT SCHWIERIGKEITEN IM RECHTSCHREIBEN IM ERSTEN SCHULJAHR SOWIE VERGLEICH DER BEIDEN AUFGESTELLTEN MODELLE 1 UND 2 388
TAB. 79: GÜTEINDIZES (GERUNDET AUF VOLLE PROZENT) ALS KLASSIFIKATORISCHE ANALYSE DER MODELLE 1 UND 2 IM BEREICH DER LESEFÄHIGKEIT ANHAND DES KRITERIUMS KNUSPEL-L (T-WERT DER KLASSENNORM, CUT-OFF: T-WERT = 40)....390
TAB. 80: GÜTEINDIZES (GERUNDET AUF VOLLE PROZENT) ALS KLASSIFIKATORISCHE ANALYSE DER MODELLE 1 UND 2 IM BEREICH DER RECHTSCHREIBFÄHIGKEIT ANHAND DES KRITERIUMS DRT1 (PROZENTRANG, CUT-OFF: PR = 10) 391
TAB. 81: PRODUKTE-MOMENT-KORRELATIONEN DES BISC 1 UND DES BISC 2 (PRÄDIKTOR) UND DEN AUßENKRITERIEN LESEN (KNUSPEL-L) UND RECHTSCHREIBEN (DRT1) 398
TAB. 82: VERGLEICH DER RATZ-WERTE DER KLASSIFIKATORISCHEN ANALYSE DER PHONOLINOMESSUNGEN UND DES BIELEFELDER SCREENINGS (BISC) ZU DEN KRITERIEN LESEN UND RECHTSCHREIBEN....401

TAB. 83: TEST DER INNERSUBJEKTEFFEKTE ZUR ÜBERPRÜFUNG DES HAUPTEFFEKTS DURCH EINE EINFAKTORIELLE VARIANZANALYSE MIT MESSWIEDERHOLUNG – VERGLEICH DES GESAMTWERTS, SOWIE DER DREI ITEMS ÜBER DIE ACHT MESSZEITPUNKTE PHONO 2 BIS PHONO 9 HINWEG 404
TAB. 84: VERGLEICH DER RATZ-WERTE ANDERER STUDIEN ZUR FRÜHERKENNUNG VON LESE- UND RECHTSCHREIBSCHWIERIGKEITEN AM ENDE DER 1. KLASSE IM VERGLEICH ZU DEN ERGEBNISSEN DER VORLIEGENDEN STUDIE (ANGELEHNT AN MARX & LENHARD 2010, S. 12) 434

Formeln

FORMEL 1: ANTEIL DER ZUFALLSTREFFER - E (AUS TRÖSTER 2009, S. 136/137) 65
FORMEL 2: SELEKTIONSQUOTE - SQ (AUS TRÖSTER 2009, S. 138)....66
FORMEL 3: GRUNDQUOTE - GQ (AUS TRÖSTER 2009, S. 138) 66
FORMEL 4: RELATIVE SENSITIVITÄT - RSE (AUS TRÖSTER 2009, S. 140)....67
FORMEL 5: RELATIVE SPEZIFIZITÄT - RSP (AUS TRÖSTER 2009, S. 141)....67
FORMEL 6: RELATIVE POSITIVE KORREKTHEIT BZW. POSITIVER PRÄDIKTIVER WERT - RPPW (AUS TRÖSTER 2009, S. 141)....68
FORMEL 7: RELATIVE NEGATIVE KORREKTHEIT BZW. NEGATIVER PRÄDIKTIVER WERT - RNPW (AUS TRÖSTER 2009, S. 142)....68
FORMEL 8: RATZ-INDEX (AUS TRÖSTER 2009, S. 143)....69
FORMEL 9: BERECHNUNG DER ITEMSCHWIERIGKEITEN....320
FORMEL 10: BERECHNUNG DER ITEMSCHWIERIGKEIT BEI MEHRSTUFIGEN ITEMS 320
FORMEL 11: VERKÜRZTE FORMEL ZUR BERECHNUNG DER ITEMSCHWIERIGKEIT BEI MEHRSTUFIGEN ITEMS....320
FORMEL 12: PRODUKT-MOMENT-KORRELATION (NACH BORTZ & DÖRING 2006, S. 219) 321
FORMEL 13: BERECHNUNG DES ALPHA-KOEFFIZENTEN (BORTZ 2005, S. 559, NACH BORTZ & DÖRING 2006, S. 199) 328
FORMEL 14: GESAMTQUADRATSUMME EINER EINFAKTORIELLEN VARIANZANALYSE MIT MESSWIEDERHOLUNG....339
FORMEL 15: BERECHNUNG INNERHALB DER EINFAKTORIELLEN VARIANZANALYSE 340
FORMEL 16: BERECHNUNG DER EFFEKTSTÄRKE NACH COHEN ÜBER DAS PARTIELLE ETA-QUADRAT 340
FORMEL 17: LOGISTISCHE REGRESSIONSFUNKTION (VGL. UNIV. ZÜRICH 2016B)....344
FORMEL 18: LOGIT (Z) – LINEARES REGRESSIONSMODELL 344
FORMEL 19: BERECHNUNG DER ODDS RATIO IN SPSS (VGL. UNIV. ZÜRICH 2016)....345
FORMEL 20: REGRESSIONSMODELL 1 FÜR DIE ACHT PHONOLINO-MESSUNGEN ALS PRÄDIKTOREN....345
FORMEL 21: REGRESSIONSMODELL 2 FÜR DIE JEWEILS ACHT PHONOLINO-MESSUNGEN UND DEN KOGNITIVEN LEISTUNGEN DER PROBANDEN IM LETZTEN KINDERGARTENJAHR ALS PRÄDIKTOREN 346
FORMEL 22: BERECHNUNG DER EFFEKTSTÄRKE NACH COHEN (1992)....348

Abkürzungsverzeichnis

ATI	Aptitude-Treatment-Interaction
BISC	Bielefelder Screening
CBA	curriculum-based assessment
CBM	curriculumbasierte Messung
CBM-LL	CBM auf der Basis der Technik des 1-Minute-Lauten-Lesens
CFT 1	Grundintelligenztest Skala 1
CIBS-R	Comprehensive Inventory of Basic Skills–Revised
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
DIBS	Dynamic Indicators of Basic Skills
DRT 1	Deutscher Rechtschreibtest für 1. Klassen
EL-IGDIs	Early literacy IGDIs
GOMs	general outcome measures
HASE	Heidelberger Auditives Screening in der Einschulungsuntersuchung
IEL-1	Inventar zur Erfassung der Lesekompetenz
IGLU	Internationale Grundschul-Lese-Untersuchung
Knuspel-L	Knuspels Leseaufgaben
LDL	Lernfortschrittsdiagnostik Lesen
LISREL	Analysis of Linear Structural Relationships
LRS	Lese-Rechtschreibschwäche
NLCB	No Child Left Behind Act
Phono	Phonolino
PISA	Programme for International Student Assessment
PR	Prozentrang
RTI	Response-To-Intervention
VSL	Verlaufsdiagnostik sinnerfassenden Lesens

Einleitung

Aus aktuellem Anlass rückt die Diskussion über lese-rechtschreibschwache Schüler[1] durch die Wiederholung der PISA[2]-Studien in 2015 (vgl. OECD 2016) und den damit einhergehenden immer nur leicht über dem OECD-Durchschnitt liegenden Ergebnissen deutscher Jugendlicher wieder in den Mittelpunkt der Öffentlichkeit. Im Allgemeinen hat sich das Durchschnittsergebnis in Deutschland seit 2006 verbessert, während der OECD-Durchschnitt gleichgeblieben ist. Dennoch konnten 16 % der Schülerinnen und Schüler in Deutschland den Anforderungen des Grundkompetenzniveaus der Stufe 2[3] in der Lesekompetenz nicht genügen (vgl. OECD 2016, S. 3). Insbesondere für die in der PISA-Studie schon mehrfach bemängelte Lesekompetenz zeigen sich Defizite in einer Kulturtechnik, von der auch der schulische und berufliche Erfolg der Schüler abhängig ist. Die Autoren der IGLU-Studie in 2011 (vgl. Tarelli, Valtin, Bos, Bremerich-Vos & Schwippert 2012, S. 13) verweisen auf ähnliche Ergebnisse: 15,4 % der Viertklässler verfügen nicht über ausreichende Lesekompetenzen. In kaum einem anderen schulischen Bereich haben Lernprobleme so weitreichende Konsequenzen für die Zukunft der Schüler wie im Lesen und Rechtschreiben. Schüler entwickeln aufgrund ihrer Schwächen ein negatives Selbstkonzept und ihre allgemeine Leistungsbereitschaft und -motivation nimmt ab (vgl. Marx 1992a, S. 251). Bereits in den ersten Jahren wird die Grundlage für späteren Erfolg oder Misserfolg in der Schullaufbahn gelegt (vgl. Bergk 1980, S. 11). Schwierigkeiten im Lesen und Schreiben können besonders in den Anfängen der Schulzeit schwerwiegende und langfristige Folgen haben, Lese- und Rechtschreibschwierigkeiten manifestieren sich und erschweren auch den Wissenserwerb in anderen Fachgebieten.

Nachgewiesene systembedingte soziale Benachteiligung (vgl. bspw. Solga & Powell 2006, S. 175ff.) und die daraus abgeleitete Forderung nach einer Chancengerechtigkeit generieren den Anspruch an eine angemessene individuelle Förderung eines jeden zukünftigen Bürgers (vgl. Jeske 1981, S. 225). Der berufliche Werdegang des einzelnen Individuums in unserer Gesellschaft basiert auf einer adäquaten schulischen Ausbildung, die allen Kindern ein Fundament zur beruflichen Weiterentwicklung bietet. Das Erlernen des Lesens und Schreibens besitzt in der heutigen Gesellschaft eine unabdingbare Relevanz und ist eine Voraussetzung zur Beschaffung von Informationen, zum Erwerb und Erhalt von Wissen und zur aktiven Teilhabe

1 Im weiteren Verlauf der Arbeit wird aus Gründen der besseren Lesbarkeit überwiegend die männliche Form verwendet.
2 PISA bedeutet "Programme for International Student Assessment".
3 Das Grundkompetenzniveau Stufe 2 entspricht einem Leseniveau, das es den Schülerinnen und Schülern ermöglicht effektiv und produktiv am Leben teilzuhaben (vgl. OECD 2016, S. 3).

am gesellschaftlichen Leben. Heutzutage darf der Schriftspracherwerb kein gesellschaftliches Privileg darstellen, sondern muss ein selbstverständlicher Bestandteil der Grundschulpädagogik im frühen Kindesalter sein. Die Teilnahme am kulturellen und politischen Leben ist nur durch ausreichende Kenntnisse der Schriftsprache möglich. Die sozialen Chancen der Menschen unserer Gesellschaft, die die Schriftsprache nur unzureichend beherrschen, sind drastisch eingeschränkt. Verlaufsbeobachtungen konnten zeigen, dass die Schüler mit Defiziten im Lesen und Schreiben häufig keinen ihren kognitiven Kompetenzen entsprechenden Schulabschluss erreichen. Mögliche Folgen von schwachen Leistungen im Lesen und Rechtschreiben sind bis ins Erwachsenenalter der Analphabetismus, zusätzlich aber auch emotionale Belastungen, eine verminderte Leistungsmotivation sowie schlechte Ausbildungs- und Berufschancen (vgl. zusammenfassend Marx 1992a, S. 251).

Die in vielen Bundesländern eingeführte Schuleingangsphase, der Verzicht auf die Zurückstellung von Kindern im schulpflichtigen Alter und der immer weiter verbreiteten Umsetzung der Inklusion stellen die Lehrer vor die Problematik einer zunehmend heterogener strukturierten Schülerschaft. Dies hat die Notwendigkeit zunehmender Individualisierung der Lernprozesse im Unterricht, aber auch vorschulischer Fördermaßnahmen zur Folge. Der Kindergarten[4] wird gesellschaftspolitisch gesehen derzeit auch als der Ort frühkindlicher Bildung betrachtet und beschreibt damit die erste Stufe unseres Bildungssystems (vgl. Liegle 2008, S. 85f.; Tietze 2010, S. 50f.). Eine besonders hohe Relevanz der Betrachtung der Bildungseinrichtung „Kindergarten“ bestätigt die hohe Besuchsquote von 90 % der drei- bis sechsjährigen Kinder (vgl. Tietze 2010, S. 51). Der Kindergarten muss durch sozial- und bildungspolitische Maßnahmen immer wieder auf neue politische Herausforderungen reagieren. Insbesondere nach der Verbreitung der Ergebnisse der PISA-Studie wurden in fast allen Bundesländern Bildungsleitlinien oder -programme für den Vorschulbereich entwickelt (vgl. a. a. O., S. 91). Die (noch) mangelnde gesellschaftliche Akzeptanz des Elementarbereichs zeigt sich unter anderem in der Bezahlung der Erzieher und in der mangelnden Ausbildungsqualität. Beides deutet darauf hin, dass die Frühpädagogik noch in ihren Anfängen steckt (vgl. Hemmerling 2007, S. 15; Tietze 2010, S. 63).

Immer stärker rückt die primäre Prävention, eine frühzeitige Förderung durch bereits im Kindergarten stattfindende Präventionsmaßnahmen, in den Mittelpunkt. Der Grundgedanke von Prävention ist es, eine schwerwiegende Beeinträchtigung gar nicht erst entstehen zu lassen. Dies ist sowohl gesellschaftlich als auch im wissenschaftlichen Bereich ein anerkanntes Vorgehen,

4 Ein Kindergarten wird hier als institutionelle Betreuungsstätte für Kinder ab 3 Jahren bis zum Schuleintritt bezeichnet (vgl. Krüger & Rauschenbach 1997, S. 96-99).

jedoch sind Präventionsprogramme, die in der Forschung ausreichend validiert und in der Praxis praktikabel sind, nur rar gesät (vgl. Hartke & Koch 2008, S. 37). Um aber eine Präventionsmaßnahme zu entwickeln, bedarf es einer wissenschaftlichen Fundierung und auch angemessener Früherkennungsverfahren der jeweiligen Schwierigkeiten. Derzeit haben sich besonders im frühkindlichen Bildungsbereich immer mehr informelle Verfahren etabliert, die aufgrund der Unzufriedenheit mit vorhandenen formellen Verfahren zahlreich genutzt werden. Eine bereichsübergreifende standardisierte durchgängig eingesetzte Diagnostik in Kindergärten konnte bis dato noch nicht institutionalisiert werden (vgl. Knauf & Schubert 2008, S. 4). Hartke und Koch (2008, S. 39) fassen die zentralen Aufgaben einer erfolgreichen Präventionsforschung zusammen:

- *„Erkenntnisgewinn über die Ursachen einer Störung und damit Theoriebildung,*
- *Entwicklung von Verfahren zur Früherkennung,*
- *Entwicklung von Handlungskonzepten und deren Evaluation,*
- *Effizienzabwägungen“*

In der hier vorliegenden Untersuchung soll ein Diagnostikkonzept zur Früherkennung von Lese- und Rechtschreibschwierigkeiten im Vorschulalter entwickelt und evaluiert werden. Im Folgenden wird die Relevanz der vorliegenden Studie begründet, die Fragestellungen und Hypothesen der Untersuchung präsentiert, sowie ein Überblick über die folgenden Kapitel gegeben.

Die Relevanz der Forschung im Bereich des Schriftspracherwerbs ist aus gesellschaftlicher Perspektive aufgrund der hohen Bedeutung der Schrift unbestritten. Es bleiben hierzu noch folgende Fragen zu klären:

1) Sind Lese-Rechtschreibschwierigkeiten so häufig und ihre Auswirkungen so schwerwiegend, dass eine frühe Diagnostik gerechtfertigt ist?
2) Gibt es relevante Prädiktoren, anhand derer dies festgemacht werden kann?
3) Wann ist die Prognose sinnvoll?
4) Gibt es Grenzen einer möglichen Prognose?

zu 1: Sind Lese-Rechtschreibschwierigkeiten so häufig und ihre Auswirkungen so schwerwiegend, dass eine frühe Diagnostik gerechtfertigt ist?

Marx (1992a, S. 250) fasst die Notwendigkeit einer frühen Identifikation von Schwierigkeiten im Lesen und Rechtschreiben wie folgt zusammen:

> - *„Langzeitprognosen von Kindern mit Lese- und Rechtschreibschwierigkeiten sind eher ungünstig.*
> - *Eine Frühförderung kann zukünftige Schwierigkeiten verringern helfen.*
> - *Die Prävalenzrate für Kinder mit Lese- und Rechtschreibschwierigkeiten ist substantiell.“*

Die Folgen von Schwierigkeiten im Lesen und Rechtschreiben in der Schule zeigen sich besonders offensichtlich in der Analphabetismus-Forschung. Kinder, die im Schriftspracherwerb in der Schule bereits Schwierigkeiten entwickeln, stehen ohne angemessene Förderung vor der Gefahr, ihre Defizite zu manifestieren und vor Schulaustritt die Schriftsprache nicht ausreichend zu erlernen. Der Bundesverband Alphabetisierung e.V. geht für die Bundesrepublik Deutschland zurzeit von vier Millionen funktionalen[5] Analphabeten aus (vgl. Döbert & Hubertus 2000, S. 20f.).

Perfetti, Georgi und Beck (1993, S. 193) beschreiben diesen Zusammenhang wie folgt:

> *“Some children will acquire the alphabetic principle on their own. Children who are advantaged by literate homes may be in a good position to discover the alphabetic principle and others will discover it despite instruction that does not make it especially easy. Still, other children will not discover the alphabetic principle and their ability to read will be severely hampered.”*

Häufig werden Probleme im Schriftspracherwerb erst sehr spät erkannt. Schwierigkeiten im Lesen und Schreiben werden oft nach einer langen Zeit der Frustration und erfolglosen Übens erst nach mehreren Jahren der Schulzeit registriert, was vielfach mit einer emotionalen, psychischen und sozialen Belastung des Kindes einhergeht. Die Angst vor dem Versagen kann sich dann auch auf andere schulische Bereiche auswirken, sie kann sogar zum Schulschwänzen oder zu massiven Verhaltensproblemen führen. In vielen Fällen treten die Schwierigkeiten bis hin in die weiterführende Schule oder in das Berufsleben auf (vgl. Esser & Schmidt 1993, S. 100ff.; Haffner, Zerahn-Hartung, Pfüller, Parzer, Strehlow & Resch 1998, S. 124ff.). Esser, Wyschkon und Schmidt (2002, S. 235ff.) konnten in einer Längsschnittstudie sogar feststellen,

5 „Funktional“ bedeutet in diesem Zusammenhang, dass die betroffenen Personen mit ihren schriftsprachlichen Fähigkeiten die Funktion der Schrift in ihrer Gesellschaft nicht nutzen können, jedoch durchaus geringe Kenntnisse der Schriftsprache vorweisen (vgl. Döbert & Hubertus 2000, S. 20f.).

dass 26 % der nun 25-jährigen als legasthen diagnostizierten Personen arbeitslos sind[6]. Weber (2003, S. 406) konnte beispielsweise in ihrer Untersuchung mit 823 Kindern feststellen, dass die Mehrheit der lese-rechtschreibschwachen Schüler nach dem Schulwechsel im fünften Schuljahr die Hauptschule besuchten. Eine erfolgreiche Therapie ist langwieriger und der Leistungsabstand zu anderen Kindern des gleichen Alters hat sich bereits enorm vergrößert. Die Intervention erweist sich häufig als kompliziert und kostspielig mit meist mäßigem Ergebnis (vgl. zusammenfassend Marx 1992a, S. 251f.). Es zeigt sich eine Nachfrage nach möglichst frühzeitiger Intervention oder wenn gar möglich einer Prävention der Lese-Rechtschreibschwäche[7], um die Belastungen der Kinder und den Aufwand einer anschließenden Förderung sowie mögliche langfristige Folgen gering zu halten. Es liegt somit auf der Hand danach zu streben, Prävention und Früherkennung im Bereich der Lese-Rechtschreibschwäche zu betreiben und Schwierigkeiten bereits in ihrer Entstehung zu erkennen. Die Früherkennung von Risikokindern ist die Voraussetzung für präventive Förderung. Das Vorschulalter und das Schuleintrittsalter sind dabei wichtige Phasen bezüglich der Entwicklung der Schriftsprache und somit auch der Entstehung von Schwierigkeiten. Unterdurchschnittliche schriftsprachliche Leistungen fallen oft erst in höheren Klassenstufen auf, wenn auch ungeübte Schriftsprachleistungen abverlangt werden und die Schüler durch reines Auswendiglernen der Diktate ihre Schwächen nicht mehr kompensieren können. Dieser Fakt spricht für eine frühzeitige Diagnostik, um Interventionsmaßnahmen zu initiieren, bevor sich Schwierigkeiten manifestiert haben.

zu 2: Gibt es relevante Prädiktoren, anhand derer dies festgemacht werden kann?

Eine Diagnostik im Vorschulalter dient nicht einer Einstufung und damit Stigmatisierung des Kindes. Sie soll eine individualisierte Förderung derjenigen Kinder vorbereiten, die Schwierigkeiten in den Vorläuferfertigkeiten des Schriftspracherwerbs haben. In einer Vielzahl von wissenschaftlichen Arbeiten ließ sich in den letzten Jahren nachweisen, dass die phonologische Bewusstheit einen Anteil der schriftsprachlichen Fähigkeiten in der Schule aufklären kann (vgl. u. a. die LOGIK-Studie von Näslund & Schneider 1993, 1996; Schneider 1989; Schneider & Näslund 1992, S. 256ff., die Untersuchungen im deutsch-österreichischen Raum von Wimmer, Landerl, Linortner & Hummer 1991; Landerl, Linortner und Wimmer 1992; Landerl & Wimmer 1994; Wimmer, Landerl & Schneider 1994, die Studien der Bielefelder Forschungsgruppe zur Früherkennung von Risiken der Lese-Rechtschreibschwäche

6 In der Kontrollgruppe waren es nur 4% (vgl. Esser, Wyschkon & Schmidt 2002, S. 235ff.).

7 In der vorliegenden Arbeit werden unter lese-rechtschreibschwachen Kindern im Folgenden die Kinder verstanden, die Schwierigkeiten im Lesen und Schreiben unabhängig von ihren kognitiven Fähigkeiten aufweisen. Auf die Terminologie der „Legasthenie“ wird aufgrund der möglichen Assoziation der in der Literatur diskutierten Diskrepanzdefinition weitestgehend verzichtet.

von Jansen, Mannhaupt, Marx & Skowronek 2002; Skowronek & Marx 1989a/b, oder die Studien der Nürnberger Gruppe von Einsiedler, Frank, Kirschhock, Martschinke & Treinies 2002; Kirschhock 2003). Der frühzeitigen Intervention vor dem Schriftspracherwerb wird eine außerordentlich hohe Bedeutung zugemessen.

zu 3: Wann ist die Prognose sinnvoll?

Eine Prognose ist dann sinnvoll, wenn es möglich ist anhand der Diagnose und Prognose Lese-Rechtschreibschwierigkeiten vorzubeugen. Dies bedeutet, dass eine Diagnostik nur dann als sinnvoll erachtet werden kann, wenn sich durch Feststellung eines Defizits auch Handlungsmaßnahmen für eine Förderung ableiten lassen und die identifizierten Kinder durch eine geeignete Frühförderung so gefördert werden können, dass sich ihre Defizite beheben lassen und somit einer Lese- und Rechtschreibproblematik vorgebeugt werden kann. Eine Vielzahl an Studien konnte zeigen, dass sich die phonologische Bewusstheit trainieren und sich dadurch die Voraussetzungen für einen erfolgreichen Schriftspracherwerb verbessern lassen (vgl. u. a. Bradley & Bryant 1991, S. 37ff.; Lundberg, Frost & Peterson 1988, S. 263ff.; Bus & Ijzendoorn 1999, S. 406ff.; Byrne & Fielding-Barnsley 1989, S. 315ff., 1991, S. 451ff., 1993, S. 104ff.; Blumenstock 1979, S. 178f.; Fröhlich, Metz & Petermann 2009, S. 204ff.; Küspert & Schneider 2006, S. 121ff.; Blaser, Preuss & Felder 2010, S. 183; Einsiedler, Frank, Kirschhock, Martschinke & Treinies 2002, S. 194ff.; Schneider, Visé, Reimers & Blaesser 1994, S. 186ff.).

zu 4: Gibt es Grenzen einer möglichen Prognose?

Die Grenzen einer möglichen Prognose wurden durch eine Untersuchung von Marx und Weber (2006, S. 194ff.) aufgezeigt. Marx und Weber (ebd.) führten eine Replikationsstudie zum Bielefelder Screening (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002) durch, um die vorschulische Prädiktion von Lese- und Rechtschreibschwierigkeiten zu prüfen. Das Bielefelder Screening (BISC) ist in der Praxis ein angesehenes und weit verbreitetes Verfahren, das bei einer ersten Untersuchung sehr gute prognostische Werte erwarten ließ. In der Replikationsstudie zeigte sich jedoch, dass die sehr guten Kennwerte (insbesondere der RATZ-Index[8], der in der Studie von 2006 sogar mit einem untersten RATZ-Wert von 16 % dargelegt wurden) der ersten Evaluierungsstudie in der Replikation nicht haltbar waren. Die Kritik der Untersuchungsleiter bezieht sich insbesondere darauf, dass sich die Vorhersage der Kinder mit einer Lese-Rechtschreibproblematik in der Schule auf ausschließlich 20 Kinder der Evaluierungsstudie beziehen (vgl. a. a. O., S. 205). Aufgrund der instabilen Messungen

8 Der RATZ-Wert ist einer der Kennwerte der prognostischen Validität eines Screening-Verfahrens und beschreibt den relativen Anstieg der Trefferquote gegenüber der Zufallsquote (vgl. dazu Kap. I.2.1.3.2).

schlussfolgern die Autoren, dass von den Ergebnissen des BISC keine Frühförderung abhängig gemacht werden sollte, da bei einer Auswahl der durch den BISC identifizierten Risikokinder die Gefahr ausgehe, dass ungefähr nur die Hälfte der Kinder, die später tatsächlich eine Problematik in den schriftsprachlichen Fähigkeiten aufweisen, gefördert werden. Trotz der massiven Kritik am Bielefelder Screening konstatieren die Autoren das Bielefelder Screening als das „am besten evaluierte Verfahren zur Vorhersage von Lese-Rechtschreibschwierigkeiten im deutschsprachigen Raum“ (vgl. a. a. O., S. 206). Dies macht die missliche Lage deutlich, dass ein nicht aussagekräftiges Verfahren zwar am besten evaluiert wurde, aber dennoch nicht zur Auswahl von Fördergruppen geeignet scheint.

Dieses Dilemma der Frühförderung bzgl. häufiger Fehlklassifikationen (vgl. auch Marx 1992a, S. 253f.) zeigt die Notwendigkeit einer anderen Möglichkeit der Identifikation von Kindern im Vorschulalter. Eine Förderung innerhalb der Gruppe beziehungsweise innerhalb eines bestimmten Bereiches kann auch als eine didaktische Förderung beschrieben werden, die auf den Lernstand der Kinder angepasst ein Angebot stellt, um einen Lernzuwachs des Kindes zu erreichen. Arnold und Hartig (2008, S. 116) weisen dabei auf das „Erfolgskriterium“ der diagnostischen Aufgabe und einer Notwendigkeit der Begleitung des Lernfortschrittes der Kinder hin. Die Wirksamkeit der Förderung sollte auch durch einen individuellen Ansatz erfolgen. Ein Erfolg einer Förderung ließe sich dann anhand von Einzelmessungen ablesen und beschreibt den Vergleich zwischen dem individuell erwarteten Lernstand und dem, der tatsächlich erreicht wurde. Diese Art von Einschätzung des Lernstands von Kindern lässt sich durch Veränderungsmessung erreichen.

Marx und Lenhard (2010, S. 20f.) beschreiben einen Überblick der Screening-Verfahren und deren Kennwerte und kommen ähnlich wie Marx und Weber (2006, S. 206) zu dem Schluss, dass eine Einordnung der Kinder anhand vorhandener Screening-Verfahren zur expliziten Förderung dieser Kinder nicht möglich sei. Sie verweisen auf Chancen, jedoch auch auf Konsequenzen durch Fehlklassifikationen eines Screenings im vorschulischen Bereich zur Vorhersage von Lese- und Rechtschreibleistungen in der Schule. Zudem wurde in der Vergangenheit die derzeit gängige Klassifikation von Risiko- und Nicht-Risikokindern anhand eines Werts (Cut-Off) kritisch diskutiert und deren Praktikabilität angezweifelt, da ein Kind mit einem Prozentrang von 25 beispielsweise als Nicht-Risikokind, ein Kind mit dem Prozentrang von 24, also einer Differenz von einem Prozentrangpunkt, als Risikokind betrachtet wurde (vgl. Brügelmann 2005, S. 146ff.).

Marx und Lenhard (2010, S. 20) empfehlen stattdessen eine lernprozessbegleitende Diagnostik:

> *„Die bestehenden Screening-Verfahren lassen keine sicheren Prognosen zu. Daher sollte die Forderung der Autoren der meisten Screening-Verfahren ernst genommen werden, dass ein auffälliges Screening-Ergebnis von einer umfangreicheren Diagnostik gefolgt werden muss. Zudem scheint eine Prozess-diagnostik erfolgversprechender als eine Prognose anhand eines einmal durchgeführten Verfahrens."*

Die Relevanz der Forschungstätigkeiten auf dem Gebiet der Früherkennung von Lese- und Rechtschreibschwierigkeiten konnten nun hinreichend argumentativ belegt werden. Im Folgenden wird kurz erläutert, warum das in dieser Untersuchung evaluierte Diagnostikkonzept namens „Phonolino" eine Neuerung auf dem Gebiet der Frühdiagnostik im Bereich des Schriftspracherwerbs darstellt und welche Fragestellungen durch die vorliegende Studie untersucht und beantwortet werden sollen.

In der wissenschaftlichen Diskussion wird in den letzten Jahren zunehmend die Forderung erhoben prozessbegleitende Verfahren zu nutzen (vgl. bspw. Klauer 2006, S. 16ff.; Walter 2008a, S. 202ff.; Strathmann, Klauer & Greisbach 2010, S. 64ff.), die dazu in der Lage sind, nicht nur als Screening-Verfahren zur Klassifikation zu dienen, sondern den Lernprozess der Kinder im letzten Kindergartenjahr begleiten und gegebenenfalls eine Interventionsmaßnahme anhand ihrer Effektivität individuell überprüfen können, so dass diese bei Misserfolg verändert wird. Dies bietet Anlass ein Messverfahren zu entwickeln, für das bessere Ergebnisse in der Vorhersagekraft nachgewiesen werden können als die neueren Studien zeigen konnten, das aber gleichzeitig auch die Kompetenzen der Kinder im letzten Kindergartenjahr dokumentiert und eine Förderung individuell evaluiert. In den Vereinigten Staaten von Amerika hat sich in der Lernverlaufsdiagnostik seit Jahrzehnten ein formatives Verfahren „curriculum-based measurement" (CBM)[9] etabliert, dass den oben angeführten Kriterien entspricht und bereits als valides und reliables Messverfahren mehrfach evaluiert (vgl. Fuchs, Deno & Mirkin 1984; Deno 1985, 2003; Fuchs, Fuchs, Hamlett, Walz & Germann 1993; Madelaine & Wheldall 1999; VanDerHeyden, Witt, Naquin & Noell 2002; Ardoin, Witt, Suldo, Connell, Koenig et al. 2004; zusammenfassend Waymann, Wallace, Wiley, Tichá & Espin 2007, S. 85ff.) und in der Praxis eingesetzt wurde. Die CBM als Art von Messungen einer formativen Evaluation der Lernentwicklung stellt eine bedeutsame Alternative zu standardisierten oder informellen Tests dar (vgl. bspw. Deno 2003, S. 184ff.; Klauer 2006, S. 16; Walter 2008a, S. 202ff.; 2008b,

9 In deutschen Untersuchungen wird zunehmen der Begriff „curriculumbasierte Messung" verwendet, der auch in der hier vorliegenden Studie unter anderem auch mit der Abkürzung „CBM" genutzt wird.

S. 62ff.), da sie dazu in der Lage ist, den Lernprozess anhand von Wiederholungsmessungen abzubilden.

Diese Methode der Lernfortschrittsmessung soll nun in der vorliegenden Studie mit den Erkenntnissen aus dem deutschsprachigen Raum zu den Vorläuferfertigkeiten der Lese- und Rechtschreibfähigkeiten vereint werden. Dazu wurde die curriculumbasierte Messung „Phonolino“ mit neun Messungen entworfen, die der Überprüfung der phonologischen Informationsverarbeitung bei Vorschulkindern zur Früherkennung von Schwierigkeiten im Lesen und Rechtschreiben dient und im Rahmen dieser Arbeit anhand einer Längsschnittstudie evaluiert wurde.

Wie von L.S. Fuchs (2004, S. 189) empfohlen, sollte die Evaluation einer Veränderungsmessung in drei Schritten erfolgen. Im ersten Schritt (Stufe 1) wird das Messverfahren zunächst bezüglich der klassischen Gütekriterien von Messverfahren, Objektivität, Reliabilität und Validität, überprüft. Sind die Voraussetzungen des ersten Schrittes erfüllt, soll im zweiten Schritt (Stufe 2) eine Überprüfung der Änderungssensibilität der Messungen stattfinden. Hierbei wird die CBM als Veränderungsmaß betrachtet, das sensibel individuelle Lernverläufe von Kindern dokumentieren kann. Nach erfolgreicher Evaluation der ersten beiden Stufen empfiehlt Fuchs (ebd.), das Verfahren in Bezug auf seine Tauglichkeit in der praktischen Anwendung zu untersuchen. Diese Stufe ist zwingend zu berücksichtigen, da sie richtungsweisend für die Interventionsentscheidungen im Alltag des Kindergartens oder der Schule ist.

In der vorliegenden Studie soll das erstellte Messverfahren „Phonolino“ auf der ersten und zweiten Stufe überprüft werden, um die Methodik des Verfahrens zu evaluieren. Der dritte Schritt (Stufe 3) wird Teil einer Replikationsstudie sein müssen.

Im Anschluss wird der Inhalt der vorliegenden CBM Phonolino überprüft. Es wird in der weitreichenden Literatur vermutet, dass nonverbale kognitive Variablen (vgl. bspw. Pfeiffer & Zielinski 1975, S. 1ff.; Deimel, Ziegler & Schulte-Körne 2005, S. 6; Ennemoser, Marx, Weber & Schneider 2012, S.63ff; Bradley & Bryant 1985, S. 163ff.; Schneider & Näslund 1992, S. 268; Vellutino et al. 2000. S. 236; Tiu, Thompson & Lewis 2003, S. 424) eine Varianzaufklärung der Lese- und Rechtschreibleistung in der Schule leisten können. Somit wird das Prognosemodell zur phonologischen Informationsverarbeitung im letzten Teil der statistischen Überprüfung um die nonverbale Kognition als weiteren Prädiktor erweitert. Dies könnte ein Fingerzeig für weitere Studien darstellen, diesen Prädiktor zur Vorhersage von schulischen Leistungen in Früherkennungsverfahren miteinzubeziehen.

Die vorliegende Arbeit hat damit folgende Zielsetzungen und Fragestellungen:

Stufe 1 nach Fuchs (2004)

F1 Ist Phonolino ein objektives Messverfahren?

F2 Sind die Messungen von Phonolino brauchbar, also trennscharf und normalverteilt, ist die Schwierigkeit der Aufgaben angemessen und bestehen sie aus den drei Dimensionen „Pseudowörter nachsprechen", „Silben klopfen" und „Lautisolation"? (Itemanalyse)

F3 Ist Phonolino ein zuverlässiges Messverfahren, d.h. misst es exakt und genau das, was es messen soll?

F4 Sind die curriculumbasierten Messungen von Phonolino valide?

Stufe 2 nach Fuchs (2004)

F5 Erfasst Phonolino Veränderungen der Leistungen der phonologischen Bewusstheit zu den acht[10] Messzeitpunkten?

Erweiterung des Prognosemodells um den Prädiktor „nonverbale kognitive Leistung"

F6 Kann die Prognose von Lese- und Rechtschreibschwierigkeiten am Ende des ersten Schuljahrs durch die curriculumbasierten Messungen von Phonolino bei Hinzunahme des Prädiktors der nonverbalen kognitiven Leistung verbessert werden? (Prognosemodell 2)

Die Arbeit leistet damit einerseits einen Beitrag zur wissenschaftlichen Untersuchung der phonologischen Informationsverarbeitung als aussagekräftigem Prädiktor sowie zur Etablierung der in den USA weitläufig bekannten curriculumbasierten Messung als wissenschaftlichem Instrument in Deutschland. Außerdem ist der Forschungsbereich explizit auf den Bereich der Frühdiagnostik und Prävention ausgerichtet und stellt damit die Frage der bestmöglichen Früherkennung und damit rechtzeitigen Förderung von Kindern bereits im Kindergarten in den Mittelpunkt. Diese Variante der Lernfortschrittsdiagnostik stellt zudem eine gänzlich andere Herangehensweise an eine Diagnostik ins Zentrum der Betrachtung. Sie hat eine effektive individuelle Förderung zum Ziel und stellt damit den Erfolg und die Fortschritte im Vergleich zu einer klassischen „Bestandsaufnahme" als Maß in den Mittelpunkt.

Ausgehend von den Forschungsergebnissen zum Bielefelder Screening aus dem Jahr 2006 (vgl. Marx & Weber 2006, S. 254ff.) wird im *ersten Kapitel* zunächst die Notwendigkeit einer frühen

10 Nicht alle Kindergärten konnten zum ersten Messzeitpunkt bereits an der Studie teilnehmen, deshalb wurde diese Messung bei der statistischen Evaluation des Verfahrens nicht berücksichtigt.

Diagnostik zur Früherkennung von Schwierigkeiten im Lese- und Rechtschreibprozess erörtert. Dazu wird zum einen das in den letzten Jahren veränderte Bildungssystem in Bezug auf strukturelle Veränderungen der Bildungsstätten Kindergarten und Schule in Anbetracht der Inklusion betrachtet, zum anderen wird das mögliche Potenzial einer Prävention erörtert. Im Anschluss wird ein Blick auf das Ausmaß möglicher Folgen für den Fall geworfen, dass keine präventiven Maßnahmen ergriffen werden. Danach werden ausgewählte Testinstrumentarien zur Früherkennung von Schwierigkeiten im Lesen und Rechtschreiben vorgestellt und anhand von Gütekriterien eingeordnet, bewertet und verglichen. Daraus ergibt sich dann die Frage nach einer sicheren Klassifikation von Risiko- und Nicht-Risikokindern zur Vorhersage von schriftsprachlichen Leistungen. Die verschiedenen Klassifikationsproblematiken werden dargestellt und in Beziehung gesetzt.

Im *zweiten Kapitel* wird dann der Frage nach denkbaren Ursachen der Klassifikationsproblematik von Risikokindern im Vorschulalter nachgegangen. Zunächst werden ausgewählte Inhalte und deren wissenschaftliches Fundament mit einem Fokus auf die in den meisten Screening-Verfahren vertretene phonologische Informationsverarbeitung geprüft. Des Weiteren ist es möglich, dass die Fehlklassifikationen von Risiko- und Schutzfaktoren herrühren, die auf die Kinder zwischen der Messung des Prädiktors oder des Kriteriums einwirken. Es werden hierzu zum einen einige klassische Ursachen einer Lese-Rechtschreibschwäche dargelegt, zum anderen aber auch endogene Einflüsse wie die sozioökonomischen Umstände des Kindes oder der veränderte Anfangsunterricht beleuchtet, die sowohl resilient als auch vulnerabel auf die Leistungen des Kindes einwirken können. Als dritter möglicher Aspekt für die mangelhafte Vorhersagegenauigkeit wird die Methodik der Früherkennungsverfahren in Frage gestellt. Dazu werden die Ziele sowie die Vor- und Nachteile einer herkömmlichen klassischen pädagogischen Diagnostik insbesondere in sonderpädagogischen Einsatzfeldern erläutert.

Im *dritten Kapitel* wird eine methodische Alternative zu den herkömmlichen Testverfahren vorgestellt. Dazu wird der Unterschied zwischen standardisierten Testverfahren und einer systematisch formativen Evaluation bzw. einem curriculumbasierten Messverfahren zusammengefasst. Die CBM als formatives Messverfahren (Ardoin, Witt, Suldo, Connell, Koenig et al. 2004, S. 220; Graney & Shinn 2005, S. 184ff.; Walter 2008a, S. 202ff., 2008b, S. 62ff., 2009a, S. 162ff.; Klauer 2006, S. 17) ist dazu in der Lage, den Lernfortschritt eines Kindes über eine Zeitspanne hinweg zu messen und zu dokumentieren. Dadurch kann die Leistungsentwicklung und die Veränderung von Leistung sensibel erfasst werden und auf deren Grundlage eine Entscheidung für bzw. eine Modifikation der Förderung getroffen werden (vgl.

Klauer 2006, S. 17). Es werden damit sowohl die diagnostischen Vorteile, wie auch die einer aufbauenden Intervention in diesem Kapitel dargelegt.

Im zweiten Abschnitt der vorliegenden Arbeit wird die durchgeführte Untersuchung erläutert.

Zunächst wird im *vierten Kapitel* die von der Autorin konzipierten curriculumbasierten Messungen „Phonolino“ vorgestellt. Die CBM wird auf der Grundlage der im theoretischen Teil erläuterten Forschungsergebnisse konzipiert und beruht auf Komponenten der phonologischen Informationsverarbeitung, die mit drei Items zu je neun Messzeitpunkten im Kindergarten zusammengefasst werden.

Im *fünften Kapitel* werden die Fragestellungen für die vorliegende Studie abgeleitet.

Anschließend wird im *sechsten Kapitel* der methodische Aufbau der über knapp zwei Jahre durchgeführten Studie beschrieben. Dazu wird der zeitlich-organisatorische Ablauf aufgeführt und die Stichprobe der 258 teilnehmenden Kinder beschrieben. Zusätzlich werden die Bildungsinstitutionen dargestellt, die methodische Durchführung der Längsschnittstudie formuliert, die eingesetzten Untersuchungsinstrumentarien vorgestellt und dann die zur Auswertung der Daten genutzte statistische Methodik dargelegt.

Im *siebten Kapitel* werden die Ergebnisse der Studie zur Reliabilität, Validität und Änderungssensibilität sowie zur Erweiterung des Prognosemodells des betrachteten Messverfahrens Phonolino vorgestellt.

Im Anschluss werden im *achten Kapitel* die gewonnenen Ergebnisse aufgrund der im Theorieteil dargelegten Forschungslage interpretiert und eingeordnet. Zusammenfassend werden dann die dieser Studie zugrunde gelegten Fragestellungen beantwortet.

Im *neunten Kapitel* wird die vorgelegte Studie kritisch betrachtet, diskutiert und mögliche Schlussfolgerungen wie weitere Forschungsschwerpunkte und ein Ausblick hinsichtlich der zukünftigen Relevanz der curriculumbasierten Messung in Deutschland erörtert.

I. Prognose schriftsprachlicher Fähigkeiten im Vorschulalter

Ein intensiver Diskurs über den Bildungsauftrag im Elementarbereich, unter anderem ausgelöst durch die Debatte um die PISA-Studien, hat in den letzten Jahren vielfältige Forschungsinitiativen und Praxisversuche angeregt. Die Voraussetzungen, Chancen und Grenzen einer frühen Diagnostik und Förderung wurden dabei in den Fokus der Betrachtung gerückt (bspw. Barth & Gomm 2005, Brügelmann 2003a, Brunner & Schöler 2008, Einsiedler 2003, Frank, Kirschhock & Martschinke 2003). Unterschiedliche präventive Maßnahmen wurden in den Bundesländern vorangetrieben, um die Vorläuferfertigkeiten des Lesens und Rechtschreibens zu trainieren und die Kinder damit auf einen barrierefreien Übergang in die Schule vorzubereiten. Der Prognose schriftsprachlicher Fähigkeiten kam aus präventiver Perspektive dabei eine zentrale Rolle zu. Früherkennungsverfahren werden derzeit in den Kindergärten meistens in Form eines Screenings eingesetzt, mit dem die Prädiktoren in standardisierter Form erfasst werden. Standardisierte Screenings und deren Möglichkeiten einer sicheren Prognose sind dabei in der letzten Zeit in wissenschaftlichen Kontexten aufgrund mangelhafter Vorhersagewerte kritisch diskutiert worden (vgl. bspw. Marx & Weber 2006, S. 254ff.).

Das Ziel eines Screenings sollte nach Tröster (2009, S. 68) sein, dass dieses ökonomisch einsetzbar ist, Frühindikatoren einer Störung erfasst, sowie dass es eine Filterfunktion darstellt, anhand deren eine Teilgruppe einer Population von Personen identifiziert werden kann, die mit hoher Wahrscheinlichkeit diese Störung entwickeln wird. Diese Kinder werden dabei auch als Risikokinder/-personen bezeichnet (vgl. Abb. 1).

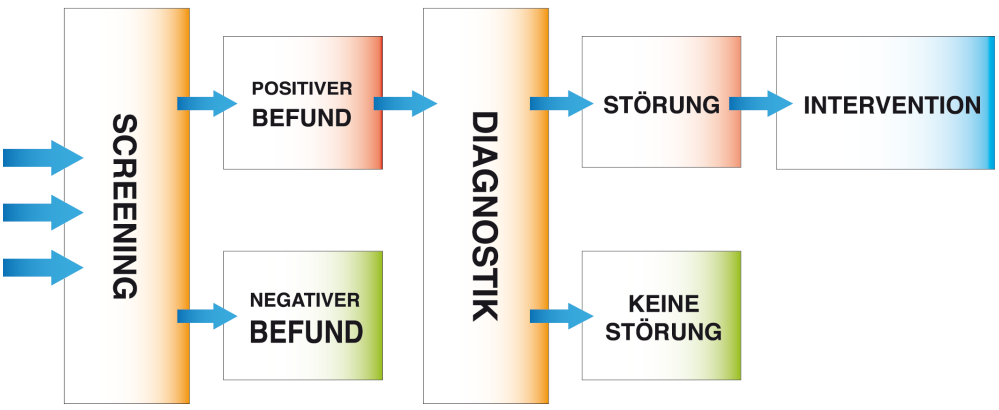


Abb. 1: Filterfunktion des Screenings (nach Tröster 2009, S. 69)

Das übliche Verfahren zur Prognose von Lese- und Rechtschreibschwierigkeiten in der Schule besteht darin, Risikokinder im Vorschulalter durch ein Screening zu identifizieren, also anhand eines Risikoindexes die Schüler einer bestimmten Gruppe zuzuordnen. Dies ist erstmalig den Entwicklern des Früherkennungsinstruments „Bielefelder Screening“ (BISC) für den Vorschulbereich gelungen (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002, S. 33ff.).

Marx und Lenhard (2010, S. 1) betonen bezüglich der Bewertung von Screening-Verfahren die Wichtigkeit der Überprüfung der prognostischen Validität der Verfahren:

> *„Vorschulische und schulische Screeningverfahren zur Vorhersage späterer Lese- und Rechtschreibschwierigkeiten sind dann wertvoll, wenn sie prognostisch valide sind, und wenn sich die Informationen für den Einsatz von Präventionsmaßnahmen nutzen lassen.“*

Die Autoren des BISC konnten diesbezüglich in der damaligen Untersuchung besonders gute Kennwerte der prognostischen Validität erreichen (vgl. Kap. I.2.2.1) und mit einem RATZ[11]-Wert von 52 % für die Rechtschreibleistung und von 71 % für die Leseleistung Werte erzielen, die beweisen, dass die Prädiktionsleistung des Bielefelder Screenings weit über den Zufallswerten liegt. Jüngere Untersuchungen (vgl. Schneider & Näslund 1993, S. 273ff.; Brügelmann 2003a, S. 1ff.; Marx & Weber 2006, S. 254ff.) zum Bielefelder Screening dagegen konnten diese Befunde nicht replizieren, die Klassifikation der Kinder mit dem BISC erbrachte schlechtere Resultate als die erste Studie. Insbesondere die Studie von Marx und Weber (2006, S. 254ff.) konnte aufzeigen, dass das Screening-Verfahren BISC doch nicht zwangsläufig detaillierte und zuverlässige Prognosen liefern kann. Die Ergebnisse zeigten, dass durch das Verfahren keine sichere Prognose der Risiko- und Nicht-Risikokinder gewährleistet werden kann. Der niedrigste RATZ-Wert ließ sich sogar mit 16 % (Kriterium Rechtschreibung) verzeichnen. Nach Jansen (1992, in Jansen, Mannhaupt, Marx & Skowronek 2002, S. 47) lässt erst ein RATZ-Wert von mehr als 34 % eine gute, aber noch unspezifische Klassifikation zu, eine sehr gute Klassifikation kann mit mehr als 66 % erreicht werden.

Aufgrund dieser Forschungserkenntnisse stellt sich im ersten Abschnitt dieses Kapitels die generelle Frage nach der Notwendigkeit einer Prognose und einer möglichen Prävention. Die Prognose von Schwierigkeiten steht im Spannungsfeld von bildungspolitischen

11 Der RATZ-Wert ist eines der Kennwerte der prognostischen Validität eines Screening-Verfahrens und beschreibt den relativen Anstieg der Trefferquote gegenüber der Zufallsquote (vgl. dazu Kap. I.2.1.3.2).

Hintergründen wie beispielsweise dem strukturellen Wandel unseres Schulsystems wie auch der Hoffnung auf eine mögliche Prävention von Schwierigkeiten im Schriftspracherwerb. Auch das Ausmaß der Problematik bei Nicht-Identifikation und ausbleibender spezifischer Förderung der Risikokinder und damit die Relevanz der Forschung im Bereich der Frühdiagnostik werden unter dem Aspekt der Lese-Rechtschreibschwäche beschrieben.

Marx und Weber (ebd.) stellen durch die Ergebnisse ihrer Studie auch die Prädiktionsleistung anderer Screening-Verfahren im Vorschulalter zur Früherkennung einer Lese-Rechtschreibschwäche in Frage. Sie diskutieren die vorhandenen Screenings und deren Gütekriterien kritisch. Fünf dieser Verfahren zur Vorhersage von Schwierigkeiten im Lesen und Rechtschreiben werden im zweiten Abschnitt dieses Kapitels vorgestellt. Es wird zu begutachten sein, wie die Prädiktionsleistung dieser Testinstrumente zu beurteilen ist und welche Unterschiede in den Verfahren zu verzeichnen sind. Im Anschluss daran wird analysiert, welche Schwierigkeiten bei der Klassifikation von Kindern zur Früherkennung der Lese- Rechtschreibproblematik auftreten.

Im ersten Kapitel sollen damit folgende zentrale Fragen diskutiert werden:

1) Welche Gründe sprechen für eine frühe Identifikation von Risikokindern? (Kap. I.1)
2) Welche Testinstrumentarien gibt es schon und welche Kennwerte für Screening-Verfahren können diese liefern? (Kap. I.2)
3) Welche Schwierigkeiten bestehen bei der Klassifikation von Risiko- und Nicht-Risikokindern? (Kap. I.3)

1 Notwendigkeit früher Diagnostik und Prävention

Infolge der ernüchternden Ergebnisse zu Screening-Verfahren zur Prognose schriftsprachlicher Fähigkeiten ist Skepsis gegenüber der Notwendigkeit einer frühen Diagnostik und Prävention im Kindergartenalter anzubringen.

Tröster (2009, S. 66f.) begründet die Notwendigkeit einer Früherkennung mit Aspekten der Entwicklungspsychologie und der Neurologie:

- *Entwicklungspsychologisch* gesehen greift das Konzept der „sensiblen Phasen“, in denen das Kind für bestimmte Lernvorgänge besonders aufnahmefähig ist. Anders als in der Verhaltensbiologie, in der ebenfalls der Begriff der Prägung beachtet werden muss, wird in der Entwicklungspsychologie die gesamte frühe Kindheit als „sensible Phase“ angenommen, die die Entwicklung des Kindes „prägt“. Werden in den frühen Stadien bestimmte Lernprozesse verpasst, ist der Aufwand, diese später nachzuholen, um einiges größer. Die Frühförderung sollte diese Phasen ausnutzen und gezielt Lernvorgänge initiieren bzw. fördern. Außerdem können durch eine frühe Förderung kumulative Effekte von Entwicklungsproblemen verhindert werden. Bei Entwicklungsrückständen bleibt häufig nicht nur dieser eine Rückstand erhalten, sondern die Kinder verändern sich auch in anderen Variablen, da sie über eine schlechtere Ausgangslage verfügen (vgl. Tröster 2009, S. 67).

- *Neurologisch* gesehen ist eine frühe Förderung aufgrund der neuronalen Plastizität sinnvoll. Die neuronale Plastizität beschreibt die Formbarkeit des Gehirns, das sich aufgrund von Erfahrungen wandeln kann. Die Plastizität nimmt mit dem Alter des Kindes jedoch ab, was beispielsweise beim Erlernen von Fremdsprachen sichtbar gemacht werden kann (vgl. Tröster 2009, S. 68).

Es gilt jedoch auch auf kritische Aspekte in der Forschung zur Früherkennung von Lese-Rechtschreibschwierigkeiten hinzuweisen. Einige Autoren bewerten die Früherkennung und eine damit möglicherweise einhergehende Selektion von Risikokindern im Vorschulalter als stigmatisierend. Sie warnen vor einem frühen „Labeling“, das die Kinder als „anders“ etikettiert und zugleich davor, dass diese in der Folge im Sinne des „self-fulfilling-prophecy“-Ansatzes (der selbsterfüllenden Prophezeiung) keine Leistungen über dem von ihnen erwarteten Niveau erbringen (vgl. Foster, Schmidt & Sabatino 1976, S. 58ff.).

Brügelmann (2005, S. 147) bspw. argumentiert in diesem Sinne:

> *„Die Kriterien für Schulerfolg lassen sich nicht unabhängig vom außerschulischen Kontext bestimmen – und auch die Suche nach den Ursachen von Misserfolg darf nicht auf individuelle ‚Voraussetzungen’ und auf innerschulische ‚Bedingungen’ verkürzt werden. Nur im Bewusstsein dieser Einschränkungen ist eine Diskussion über Risikoprognosen des Schulerfolgs angemessen zu führen.“*

Marx (1992a, S. 253f.) führt, ähnlich wie Brügelmann, zwei Dilemmata der Früherkennung an:

- Eine erfolgreiche Frühidentifikation und -förderung der gefährdeten Kinder kann eine Veränderung der Rangplätze mit sich bringen, so dass vorherige Grenzfälle nun laut Kriterium zu Problemfällen werden.
- Eine vollständig richtige Klassifikation der Kinder ist nicht möglich, da Prognosen grundsätzlich auch immer Fehler mit sich bringen und nicht alle Einflüsse kontrolliert werden können.

Im Folgenden wird die Notwendigkeit einer frühen Förderung unter drei Aspekten betrachtet. Zuerst werden neue Herausforderungen in der Diagnostik und Frühförderung erörtert, die auf zunehmende inklusive Tendenzen und Entwicklungen im deutschen Bildungssystem zurückzuführen sind. Im Anschluss werden die Chancen und Grenzen einer frühkindlichen Förderung zur Prävention von Schwierigkeiten im Schriftspracherwerb erörtert. Zum Abschluss soll das Ausmaß der Folgen nicht wahrgenommener präventiver Maßnahmen anhand der Stabilität und Persistenz der Lese-Rechtschreibschwäche beschrieben werden.

1.1 Veränderungen im Bildungssystem

Die erste PISA-Studie im Jahr 2000 (vgl. Baumert, Klieme, Neubrand, Prenzel, Schiefele, Schneider, Stanat, Tillmann & Weiß 2001) erzeugte einen gewaltigen Widerhall in der nationalen wie internationalen Bildungspolitik, genauso wie auch in der wissenschaftlichen Diskussion und in der pädagogischen Praxis. So wurde für den deutschen Teil der Studie nachgewiesen, dass bereits das Kompetenzniveau der 15-Jährigen im Lesen unter dem Mittelwert der OECD-Staaten lag und die Streuung der Werte überaus groß war. Es wurden immense Leistungsunterschiede zwischen sogenannten „Risikokindern“ und Kindern aus bildungsnahen Elternhäusern festgestellt und die nicht vorhandene Chancengleichheit wurde insbesondere für das deutsche Bildungssystem bemängelt (vgl. Klieme, Jude, Baumert & Prenzel 2010, S. 277ff.).

In vielen Bundesländern wurden administrative Veränderungen angestrebt und schul-organisatorische Konsequenzen gezogen. Das traditionelle dreigliedrige Schulsystem wurde in den Folgejahren schrittweise in ein zweigliedriges System umgewandelt – im Rahmen eines nahezu umfassenden parteienübergreifenden Konsenses. Die vormaligen Schularten

Hauptschule und Realschule entwickelten sich z.B. zur Regional- und Gemeinschaftsschule (Schleswig-Holstein), zur Werkrealschule (Baden-Württemberg) oder zur Mittelschule (Sachsen). Immer mehr Schüler werden über einen immer länger währenden Zeitraum gemeinsam unterrichtet, in der Hoffnung, dass sich aus der Systemveränderung den Schülern bildungsferner Elternhäuser bessere Chancen für ihren Bildungsweg eröffnen.

Auch die Strukturen innerhalb der Schulen veränderten sich bzw. sind bis heute noch im Wandel. Bildungsstandards auf Bundesebenen wurden eingeführt, Vergleichsarbeiten zur Überprüfung und zum Vergleich der Schülerleistungen etabliert und die Idee des kompetenzorientierten Unterrichts forciert (vgl. a. a. O., S. 278f.).

Progressive Veränderungen im deutschen Bildungssystem ergaben sich insbesondere aber auch im Bereich der Sonderbeschulung von Kindern durch das „Übereinkommen der Vereinten Nationen über die Rechte von Menschen mit Behinderungen“ (vgl. KMK 2010, S. 7ff.), das seit dem 26. März 2009 auch für Deutschland völkerrechtlich verbindlich ist. Hiermit rücken die Rechte der Menschen mit Behinderungen stärker in den Fokus der gesellschaftlichen Aufmerksamkeit. Unter Berücksichtigung dieses Abkommens haben die Bundesländer mit ihren Empfehlungen zur inklusiven Bildung von Kindern und Jugendlichen (vgl. KMK 2011, S. 2) richtungsweisende Reformen angestoßen. Hiermit wurden die Grundlagen für die Teilhabe von Menschen mit Beeinträchtigungen an Bildung geschaffen, die eine zunehmend inklusive Praxis in deutschen Bildungsinstitutionen, in Kindergärten wie in allgemeinbildenden und beruflichen Schulen mit sich bringt. Dabei wird die Zuständigkeit der allgemeinen Bildungseinrichtungen für *alle* Kinder und Jugendliche besonders hervorgehoben (vgl. a. a. O., S. 3f.).

Die damit einhergehende zunehmende Heterogenität der Schülerschaft stellt eine große Herausforderung für einen professionellen Umgang des pädagogischen Personals im Unterricht und auch schon im Kindergarten mit den individuell sehr unterschiedlich ausgeprägten Lernvoraussetzungen dar. Individualisierte Lernangebote werden notwendig, eine differenzierte Diagnostik zur Feststellung der unterschiedlichsten Lernausgangslagen und -entwicklungen rückt zunehmend in das Zentrum der Aufmerksamkeit. In neueren Abhandlungen (vgl. bspw. Liebers & Seifert 2012, S. 1ff.) wird immer mehr in Bezug auf diese inklusiven Ansätze auch die inklusive Diagnostik beleuchtet. Für eine Professionalisierung der pädagogischen Arbeit ist eine inklusive Diagnostik unerlässlich, um Alltagshypothesen selbstreflexiv zu überprüfen.

Die pädagogischen Zielsetzungen einer solchen Diagnostik sind auf die optimale Entwicklung aller Kinder und Jugendlichen gerichtet, um diese individuell zu fördern und ihnen damit ihre individuell bestmöglichen Chancen der gesellschaftlichen Partizipation durch eine Befähigung zum eigenständigen Handeln zu ermöglichen. Dieses Ziel lässt sich insbesondere durch eine frühzeitige Erkennung von Schwierigkeiten in der Lese- und Rechtschreibentwicklung aller Kinder erreichen. Es bedarf dazu zeitökonomisch praktikabler und dennoch den wissenschaftlichen Gütekriterien gerecht werdender, einsetzbarer Screening-Verfahren schon im Vorschulalter. Diagnostikverfahren, die lediglich zur Selektion der Kinder für die Überweisung an eine Sonderschule dienen, werden zukünftig in dieser Form der Vergangenheit angehören (vgl. ebd.).

Insbesondere für die pädagogischen Fachkräfte im Elementar- und Primarbereich werden unterstützende Verfahren benötigt, die es ihnen ermöglichen schon im Frühstadium der Bildungsprozesse der Kinder sich abzeichnende Schwierigkeiten zu erkennen, sie fachgerecht zu diagnostizieren und daraus angemessene Fördermöglichkeiten abzuleiten. In der Übergangssituation zwischen Kindergarten und Schule wird dies besonders deutlich:

> *„Zur Einschulung erscheinen dort einige Kinder mit bereits aktenkundigem Förderbedarf, mit denen die Grundschule bisher nicht in Berührung kam, weil diese Kinder – meist aus der Frühförderung bekannt – direkt einer Sonder-/Förderschule zugeführt wurden.“ (Probst & Euker 2012, S. 198)*

Professionelles Handeln der Grundschullehrer ist in dieser Praxis auf eine gute Dokumentation der Fähigkeiten angewiesen, die die Kinder aus dem Kindergarten bereits mitbringen. Um dabei über allgemeine, subjektive Einschätzungen bezüglich des Förderpotenzials von Kindern hinaus zu gelangen („Jakob konnte seine Fähigkeiten im Silbenklatschen nicht verbessern.“) sind standardisierte Testverfahren unabdingbar.
Für zielgerichtete Förderangebote kommt dabei der Vergleichbarkeit der Leistungen eine hohe Bedeutsamkeit zu:

> *„Die Vergewisserung über den Erfolg einer Fördermaßnahme kann nur gelingen, wenn Veränderungen der Schüler/innen (oder ihre Stagnation) beim wiederholten Anlegen derselben Messlatte ersichtlich sind.“ (Probst & Euker 2012, S. 201)*

Aber auch für die notwendige enge Zusammenarbeit von Lehrern, Eltern und Kindern, wie auch für einen intensiveren Austausch zwischen den Bildungsinstitutionen Kindergarten, Grundschule und weiterführende Schule sind diagnostische und präventive Anstrengungen

elementar. Eine frühzeitige Diagnostik der Vorläuferfertigkeiten des Lesens und Schreibens sowie das Ergreifen präventiver Maßnahmen bereits im Kindergarten können späteren Problemen im Lesen und Schreiben in der Schule vorbeugen und sie vielleicht sogar vermeiden. Andererseits ist eine pädagogische Diagnostik mit kontinuierlicher Dokumentation der frühen Fertigkeiten der Kinder ebenfalls bildungsbiografisch gesehen ein bedeutsamer Aspekt. Die Datenbasis und die Förderansätze aus dem Kindergarten können an die Grundschulen weitergeleitet werden, damit die Pädagogen begonnene Präventionsmaßnahmen gezielt fortsetzen können. Kindern wird durch diese Möglichkeit der Prävention die Chance gegeben, ihre Fähigkeiten auszubilden und ihre Potenziale auszuschöpfen, ohne einen Bruch zwischen Kindergarten und Schule zu erleben.

Aufgrund der oben angeführten Zweifel an bisher auf dem Markt verfügbaren Screening-Verfahren zur vorschulischen Früherkennung von Schwierigkeiten im Lesen und Rechtschreiben ist es nun auch in Anbetracht der schulstrukturellen Veränderungen in Deutschland zu überdenken, welche Diagnostikkonzepte einer Prävention und einer theoriegeleiteten Planung, Durchführung und Überprüfung von Fördermaßnahmen förderlich sind. Screenings sind im Vorschulalter insbesondere bei sogenannten „Risikokindern“ als sinnvoll zu erachten, wenn sie mit geeigneten Förderkonzepten einhergehen. Zu beachten sind jedoch auch ressourcenorientierte Vorgaben. Eine Früherkennung sollte zwar primär mit einem pädagogischen Fokus ausgestattet sein, bildungspolitische, strukturelle und ressourcenorientierte Gesichtspunkte müssen jedoch auch in ein Konzept für eine umfassende Prävention von Problemen im Schriftspracherwerb miteinbezogen werden.

Die Notwendigkeit, aber auch die Möglichkeiten einer präventiven Arbeit im Kindergarten zur Vermeidung von Schwierigkeiten im Lesen und Rechtschreiben in der Schule werden im nächsten Abschnitt erörtert.

1.2 Potenzial der Prävention

Tröster (2009, S. 66) konstatiert:

> *„Früherkennung ist nur sinnvoll, wenn eine frühzeitige Intervention mehr Erfolg verspricht als ein späterer Interventionsbeginn.“*

Früherkennung ist demnach umso besser, je länger das Frühstadium der Störung andauert (vgl. a. a. O., S. 63; vgl. Abb. 2). Damit ist zu hinterfragen, ob eine frühe Förderung der

Vorläuferfertigkeiten des Lesens und Rechtschreibens sich tatsächlich auf spätere schriftsprachliche Leistungen auswirkt.

Ansatzpunkt der neueren Präventionsforschung ist, dass davon ausgegangen wird, dass die Einschulung der Kinder nicht die „Stunde Null“ des Schriftspracherwerbs ist. Es wird in der entwicklungspsychologischen Forschung vermutet, dass nicht nur allgemein kognitive Fähigkeiten für Erfolg im Schriftspracherwerb sprechen, sondern dieser auf den Vorkenntnissen und den Fähigkeiten der Kinder aufbaut. Lese- und Rechtschreibschwierigkeiten können zu schulischem Versagen führen (vgl. auch Kap. I.1.3.2) - dem sollte möglichst früh durch geeignete Präventionsmaßnahmen entgegengewirkt werden. Anhand einer qualitativ hochwertigen und vor allem frühzeitigen Diagnostik können schon früh Daten gewonnen werden, anhand derer Präventionsmaßnahmen abgeleitet werden können.

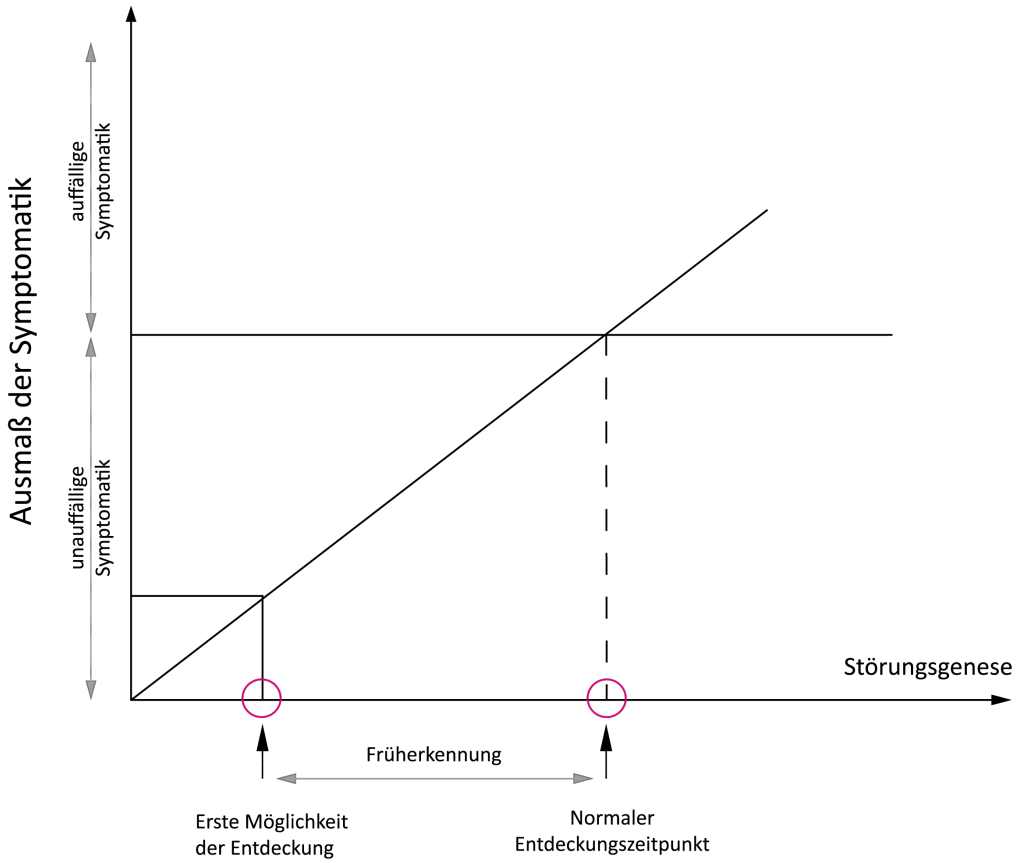


Abb. 2: Früherkennungsschema (nach Tröster 2009, S. 62)

Mannhaupt (2006, S. 105f.) betont zudem die Ausrichtung der Förderung an den aktuellen Entwicklungsständen und -möglichkeiten der Schüler. Eine Intervention mit nicht ausreichender Berücksichtigung von Vorkenntnissen der Kinder ist nicht erfolgreich. Eine

individualisierte Vorgehensweise dagegen ist eine gute Voraussetzung für eine erfolgreiche Förderung (vgl. Machemer 1972, S. 336ff.).

Es ist relativ leicht zu testen, ob ein Kind Schwierigkeiten im Lesen und Schreiben hat, indem man überprüft, inwieweit es der Schriftsprache mächtig ist und diese Ergebnisse mit anderen Daten von Schülern der gleichen Altersgruppe vergleicht. Umso schwerer ist es jedoch die Leistung vorherzusagen, bevor ein Defizit aufgetreten ist und festzustellen, ob das Kind in der Zukunft Schwierigkeiten im Schriftspracherwerb bekommt.

Erste Hinweise liefern die Erkenntnis, dass sich schon zum Ende der zweiten Klasse eine stabile Lernleistungsrangreihe entwickelt hat, die die Schüler anhand ihres Erfolgs im Lesen und Schreiben klassifiziert (vgl. Klicpera & Gasteiger-Klicpera 1995, S. 95ff.). Es konnte auch in der längsschnittlichen Untersuchung zum „Modell Schriftsprach-Moderatoren" (MSM) von Deimel, Ziegler und Schulte-Körne (2005, S. 7) mit 69 Grundschulklassen gezeigt werden, dass Kinder, die durch das Bielefelder Screening als Risikokinder identifiziert worden sind, auch im unteren Leistungsbereich in den Rechtschreibleistungen in der Schule bis Ende Klasse 2 bleiben. Marx (1992a, S. 251) merkt ebenfalls an, dass sich bei älteren lese-rechtschreibschwachen Kindern weniger häufig Erfolge bei der Förderung des Lesens und Schreibens einstellen. Er verweist auf eine Untersuchung von Tramontana, Hooper und Selzer (1988, S. 89ff.), der zufolge jüngere Kinder besser auf Interventionsmaßnahmen ansprechen als ältere.

Korrelative Längsschnittstudien und experimentelle Trainingsstudien konnten zudem in Bezug auf die phonologische Informationsverarbeitung zeigen (vgl. i. F. Kap. II.1.2), dass diese im Vorschulalter eine kausale Beziehung zu den schulischen schriftsprachlichen Leistungen aufzeigt und die Fähigkeit der phonologischen Bewusstheit auch effektiv gefördert werden kann (vgl. bspw. Lundberg, Frost & Peterson 1988, S. 263ff.; Bus & Ijzendoorn 1999, S. 406ff.; Fröhlich, Metz & Petermann 2009, S. 204ff.).

Das Training der phonologischen Informationsverarbeitung als Vorläuferfertigkeit der Lese- und Rechtschreibleistungen in der Schule scheint möglich zu sein (vgl. auch Kap. II.1 zu möglichen Prädiktoren von Lese-Rechtschreibleistungen). Somit bedarf es zu präventiven Zwecken einer Frühdiagnostik, um frühzeitige Interventionen treffen zu können.

Im folgenden Abschnitt wird das Ausmaß möglicher Konsequenzen ohne frühzeitige Prognose und Förderung erörtert, um aufzuzeigen, inwieweit die präventive Forschung zur Vermeidung dieser Folgen in dem Fachgebiet sinnvoll ist.

1.3 Mögliche Folgen ohne frühzeitige Prognose

Die Relevanz der Forschung im Bereich der Prädiktion und Prävention von Lese-Rechtschreibschwäche wird insbesondere auch durch deren Verlauf, Stabilität, dem Ausmaß der Schwierigkeiten und den Zukunftsaussichten für lese-rechtschreibschwache Kinder aufgeklärt. Die heute praktizierte Diagnostik im Bereich Legasthenie oder Lese-Rechtschreibschwäche wird zumeist erst im zweiten bis vierten Schuljahr initiiert. Die Erkennung der Schwierigkeiten erfolgt somit oft zu spät. Es bedarf einer Diagnostik, die bereits im vorschulischen Bereich Schwierigkeiten aufdeckt, die auf Probleme im Schriftspracherwerb im späteren schulischen Verlauf hinweisen, die dann durch eine abgestimmte individuelle Förderung der Vorläuferfertigkeiten geschwächt oder gar eingestellt werden können.

Barth (1997, S. 8) zählt folgende Gründe für die Notwendigkeit einer Früherkennung auf:

- Lese-rechtschreibschwache Kinder haben einen mittel- und langfristig schlechteren Schulerfolg als durchschnittliche Kinder (nur 12,5 % besuchen das Gymnasium).
- Kinder mit Lese-Rechtschreibschwäche entwickeln häufig Verhaltensauffälligkeiten, meist im dissozialen Bereich (Schulschwänzen, Alkohol- oder Drogenmissbrauch, Zerstörung).
- Die Besserungsrate liegt bei höchstens 34 % der Kinder mit Lese-Rechtschreibschwäche. Bereits Ende der zweiten Klasse hat sich eine relativ stabile Leistungsrangfolge etabliert.

Im folgenden Abschnitt werden die Prävalenz, die Komorbidität sowie der Verlauf und die Stabilität der Lese-Rechtschreibschwäche betrachtet, um die Auswirkungen über das Ausmaß der Schwierigkeiten abwägen zu können, die eintreten können, wenn keine Fördermaßnahmen ergriffen werden.

1.3.1 Prävalenz und Komorbidität der Lese-Rechtschreibschwäche

Die Prävalenz und Komorbidität der Lese-Rechtschreibschwäche soll nun genauer ins Blickfeld genommen werden, da sie das Ausmaß der Problematik und weiterer möglicher Folgen einer Lese-Rechtschreibschwäche erklären, wenn keine präventiven Bemühungen vorgenommen werden.

Zunächst wird die Prävalenz der Lese-Rechtschreibschwäche beschrieben. Aufgrund uneinheitlicher Begrifflichkeiten und Inkonsistenzen bei Definitionen ergeben sich starke Unterschiede in den Prävalenzraten. Aus diesem Grund werden hier vorerst die verschiedenen Begriffe und Definitionen der Lese-Rechtschreibschwäche geklärt.

Die Terminologie der Lese-Rechtschreibschwäche wurde in den Untersuchungen in der Vergangenheit oft uneinheitlich genutzt. Es existieren sowohl eine Reihe diverser Begrifflichkeiten, die für die Beschreibung des gleichen Konstrukts genutzt werden, als auch eine Vielzahl u. a. von den Forschungsdisziplinen abhängige Definitionen mit differenten Merkmalen der Lese-Rechtschreibschwäche. Auch in der pädagogisch-psychologischen Forschung wurde der Begriff der Lese-Rechtschreibschwäche nicht einheitlich verwendet. Die Definitionen sind je nach angewandter Theorie sogar widersprüchlich zueinander (vgl. Weber 2003, S.91; S. 116f.). Der Neurologe und Psychiater P. Ranschburg formte (1916, S. 3) den Begriff „Legasthenie“ (griech. legein = lesen; asthenia = Schwäche), der heute oft synonym zur „umschriebenen Lese-Rechtschreibschwäche“ genutzt wird (vgl. Warnke 1996, S. 21ff.). Die Unterschiede zwischen einer „Lese-Rechtschreibschwäche“ und einer „Legasthenie“ sind bei einigen Autoren jedoch auch durch eine defizitäre intellektuelle Grundlage der betroffenen Kinder ausdifferenziert, was sich unter dem Begriff „Diskrepanzdefinition“ in der Wissenschaft etabliert hat.

Die Diskrepanzdefinition schlug sich ebenfalls in den älteren Versionen der internationalen Klassifikationsschemata (ICD-10[12] oder DSM-IV-TR[13]) nieder. Die Lese-Rechtschreibstörung (LRS) gehörte im internationalen Klassifikationsschema psychischer Störungen (ICD-10, vgl. Dilling, Mombour & Schmidt 1994, S. 257-258) zu den umschriebenen Entwicklungsstörungen schulischer Fertigkeiten (Störungsnummer F81.0 und F81.1). Die betroffenen Schüler sind nicht dazu in der Lage das Lesen und Schreiben, trotz ausreichender Unterrichtung und mindestens durchschnittlichen kognitiven Fähigkeiten, sowie adäquaten Hör- und Sehleistungen, zu erlernen:

> *„Das Hauptmerkmal dieser Störung ist eine umschriebene und eindeutige Beeinträchtigung in der Entwicklung der Lesefertigkeiten, die nicht allein durch das Entwicklungsalter, durch Visusprobleme oder unangemessene Beschulung erklärbar ist. Das Leseverständnis, die Fähigkeit, gelesene Worte wiederzugeben, vorzulesen und die Leistungen bei Aufgaben, für welche Lesefähigkeit benötigt werden, können*

12 ICD-10 ist die „Internationale Klassifikation psychischer Störungen.“ (vgl. WHO 2010, 2015).
13 DSM-IV ist das „Diagnostische[s] und Statistische[s] Handbuch Psychischer Störungen“ (vgl. Saß, Wittchen, Zaudig & Houben 2003).

> *sämtlich betroffen sein. Mit Lesestörungen gehen häufig Rechtschreibstörungen einher. Diese persistieren oft bis in die Adoleszenz, auch wenn im Lesen einige Fortschritte gemacht wurden […]. Die Leseleistungen des Kindes müssen unter dem Niveau liegen, das auf Grund des Alters, der altersgemeinen Intelligenz und der Beschulung zu erwarten ist […]. In der späteren Kindheit und im Erwachsenenalter sind die Rechtschreibprobleme meist größer als Defizite in der Lesefähigkeit." (ICD-10, Dilling, Mombour & Schmidt 1994, S. 257-258)*

Bei einer Lese-Rechtschreibschwäche sei die Lernfähigkeit im Allgemeinen nicht eingeschränkt. Wie in den älteren Fassungen der ICD-10 beschrieben zeichnete sich die Lese-Rechtschreibschwäche durch das doppelte Diskrepanzkriterium aus: Die Lese- und Schreibleistungen sind im Vergleich zu altersgleichen Mitschülern diskrepant und die Lese-Rechtschreibleistung ist im Niveau deutlich niedriger als das gemessene Intelligenzniveau. Die Diskrepanz wird meistens mit 1 bis 1,5 Standardabweichungen gemessen (vgl. Warnke, Hemminger & Plume 2004, S. 68f.).

Bereits in den 70er-Jahren haben Forscher befürwortet, das Diskrepanzkriterium aus den offiziellen Richtlinien der Bundesländer und auch aus dem internationalen Klassifikationsschema zu entfernen und stattdessen alle Schüler mit Schwächen im Lesen und Schreiben als lese-rechtschreibschwach zu beschreiben (vgl. Schlee 1976 in seinem Werk „Legasthenieforschung am Ende?").

Zwei zentrale Forschungsbefunde konnten in zahlreichen Studien (vgl. bspw. Tacke 2008, S. 152; Walter 2000, S. 381; Klicpera & Gasteiger-Klicpera 2001, S. 41ff.; Metz, Marx, Weber & Schneider 2003, S. 132, Schulte-Körne, Schäfer, Deimel & Remschmidt 1997, S. 51; Tacke, Nock & Staiber 1987, S. 45ff.; Zielinski & Schneider 1986, S. 38ff.; Siegel 1989, S. 469ff.) zeigen, dass sich die IQ-Diskrepanzdefinition als nicht haltbar erwies (vgl. Tacke 2008, S. 152):

1) Es gibt keine Ursachen der Lese-Rechtschreibschwäche, die die über- oder unterdurchschnittlich intelligenten Schüler voneinander unterscheiden.
2) Es gibt keine Unterschiede bei der Wirkung von Fördermaßnahmen zwischen intelligenten und weniger intelligenten Schülern.

Siegel (1989, S. 469) fasst dies auf der Grundlage ihrer Studien zusammen: „On logical and empirical grounds, IQ test scores are not necessary for the definition of learning disabilities."

Die fortgesetzte Resistenz des Diskrepanzkriteriums über viele Jahre in der Forschung, wie auch in der praktischen Anwendung lässt einen Mangel an Alternativen und an anderen entsprechend erforschten Theorien vermuten. Durch die Vielfalt an unterschiedlichen Definitionen scheint die Diskrepanz zwischen Lese-Rechtschreibleistungen und Intelligenz einen Teilbereich der Forschung darzustellen, der sich in der Diskussion der unterschiedlichen Theorien vermehrt widerspiegelt und dadurch eine Art „Scheinrelevanz" aufrecht erhält, die sich jedoch bei der genaueren Betrachtung der Studien als nicht gerechtfertigt erweist (vgl. Weber 2003, S. 117). In der aktuellen Fassung der DSM-V (2015) sowie in der aktuellen ICD-10 (2016) wurde das IQ-Diskrepanzkriterium aufgegeben. Außerdem ergänzt die ICD-10, dass Störungen des Sprechens und der Sprache der Lese- Rechtschreibstörung voraus gehen:

> *„Das Hauptmerkmal ist eine umschriebene und bedeutsame Beeinträchtigung in der Entwicklung der Lesefertigkeiten, die nicht allein durch das Entwicklungsalter, Visusprobleme oder unangemessene Beschulung erklärbar ist. Das Leseverständnis, die Fähigkeit, gelesene Worte wieder zu erkennen, vorzulesen und Leistungen, für welche Lesefähigkeit nötig ist, können sämtlich betroffen sein. Bei umschriebenen Lesestörungen sind Rechtschreibstörungen häufig und persistieren oft bis in die Adoleszenz, auch wenn einige Fortschritte im Lesen gemacht werden. Umschriebenen Entwicklungsstörungen des Lesens gehen Entwicklungsstörungen des Sprechens oder der Sprache voraus. Während der Schulzeit sind begleitende Störungen im emotionalen und Verhaltensbereich häufig." (vgl. WHO in dimdi 2015)*

Dies beschreibt eine Annäherung der medizinischen zu den pädagogisch-psychologischen Ansichten. In der vorliegenden Arbeit wird dennoch vorwiegend der Begriff „Lese-Rechtschreibschwäche" verwendet, auf den Terminus „Legasthenie" wird aufgrund der möglichen Assoziation zur Diskrepanzdefinition verzichtet.

Die Angaben zur Prävalenz der Lese-Rechtschreibschwäche sind in diversen epidemiologischen Untersuchungen schwankend. Die unterschiedlichen Definitionen einer Lese-Rechtschreibschwäche, sowie Normierungsprobleme in Testverfahren zum Lesen und Schreiben, der Einsatz unterschiedlicher Diagnostikverfahren und unterschiedliche Grenzwerte für die Schwere einer Lese- und Rechtschreibschwäche beeinflussen die Untersuchungsergebnisse. Eine genaue Messung der Prävalenzrate ist dadurch nicht möglich.

Marx (1992a, S. 252f.) führt eine hohe Prävalenzrate zwischen 5 und 15 % an (die allerdings nur als relativ zu betrachten ist), die er als substantiell bezeichnet und aufgrund derer die Notwendigkeit einer Förderung auf der Hand läge. Rispens und van Yperen (1990, zit. in Weber 2003, S. 117) führten eine Untersuchung mit fünf unterschiedlichen Diagnostikmethoden durch und verglichen den Anteil der jeweils erfassten Kinder mit Lese-Rechtschreibschwäche. Je nach Diagnostikverfahren schwankte der Anteil der legasthenen Kinder zwischen 4 % und 18,3 %. Internationale Schätzungen liegen bei 2 bis 4 % lese-rechtschreibschwacher Schüler in den Industrienationen (vgl. Gasteiger-Klicpera & Klicpera 2004, S. 47). Nach Warnke, Hemminger und Plume (2004, S. 4f.) entwickeln ungefähr 4 bis 8 % aller Kinder eine Lese-Rechtschreibstörung. Warnke und Roth (2000, S. 458) machen eine Prävalenzrate von 2,7 % im deutschen Sprachraum für Kinder im Alter von acht Jahren aus. Von der Diskrepanzdefinition der Lese-Rechtschreibschwäche ausgehend konnte Schulte-Körne eine Prävalenzrate von 3 bis 5 % feststellen. Dies sind für das Grundschulalter in Deutschland knapp 200 000 Schüler (vgl. Schulte-Körne 2004, S. 65). Strehlow und Haffner (2002, S. 122) berechneten die Prävalenzrate nach verschiedenen Forschungskriterien. Die empirischen Daten reichen von einer Prävalenzrate von 2 % (ICD-10-Kriterium) bis zu einer von über 10 % nach den unterschiedlichen klinischen Definitionen und Diagnostikinstrumenten. Thomas, Schulte-Körne und Hasselhorn (2015, S. 437) fassen zusammen, dass die Prävalenzrate altersabhängig ist. Im Alter von acht Jahren wären 7 bis 8 % der Kinder betroffen, im Alter von zwölf Jahren nur noch 6 %.

Jungen werden drei- bis viermal häufiger diagnostiziert, es gibt jedoch Hinweise darauf, dass Mädchen die Lese-Rechtschreibstörung schneller und besser überwinden und sich die Rate deshalb wieder ausgleicht (vgl. Warnke, Hemminger & Plume 2004, S. 5). Auch Vellutino (1979) konnte von Verhältnissen von 4:1 bis hin zu 8:1 zu Lasten der Jungen berichten, dies konnte aber nicht in allen Studien bestätigt werden. Oft ließen sich keine signifikanten Unterschiede aufweisen (vgl. Küspert 1998, S. 112). Mannhaupt (1994, zit. in Schneider, Stefanek & Dotzler 1997, S. 115) analysierte die Daten der Bielefelder Längsschnittstudie und machte einen größeren Anteil an Jungen in der Stichprobe aus, die lese-rechtschreibschwach waren (ca. 70 %). Thomas, Schulte-Körne und Hasselhorn (2015, S. 437) beschreiben, dass der erhöhte Anteil der Jungen eher im Bereich des Rechtschreibens auftreten würde.

Die Lese-Rechtschreibschwäche verursacht meist auch Schwierigkeiten in anderen Leistungsbereichen, so dass sich häufig andere komorbide Problematiken daraus ergeben oder mit der Störung einhergehen. Verschiedene Autoren bestätigen, dass sich eine Lese-Rechtschreibstörung vielfältig auf diverse Bereiche des Lebens auswirkt (vgl. Warnke,

Hemminger & Plume 2004, S. 23fff.; Barth & Gomm 2008b, S. 9; Gasteiger-Klicpera & Klicpera 2004, S. 47f.).

Zusätzlich zum Schulversagen gehen auch noch andere komorbide Entwicklungsstörungen mit der Lese-Rechtschreib-Schwäche einher. Im Vorschulalter können folgende weitere Auffälligkeiten festgestellt werden (zusammenfassend in Warnke, Hemminger & Plume 2004, S. 24ff.; Schulte-Körne 2004, S. 66f.):

- Aufmerksamkeits- oder Hyperaktivitätsstörung
- Entwicklungsstörungen der Sprache
- Visuo-motorische Beeinträchtigung
- Schwächen in der Figur-Grund-Wahrnehmung, Raum-Lage-Wahrnehmung
- Störungen im sozialen Verhalten

Im Schulalter können weitere Auffälligkeiten hinzukommen:

- Hyperkinetische Störung
- Emotionale Störung
- Störung des Sozialverhaltens
- Psychische Probleme (Versagensängste, motorische Unruhe, Motivationsverlust)
- Beeinträchtigung des schulischen Fortkommens (z. B. Klassenwiederholung, oder niederen Schulabschluss im Vergleich zu gleichintelligenten Mitschülern)
- Hausaufgabenkonflikte
- Motorische Unruhe
- Konzentrationsstörungen
- Depressive und ängstliche Anpassungsstörungen

Im Beruf sind folgende Konsequenzen zu einer erhöhten Wahrscheinlichkeit möglich:

- Höhere Arbeitslosigkeit
- Straffälligkeit (ggf. mit Gefängnisaufenthalt)

In der klinischen Studie von Georgiewa, Grünling, Ligges, Filz, Möller und Blanz (2004, S. 285f.) wurde die Leseleistung von Kindern, Jugendlichen und Erwachsenen mit und ohne Lese-Rechtschreibschwäche verglichen. Zwei wesentliche Befunde konnten sie vermerken: Personen mit Lese-Rechtschreibschwäche machen mehr Fehler und benötigen mehr Zeit zum Lesen. Außerdem erreichen Erwachsene mit Lese-Rechtschreibschwäche beim Wortlesen (Fehler und Zeit) nicht das Niveau von normallesenden Jugendlichen und beim

Pseudowortlesen nicht das Niveau von normallesenden Kindern. Die Autoren vermuten ein Defizit im sublexikalischen Bereich, da die Aufgaben mit den häufigsten Fehlern inhaltlich auf den Graphem-Phonem-Korrespondenzen beruhen. Es wird vermutet, dass mit der Lese-Rechtschreibschwäche komorbid auch kognitive Schwierigkeiten einhergehen (vgl. dazu Kap. II.2.1.2.3 zur Intelligenz, S. 182).

Nach einer Metaanalyse von McArthur, Hoghen, Edward, Heath und Mengler (2000, S. 869ff.) entwickeln 40 – 50 % der sprachentwicklungs- oder aussprachegestörten Kinder eine Lese-Rechtschreibschwäche (vgl. dazu Kap. II.1.4 zur Sprache, S. 156). Zusätzlich ist eine schwierige sozioökonomische Situation im Elternhaus ein Risikofaktor für die Entwicklung von Lese-Rechtschreibleistungen. In vielen Studien wurde auch ein Zusammenhang mit Rechenstörungen festgestellt (vgl. Barth & Gomm 2008a, S. 4). Schwenck und Schneider (2003, S. 212ff.) konnten sogar einen Anteil von 20 – 60 % der betroffenen Kinder feststellen, die neben einer Lese-Rechtschreibschwäche auch eine Rechenschwäche aufwiesen. Schwenck und Schneider (a. a. O., S. 216ff.) konnten ebenfalls die Existenz einer Gruppe von Kindern bestätigen, die sowohl Schwierigkeiten in den schriftsprachlichen als auch in den mathematischen Fähigkeiten aufweisen. Es existierte jedoch gleichermaßen auch eine Gruppe von Kindern, die nur Schwierigkeiten im Lesen und Schreiben oder nur im Rechnen aufwies.

Im Anschluss wird der Verlauf und die Stabilität der Lese-Rechtschreibschwäche beschrieben, um auf die Notwendigkeit einer frühen Identifikation von Schwierigkeiten im Schriftspracherwerb hinzuweisen.

1.3.2 Verlauf und Stabilität der Lese-Rechtschreibschwäche

Die individuelle Entwicklung des einzelnen Kindes mit Lese-Rechtschreibschwäche ist sehr unterschiedlich. Die Lese-Rechtschreibschwäche ist meist ab dem zweiten Schuljahr eine über mehrere Schuljahre hinweg stabil messbare Größe (vgl. Strehlow 1998, S. 31). Es konnte nachgewiesen werden, dass rechtschreibschwache Schüler, die in der zweiten Jahrgangsstufe zu den schwächsten Schülern gehörten, dies auch in der achten Jahrgangsstufe waren. Von den sehr guten Rechtschreibern blieben 75 % in der gleichen Kategorie (vgl. Schneider 2008, S. 148f.). Ein Schüler am Ende der 1. Klasse mit Lese-Rechtschreibschwäche wird ohne Förderung über die gesamte Schulzeit hinweg Schwierigkeiten haben (vgl. Gasteiger-Klicpera & Klicpera 2004, S. 49). In der von Schneider (2008, S. 148f.) erwähnten LOGIK-Studie (Longitudinalstudie zur Genese individueller

Kompetenzen) ergaben sich ähnliche Ergebnisse mit Kennwerten, die ab der dritten Jahrgangsstufe über r = .7 lagen und sogar die individuellen Stabilitäten der Intelligenz leicht übertrafen. Dies konnten auch Esser und Schmidt (1993, S. 100ff.) in ihren Untersuchungen bestätigen, in denen sich nur ein Drittel der Schüler zwischen dem achten und dreizehnten Lebensjahr signifikant in ihren Rechtschreibleistungen verbessern konnten.

Auch die PISA- und die IGLU-Studie (Internationale Grundschul-Lese-Untersuchung) konnten bedeutend dokumentieren, dass sich die Stabilität der Lese-Rechtschreibschwäche als konstant erwies:

> *„Das auf der Ebene der Grundschule erreichte Leistungsniveau hat wesentlichen Einfluss auf den weiteren Verlauf von Bildungskarrieren. Insbesondere Schülerinnen und Schüler, die im Rahmen von IGLU zur unteren Leistungsgruppe gehören, werden den Anschluss an das Lernen der Sekundarstufe 1 nur mit Mühe finden […]“ (Bos, Lankes, Prenzel, Schwippert, Valtin & Walther 2003, S. 38).*

Anfänglich sind die Schwierigkeiten noch gering, die Leistungsrückstände der Schüler nehmen dann meist massiv zu und sind ohne Förderung kaum mehr aufzuholen. Die Schwierigkeiten bleiben stabil und führen häufig zu Aversionen gegenüber schriftsprachlichen Leistungen. Die anfänglichen Schwierigkeiten in der Schriftsprache finden sich häufig im gesamten Schulverlauf wieder. Dieser Effekt wird auch als „Matthäuseffekt“ bezeichnet (vgl. Stanovich 1986, S. 360ff.). Gute Leser haben Spaß am Lesen und bauen ihre Fähigkeiten geschwind aus, während schwache Leser und Schreiber sich umso weniger mit der Schriftsprache befassen und damit ihren Rückstand noch vergrößern. Leistungsunterschiede zwischen den Schülern mit Lese-Rechtschreibschwäche und den anderen Schülern bleiben meistens konstant (vgl. Strehlow 1994, S. 219) bzw. nehmen noch zu. Schneider und Stefanek (2007, S. 79f.) stellten aus Befunden der LOGIK-Studie im Zeitraum von 1984 bis 2004 fest, dass sich eine hohe Zeitstabilität einstellt und die individuellen Unterschiede im Rechtschreiben erhalten bleiben. Weibliche Teilnehmer, sowie Teilnehmer mit einem höheren Bildungsniveau konnten Vorteile bei der Entwicklung verzeichnen. Es bleibt anzumerken, dass die Teilnehmer im Jahr 2004 sich von ursprünglich 204 Kindergartenkindern auf 152 Erwachsene im Jahr 2004 reduzierten, unter denen Personen mit höherem Bildungsabschluss überrepräsentiert waren. Strehlow, Kluge, Müller und Haffner (1992, in Hasselhorn, Schneider & Marx 2000, S. 3) konnten in einer Längsschnittstudie zeigen, dass sich die Leistungsrückstände von Kindern mit Lese-Rechtschreibschwäche nach zwölf Jahren nochmals vergrößert hatten. Barth und Gomm (2008b, S. 8f.) fassen zusammen, dass die schulischen und beruflichen Perspektiven der Schüler mit

Lese- und Rechtschreibschwierigkeiten eingeschränkt sind. Sie sind langfristig ihren Mitschülern gegenüber benachteiligt und erreichen signifikant seltener als Mitschüler mit gleicher Intelligenz einen höheren Schulabschluss (vgl. a. a. O., S. 9). Esser, Wyschkon und Schmidt (2002, S. 238ff.) konnten sogar feststellen, dass lese-rechtschreibschwache Personen im beruflichen Verlauf zu einem deutlich erhöhten Anteil Arbeitslosigkeit zu verzeichnen haben.

Strehlow und Haffner (2002, S. 124) analysieren diverse Studien und bestätigen durch die eigene Untersuchung, dass die Lese-Rechtschreibschwäche kein lineares Zusammenhangsgefüge darstellt, sondern dass es sich um Wechselwirkungsprozesse handelt. Einerseits hätten Schüler mit einer Lese-Rechtschreibschwäche zum Ende der Grundschulzeit schlechtere Chancen an ein Gymnasium oder eine Realschule überwiesen zu werden, andererseits verbessern sich ihre Lese-Rechtschreibleistungen an diesen Schulen signifikant, wenn sie diese dann doch besuchen sollten. Dies konnten auch Hofmann und Sasse (2006a, S. 96) bestätigen. Sie untersuchten Erwachsene, bei denen als Kind eine Lese-Rechtschreibschwäche identifiziert worden ist und die im Raum Bochum/Dortmund an einem Förderprogramm teilgenommen hatten. 89 % der 187 Teilnehmer konnten bis zu 14 Jahre nach Abschluss des Förderprogramms einen höheren Bildungsabschluss vorweisen (Realschulabschluss, Fachhochschul- oder Hochschulreife). Im Bundesdurchschnitt sind es dagegen nur 37 % der Personen mit Lese-Rechtschreibschwäche.

Außerdem konnten Klicpera und Gasteiger-Klicpera (1993) feststellen, dass der Abstand der lese-rechtschreibschwachen Kinder zu anderen Kindern in der gleichen Altersstufe sehr groß ist. 15 % der schwächsten Leser der vierten Klasse wiesen die Fähigkeiten von durchschnittlichen Kindern am Ende der ersten Klasse auf. In der achten Klasse der Hauptschule liegen die schwächsten Leser sogar bis zu sechs Jahre in ihrer Leseleistung zurück. Viele verlassen die Schule dann als totale oder als funktionale Analphabeten. Die Schätzung der Analphabeten im Jahr 2011 liegt bei ca. 3,9 %, also bei ca. 2 Mio. Menschen in Deutschland, für funktionale Analphabeten sogar bei 14,5 % der Erwachsenen mit ca. 7,5 Mio. Menschen (vgl. Grotlüschen & Riekmann 2011, S. 4).

Es zeigt sich anhand der Ergebnisse, wie wichtig eine frühzeitige Förderung der Kinder ist, bevor sich die Lese-Rechtschreibschwäche auf andere Bereiche des Lebens auswirkt. Durch frühzeitige Diagnostik kann die Förderung beginnen, bevor sich eine Sekundärsymptomatik ausgebreitet hat. Je früher eine Intervention erfolgt, desto größer ist die Wahrscheinlichkeit, dass sich die Schwierigkeiten nicht schon negativ auf das Kind ausgewirkt haben. Die sehr hohen Prävalenzraten und auch die Stabilität der Lese-Rechtschreibschwäche veranlassen

dazu, sich mit Prognose, Prävention und Förderung zu beschäftigen. Die Relevanz und die Brisanz des Themas werden durch die schlechten Zukunftsperspektiven für lese-rechtschreibschwache Schüler noch verschärft. Die Notwendigkeit der Implementierung von präventiven spezifischen Maßnahmen in die Praxis zur Reduzierung der Kinder mit Lese-Rechtschreibschwäche liegt auf der Hand. Die Prävention und die damit verbundene Früherkennung von Kindern, die Gefahr laufen, in der Grundschule eine Lese-Rechtschreib-schwäche zu entwickeln, stehen dabei im Vordergrund. Insbesondere ist es die Zielsetzung der aktuellen Forschung, Verfahren zu finden, die zum einen förderbedürftige Kinder erkennen, zum anderen aber auch Hinweise für die Förderung geben und diese Interventionsmaßnahme ggf. auch begleiten bzw. deren Effektivität beweisen können.

Im dargestellten Abschnitt konnte anhand der immensen strukturellen Veränderungen im Bildungssystem insbesondere in Bezug auf die Inklusion und die zentrale Rolle der Diagnostik in den systemveränderten Bildungsstätten dargestellt werden, dass die Not-wendigkeit einer Früherkennung und damit einhergehenden präventiven Maßnahmen essentiell ist. Durch die Persistenz, die Stabilität und komorbid einhergehende schwere Folgeproblematiken einer Lese-Rechtschreibschwäche im Sinne einschneidender, lebens-verändernder und gravierender Konsequenzen, wird die Notwendigkeit der weiteren Forschung im Bereich der Prognose von Schwierigkeiten im Lesen und Rechtschreiben bereits im Vorschulalter deutlich.

Im deutschsprachigen Raum wurde von Jansen, Mannhaupt, Marx und Skowronek (1999) auf der Grundlage skandinavischer Forschungen (vgl. Lundberg, Frost & Peterson 1988, S. 263ff.) ein Instrument zur Prädiktion von Lese-Rechtschreibschwierigkeiten entwickelt, an dessen Beispiel im späteren Verlauf weitere Verfahren geschaffen wurden (vgl. bspw. Martschinke, Kirschhock & Frank, 2001; Barth & Gomm 2005). Fünf der derzeit zentral diskutierten Früherkennungsverfahren werden im folgenden Abschnitt vorgestellt, diskutiert und bezüglich ihrer Relevanz, wissenschaftlichen Fundierung und Vorhersagegenauigkeit beurteilt.

2 Testinstrumentarien zur Prognose schriftsprachlicher Fähigkeiten im Vorschulalter

Eine Reihe von Längsschnittstudien (vgl. bspw. LOGIK – Longitudinalstudie zur Genese individueller Kompetenzen von Helmke und Schrader ab 1984, oder auch die Studien zum Bielefelder Screening von Jansen, Mannhaupt, Marx und Skowronek 1999) konnte in den

vergangenen Jahren die Verschiebung der Forschungsbemühungen auch auf den Bereich der Frühdiagnostik bewirken. Der Grundgedanke der Frühdiagnostik fokussiert sich dabei auf die Erkennung von Schwierigkeiten vor Beginn des formellen Schriftspracherwerbs. Aufgrund der heute verfügbaren Testverfahren zur Früherkennung von Legasthenie hat sich das Interesse der Forschung weitgehend verschoben. Die Frühdiagnostik ist als Möglichkeit der Prognose und Prävention von Lese-Rechtschreibschwäche in den Mittelpunkt des wissenschaftlichen Interesses gerückt. Aufgrund der bereits oben beschriebenen Problematik der schlechten Prädiktionsleistung des Bielefelder Screenings werden im Folgenden zunächst in einem kurzen Exkurs die geforderten Hauptgütekriterien der klassischen Testtheorie insbesondere in Bezug auf Screening-Verfahren vorgestellt. Im Anschluss werden dann das Bielefelder Screening sowie vier weitere Verfahren zur Prognose von Lese- und Rechtschreibleistungen präsentiert, sowie auf der Grundlage ihrer Gütekriterien bewertet und anhand aktueller Studien beleuchtet.

2.1 EXKURS: Gütekriterien für Screening-Verfahren

Neben den allgemeinen Gütekriterien für Testverfahren gibt es spezifische Gütekriterien explizit für die Vorhersagegenauigkeit von Screening-Verfahren, außerdem vom Zufall bereinigte Gütekriterien, die insbesondere zum Vergleich unterschiedlicher Screening-Verfahren dienlich sind. Die allgemeinen und auch die speziellen Gütekriterien werden im folgenden Abschnitt vorgestellt, um die darauf folgenden vorgestellten Screening-Verfahren wissenschaftlich bewerten und vergleichen zu können. Dieser Abschnitt ist für den Leser als Exkurs gekennzeichnet, da die klassischen Testgütekriterien der Objektivität, Reliabilität und Validität in der Fachliteratur hinreichend bekannt sind. Auf die Besonderheit der zusätzlichen Indizes zur Auswertung der klassifikatorischen Validität sei an dieser Stelle hingewiesen (vgl. Abschnitt 2.1.3.2).

Ein diagnostisches Verfahren kann unterschiedliche Informationen enthalten und auch von unterschiedlicher Qualität sein. Die methodischen Gütekriterien definieren die Qualität der Messung. Hierbei werden insbesondere die Genauigkeit und die Verlässlichkeit der Ergebnisse betrachtet.
In der pädagogisch-psychologischen Diagnostik (vgl. Ingenkamp & Lissmann 2008, S. 51ff.) werden die wichtigsten Gütekriterien wie folgt beschrieben (vgl. Abb. 3):

- Objektivität
- Reliabilität
- Validität

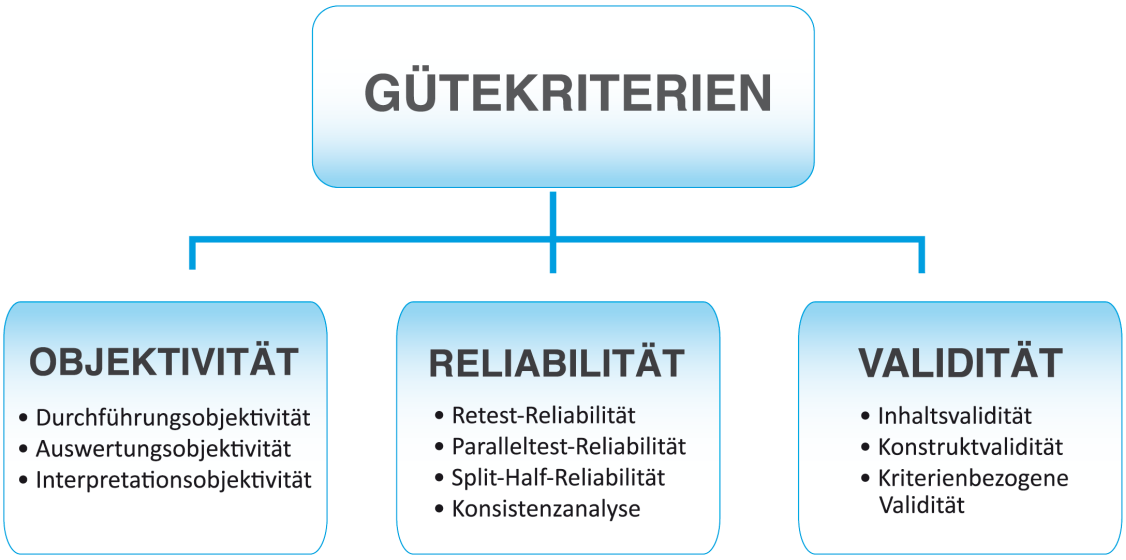


Abb. 3: Gütekriterien: Objektivität, Reliabilität und Validität

2.1.1 Objektivität

Die Objektivität beschreibt die Unabhängigkeit der Messergebnisse vom Testleiter, dessen intersubjektive Einflüsse möglichst unter Kontrolle gehalten werden sollten:

> *„Unter Objektivität eines Tests verstehen wir den Grad, in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind. Ein Test wäre demnach vollkommen objektiv, wenn verschiedenen Untersucher bei denselben Probanden zu gleichen Ergebnissen gelangten. Man spricht deshalb auch von 'interpersoneller Übereinstimmung' der Untersucher.“ (Lienert 1967, S. 13)*

Die Objektivität wird meistens in drei Bereiche eingeteilt: Durchführungs-, Auswertungs- und Interpretationsobjektivität.

Die *Durchführungsobjektivität* stellt sicher, dass alle Lernenden in der Diagnostiksituation den gleichen Anforderungen unterliegen. Dies wird zumeist durch die Vereinheitlichung von Aufgabenstellung, Instruktion, Bearbeitungszeit oder zulässigen Utensilien sichergestellt. Bei Screening-Verfahren zur Früherkennung von Lese-Rechtschreibschwäche im Vorschulbereich wird die Durchführungsobjektivität meist über die Vorgabe der Items per CD gewährleistet.

Die *Auswertungsobjektivität* beschreibt die Bewertung des Verhaltens des Lernenden. Die Auswertungsobjektivität wird oft durch schriftliche Aufgaben gewährleistet, die mit einem Punktesystem für alle Probanden gleich bewertet werden.

Durch eine gute *Interpretationsobjektivität* können unterschiedliche Personen die Ergebnisse einer Messung gleich interpretieren. Dies wird meistens durch gezielte Hinweise zur Interpretation im Handbuch oder durch die Einteilung in Normwerte garantiert.

Die Objektivität kann als Voraussetzung für die Reliabilität und Validität bewertet werden. Eine Messung, die der Objektivität nicht genügt, wird somit auch der Reliabilität und Validität nicht genügen (vgl. Ingenkamp & Lissmann 2008, S. 53f.; Wild & Krapp 2001, S. 524f.).

2.1.2 Reliabilität

Die Reliabilität beschreibt die Zuverlässigkeit, also den Grad der Genauigkeit, mit der ein Merkmal gemessen werden kann. Den Grad der Genauigkeit beschreibt der Zuverlässigkeitskoeffizient, der angibt, in welchem Maße das Testergebnis beim gleichen Probanden reproduzierbar ist. Die Reliabilität kann nicht aus einer Einzelmessung bestimmt werden, sondern wird anhand von repräsentativen Untersuchungen statistisch geschätzt (vgl. Wild & Krapp 2001, S. 524). In der Testkonstruktion werden zur Überprüfung dieses Kriteriums meistens die Retestmethode, die Paralleltestmethode, die Split-Half-Methode und die Konsistenzanalyse angewendet (vgl. Abb. 4).

Bei der *Retestmethode* werden die gleichen Aufgaben zu verschiedenen Zeitpunkten von den gleichen Personen bearbeitet. Die Korrelation der Ergebnisse wird im Zuverlässigkeitskoeffizienten ausgedrückt, der die zeitliche Stabilität eines Merkmals beschreibt. Aufgrund der Wiederholungsmessung kann es bei dieser Methode jedoch zu Übungseffekten kommen.



Abb. 4: Überprüfungsmöglichkeiten der Reliabilität (nach Lukesch 1998, S. 45)

Bei der *Paralleltestmethode* werden zwei Parallelformen des Tests verglichen, dessen Korrelationen wieder anhand eines Zuverlässigkeitskoeffizienten dargestellt werden können. Die Übungseffekte könnten durch die *Split-Half-Methode* kontrolliert werden, indem die Testergebnisse halbiert werden und die Korrelation der beiden Testhälften berechnet wird. Bei der *Konsistenzanalyse* unterteilt man den Test in alle seine Aufgaben und berechnet die Korrelationen, so wird die Konsistenz, also die Homogenität der Testitems, des Tests bestimmt (vgl. Lukesch 1998, S. 51).

Die Reliabilität ist Voraussetzung für die Validität (vgl. Ingenkamp & Lissmann 2008, S. 56). Der Zuverlässigkeitskoeffizient sollte nach Diekmann (2010, S. 253) möglichst nicht geringer als r = .80 ausfallen.

2.1.3 Validität

Die Validität beschreibt die Gültigkeit des Testverfahrens. Sie beschreibt, inwieweit das Kriterium gemessen wird, das gemessen werden soll. Die Validität ist das zentrale Gütekriterium für die Qualität der Messung (vgl. Wild & Krapp 2001, S. 525). Es wird überprüft, ob die Messung das richtige Merkmal erfassen konnte.

In der Testmethodik werden drei verschiedene Arten beschrieben: die Inhaltsvalidität, die Konstruktvalidität und die kriterienorientierte Validität (vgl. Lienert & Raatz 1998, S. 10; Lukesch 1998, S. 55ff.). Im Anschluss wird die prognostische Validität als zentrales Gütekriterium für Screening-Verfahren, die eine Vorhersage für bestimmte Leistungen treffen und deren Kennwerte separat beschrieben.

2.1.3.1 Inhaltliche Validität, Konstruktvalidität und kriterienorientierte Validität

Die *Inhaltsvalidität* zeigt, dass der Test die zu messenden Persönlichkeitsmerkmale optimal darstellt. Die inhaltliche Validität wird meistens durch Expertenurteile sichergestellt. Sie kann durch eine inhaltliche Analyse des Messverfahrens geprüft werden. Sie wird durch bisher vorhandene wissenschaftliche Erkenntnisse belegt und die Items werden mit dem aktuellen Stand der Forschung verglichen.

Die *Konstruktvalidität* zeigt auf, dass hinter dem Test eine theoretische Konstruktion vorhanden ist, durch die die Gültigkeit des Verfahrens bestimmt werden kann. Konstrukte sind „[...] abgeleitete, nicht unmittelbar fassbare, latente, komplexe Merkmale, die als relativ überdauernd angesehen werden und die unser Verhalten beeinflussen." (Ingenkamp & Lissmann 2008, S. 59). Für die Konstruktvalidität werden meist die Korrelationen zwischen den Kriterien eines Konstrukts auf Angemessenheit gemäß der Theorie überprüft. Die Bedeutung der Konstruktvalidierung „[...] liegt daher nicht in einer unmittelbaren praktisch-diagnostischen Verwertbarkeit, sondern in der theoretischen Klärung dessen, was der betreffende Test misst." (Lienert & Raatz 1998, S. 226).

Die *kriterienorientierte Validität* (auch empirische Validität) kontrolliert die Übereinstimmung mit einem Außenkriterium, welches das im Test zu untersuchende Merkmal abbildet. Dies ist beispielsweise durch einen vorhandenen standardisierten Test möglich. Die kriterienbezogene Validität wird empirisch geprüft. Dabei existieren so viele Arten von Validitäten wie angemessene Kriterien gefunden werden (vgl. Lienert & Raatz 1998, S. 10f.). Lienert (1969, S. 13) unterteilt die kriterienbezogene Validität in zwei Aspekte:

(a) die innere bzw. äußere kriterienbezogene Validität
(b) die Übereinstimmungsvalidität bzw. prognostische Validität

Bei einer Korrelation zwischen äußerem Kriterium mit einem Testpunktwert handelt es sich um eine *äußere Validität.* Bei einer Korrelation von Testpunktwert mit dem zu überprüfenden Merkmal wird von *innerer Validität* gesprochen.

Bei der *Übereinstimmungsvalidität* wird gemessen, inwieweit die Daten mit Angaben übereinstimmen, die auf andere Weise beschafft worden sind, beispielsweise mit einem anderen Testverfahren. Die Kriteriumswerte werden mit den Testwerten gleichzeitig erfasst. Es wird unabhängig von der Zeit von den Testleistungen auf Verhalten außerhalb des Tests übertragen.

Im Folgenden wird die *prognostische Validität*, die bei den Früherkennungsverfahren zur Prognose einer Lese-Rechtschreibschwäche in die Kritik geraten ist, gesondert hervorgehoben und eingehend bezüglich zu berechnender Kennwerte beschrieben.

2.1.3.2 Prognostische Validität

Die Vorhersagevalidität (oder auch prognostische Validität) wird gemessen, indem Korrelationen zwischen den Daten und den Daten eines vorherigen Messzeitpunkts verglichen werden. Die tatsächlich eingetretenen Werte werden mit den vorhergesagten Werten verglichen. Das zu erhebende Merkmal wird also mit Daten in der Zukunft in Bezug gesetzt. Die prognostische Validität ist ein wichtiges Gütekriterium für selektive Maßnahmen. Sie wird durch einen Korrelationskoeffizienten dargestellt, der den Zusammenhang zwischen Kriteriumsmerkmal und der diagnostischen Eignung beschreibt. Je näher dieser Koeffizient an 1 herankommt, desto niedriger ist auch die Irrtumswahrscheinlichkeit der Prognose (vgl. Wild & Krapp 2001, S. 527). Die prognostische Validität steht bei Screening-Verfahren im Mittelpunkt der Betrachtung. In der Forschung wird zwischen Prädiktoren, also den Merkmalen eines Kindes, mit denen zukünftige Leistungen vorhergesagt werden und Kriterien, also den Leistungen, die durch den Prädiktor vorhergesagt werden, unterschieden. Bei den Prädiktoren wird in Risiko- und Nichtrisikokindern unterschieden. Bei den Kriterien wird überprüft, ob ein Problem (z. B. Lese-Rechtschreibschwäche) eingetreten ist, oder nicht (vgl. Marx, Jansen & Skowronek 2000, S. 11). Identifizierte Risikokinder im Vorschulalter sollten mit hoher Wahrscheinlichkeit (ohne Förderung) in der Schule eine Lese-Rechtschreibproblematik entwickeln, während sogenannte Nicht-Risikokinder wiederum mit einer hohen Wahrscheinlichkeit im späteren Schulverlauf keine Schwierigkeiten mit dem Lesen und Rechtschreiben haben sollten (vgl. Marx und Lenhard 2010, S. 9).

Mit welcher Sicherheit eine Störung mit einem Screening erkannt wird, ist von den Güteeigenschaften abhängig. Die Typen der Richtig- bzw. Falsch-Klassifikationen der Kinder, die durch ein Messverfahren zustande kommt, werden meist durch ein Vierfelderschema verdeutlicht (vgl. Tab. 1).

- *Valid positiv* klassifizierten Kindern wurde durch den Prädiktor ein Risiko vorhergesagt, das sich auch durch das Kriterium bestätigte.
- *Falsch positiv* klassifizierten Kindern wurde ein Risiko durch das Messinstrument vorhergesagt, dieses konnte aber im weiteren Verlauf durch das Kriterium nicht bestätigt werden (Fehlerklassifikation erster Art).
- *Falsch negativ* klassifizierte Kinder wurden fälschlicherweise als unauffällig vorhergesagt, das Kriterium konnte später ein Problem feststellen (Fehlerklassifikation zweiter Art).
- *Valid negativ* klassifizierte Kinder wurden vom Prädiktor als unauffällig vorhergesagt, das sich durch das Kriterium bestätigte.

Tab. 1: Vierfelderschema zur Klassifikation (Marx 1992a, S. 255)

		Kriterium	
		Problem	Keine Probleme
Prädiktion	Risiko	Valid positiv (a)	Falsch positiv (b)
	Kein Risiko	Falsch negativ (c)	Valid negativ (d)

Um die Prognoseleistung des Diagnostikinstruments interpretieren zu können, werden bei standardisierten Verfahren üblicherweise die Kennwerte für die Güte der Vorhersagegenauigkeit berechnet, die nun vorgestellt werden. Die Vorhersagegenauigkeit wird durch die in Tab. 2 dargestellten Güteindizes bewertet.

Sensitivität und *Spezifität* zeigen die Vorhersagewahrscheinlichkeit bei den Personen auf, die die Störung aufweisen (Sensitivität) oder nicht aufweisen (Spezifität).

Die *positive und negative Korrektheit* beschreiben, wie hoch die Wahrscheinlichkeit ist, dass ein positiver oder negativer Screening-Befund auftritt.

Die *Gesamttrefferquote* dagegen zeigt den Anteil der richtigen Vorhersagen an der Gesamtheit auf, ist jedoch stark von der Grundquote abhängig.

Die *Aufdeckungsrate* ist zur Charakterisierung der Testvalidität gegeben und ebenfalls stark von der Prävalenz der Störung abhängig. Sie gibt den Anteil der durch das Screening entdeckten Störungen an. Je höher die Prävalenzrate ist, umso höher ist auch der Anteil der entdeckten Störungen. Ist die Störung jedoch nur selten vertreten, so ist auch der Anteil der Aufdeckungsrate geringer (vgl. Tröster 2009, S. 93). Die oberste Grenze der Aufdeckungsrate ist die Grundquote. Sie berücksichtigt nicht die falsch-positiven Befunde. Grundtrefferquote und Aufdeckungsrate eignen sich aufgrund der starken Abhängigkeit von der Prävalenz nicht sonderlich gut zur Überprüfung der Leistungsfähigkeit eines Screenings, insbesondere wenn die Störung eher selten auftritt.

Tab. 2: Bedeutung und Berechnung der Güteindizes (angelehnt an Marx, Jansen & Skowronek 2000, S. 46ff. und Tröster 2009, S. 105)

Güte-indizes	Bedeutung	Formel
Sensitivität (SE)	Die Sensitivität erfasst den prozentualen Anteil der Risikokinder, die durch den Prädiktor korrekt vorhergesagt wurden.	$SE = \frac{a}{a+c} \cdot 100$
Spezifizität (SP)	Die Spezifität erfasst den prozentualen Anteil der Nicht-Risikokinder, die durch den Prädiktor prognostiziert werden konnten.	$SP = \frac{b}{b+d} \cdot 100$
Grundrate (GR)	Die Grundrate beschreibt die prozentuale Häufigkeit von Risikokindern.	$GR = \frac{a+c}{N} \cdot 100$
Selektionsrate (SR)	Die Selektionsrate gibt die prozentuale Häufigkeit der Risikokinder in der Stichprobe an.	$SR = \frac{a+b}{100} \cdot 100$
Gesamttrefferquote (GQ)	Die Gesamttrefferquote bezeichnet den Anteil der valide positiv und valide negativ klassifizierten Probanden in der Stichprobe.	$GQ = \frac{a+d}{N} \cdot 100$

Zufalls-treffer-quote (ZQ)	Die Zufallstrefferquote ist die Quote der richtigen Vorhersagen, die auf Grundlage der jeweiligen Verteilung bei einer Zufallszuweisung zu erwarten sind.	$ZQ = \frac{\frac{(a+b)\cdot(a+c)}{N} + \frac{(c+d)\cdot(b+d)}{N}}{N} \cdot 100$
Maximal-treffer-quote (MQ)	Die Maximaltrefferquote begrenzt die Grundrate nach oben und ist die Differenz von Selektionsrate und Grundrate.	$MQ = 1 - (\lvert SR - GR \rvert)$
Relativer Steiger-ungsbetrag gegenüber dem Zufall (RATZ)	Der RATZ-Index bestimmt die Dominanz der Trefferquote im Gegensatz zur Zufallsquote mit Rücksichtnahme auf die Maximaltrefferquote. Der RATZ-Wert ist ein standardisierter Wert, der die Ergebnisse der individuellen Voraussage erfasst.	$RATZ = \frac{GQ - ZQ}{MQ - ZQ} \cdot 100$
Positive Korrekt-heit (PPW)	Die positive Korrektheit (oder auch positiver prädikativer Wert, PPW) beschreibt den prozentualen Anteil an valid positiven Probanden. Diese Kinder wurden sowohl durch den Prädiktor als auch durch das Kriterium als auffällig identifiziert.	$PPW = \frac{a}{(a+b)} \cdot 100$
Negative Korrekt-heit (NPW)	Die negative Korrektheit (oder auch negativer prädikativer Wert, NPW) beschreibt den prozentualen Anteil der valid negativen Probanden. Diese Kinder wurden sowohl durch den Prädiktor als auch durch das Kriterium als unauffällig identifiziert.	$NPW = \frac{d}{(c+d)} \cdot 100$
Youden-Index (YI)	Maß für die Trennschärfe des Screenings	$YI = \frac{SE + SP}{100} - 1$
Likelihood Ratio (LR^+ bzw. LR^-)	Der Likelihood-Ratio eines positiven Screening-Befundes (LR^+) gibt die Wahrscheinlichkeit an, um wie viel mal wahrscheinlicher ein	$LR^+ = \frac{SE}{100 - SP}$

	positiver Befund bei Kindern ist, die Probleme aufweisen im Vergleich zu Kindern, die diese Probleme nicht aufzeigen.	
	Der Likelihood-Ratio für einen negativen Screening-Befund (LR^-) zeigt dagegen an, wie wahrscheinlich es ist, dass Kinder einen negativen Befund erhalten, die Probleme aufweisen, im Vergleich zu denen, die die Schwierigkeiten nicht aufweisen.	$LR^- = \frac{100 - SE}{SP}$
Odds Ratio (OR)	Der Odds Ratio gibt an, wie sehr das Verhältnis zwischen Kinder mit einem Problem zu Kindern ohne ein Problem bei einem positiven bzw. einem negativen Screening-Befund wächst.	$OR = \frac{a \cdot d}{c \cdot b}$

Anmerkung: a = Anzahl richtig-positiver Screening-Befunde, b = Anzahl falsch-positiver Befunde, c = Anzahl falsch-negativer Befunde, d = Anzahl richtig-negativer Befunde

Der *Youden-Index* stattdessen ist unabhängig von der Grundquote und wird aus der Sensitivität und Spezifität berechnet. Es werden falsch negative und falsch positive Klassifikationen berechnet. Der Index kann Werte zwischen -1 und +1 annehmen. Wenn das Screening trennscharf ist, ist der Youden-Index positiv. Die Trennschärfe wird umso besser, je näher sich der Wert der 1 nähert.

Auch das *Likelihood Ratio* ist ein Wert, der sowohl Spezifität als auch Sensitivität ohne Abhängigkeit zur Grundquote berücksichtigt. Ein positives Likelihood Ratio (LR^+) ist die Wahrscheinlichkeit eines positiven Screening-Befundes, wenn die Störung nicht vorliegt. Je größer der LR^+ ist, umso mehr kann auf das Vorhandensein einer Störung geschlossen werden (vgl. a. a. O., S. 95). Der negative Likelihood Ratio (LR^-) gibt das Verhältnis zwischen den falsch-negativen und den richtig-negativen Befunden des Screenings an. Ist der Wert kleiner als 1, so ist die Wahrscheinlichkeit eines negativen Befundes bei Personen, die Probleme aufweisen, geringer als bei Personen, die eine Störung aufweisen. Je mehr sich der Wert an Null annähert, umso höher ist die Wahrscheinlichkeit, dass durch einen negativen Befund die Störung ausgeschlossen werden kann. Für die Beurteilung des Likelihood Ratio kann auf Tab. 3 zurückgegriffen werden.

Tab. 3: Interpretation der Testeffizenz des Likelihood Ratio (nach Bender 2001, S. 118)

LR^+	LR^-	Testeffizienz
> 10	< 0.10	sehr gut
5 – 10	0.1 – 0.2	gut
2 – 5	0.2 – 0.5	mäßig
< 2	> 0.5	schlecht

Das *Odds* bezeichnet die Chance für das Eintreten eines bestimmten Ereignisses. Das Odds eines positiven Screening-Befundes (OD^+) ist das Verhältnis von richtig-positiven zu den falsch-negativen Screening-Befunden. $OD^+ = 5$ bedeutet, dass eine Chance von 1 zu 5 besteht, dass die Störung bei einem positiven Screening eintritt, also würden vier von fünf Kindern eine Störung aufweisen, während eine Person diese nicht aufzeigt. Bei der Bewertung eines Screening-Tests sollte das Odds damit bei einem positiven Befund möglichst groß sein, was bedeutet, dass die Chance bei einem positiven Befund größer ist, dass die Störung auch wirklich auftritt. Das Odds bei einem negativen Befund sollte möglichst klein sein, so ist die Chance umso geringer, dass bei einem negativen Befund die Störung auftritt. Das *Odds Ratio* wiederum beschreibt das Chancenverhältnis der positiven und negativen Odds. Tab. 4 beschreibt die Bewertung des Odds Ratio-Wertes.

Tab. 4: Interpretation des Odds Ratio (nach Tröster 2009, S. 103)

OR = 1	Die Chance, dass die Störung auftritt, ist bei einem positiven und negativen Screening-Befund gleich groß.
OR < 1	Die Chance, dass die Störung auftritt, ist bei einem positiven Screening-Befund geringer als bei einem negativen Screening-Befund.
OR > 1	Die Chance, dass die Störung auftritt, ist bei einem positiven Screening-Befund größer als bei einem negativen Screening-Befund.

Das *Likelihood Ratio* (Wahrscheinlichkeitsverhältnis) ist für die Praxis eine bessere Entscheidungshilfe, z. B. für die Implementation von Interventionsmaßnahmen. Der Vorteil des Odds Ratio besteht in der Darstellung der Vorhersagegenauigkeit in einer einzigen Maßzahl, es wird häufig zum Vergleich von Screening-Verfahren genutzt (vgl. Tröster 2009,

S. 104). Beide symbolisieren die Prognosesicherheit ohne Abhängigkeit von der Prävalenzrate.

Bei einem Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten im Vorschulalter sollte insbesondere die Sensitivität des Verfahrens sehr hoch sein, damit man nicht Gefahr läuft eines der eigentlichen Risikokinder mittels des Screenings nicht zu erkennen. Zudem sollte das negative Likelihood Ratio möglichst klein sein, das symbolisiert, mit welcher Sicherheit durch einen negativen Befund die Lese-Rechtschreibschwierigkeiten ausgeschlossen werden können (vgl. a. a. O., S. 106). Marx und Lenhard (2010, S. 9ff.) stellten die Sensitivität, Spezifizität und die Prädiktortrefferquote als wesentliche Kennwerte heraus. Die Gesamttrefferquote bewerten sie dagegen als nicht sonderlich aussagekräftig. Die Trefferquote sollte immer im Vergleich mit der Zufallstrefferquote betrachtet werden. Dies weist darauf hin, dass es zum Vergleich von Screening-Verfahren noch weitere Güteindizes zu betrachten gibt.

Die Güteindizes an sich reichen noch nicht aus, um die Leistungsfähigkeit eines Screenings zu beurteilen. Es bedarf Beurteilungskriterien, die berücksichtigen, dass ein bestimmter Anteil der identifizierten Risikokinder hätte durch den Zufall identifiziert werden können. Entwickeln bspw. nur 10 % der Gesamtheit der Kinder ein bestimmtes Defizit, dann ist die Zufallsquote, um durch Raten einen Treffer für ein Kind ohne Defizit zu erzielen, bei 90 %. D. h. ein Screening mit einer negativen Korrektheit von 90 % ist genauso zu bewerten, wie ein Zufallstreffer. Das Screening sollte also bestenfalls mehr richtige Vorhersagen liefern, als eine Zufallsvorhersage es machen würde.

Der *Anteil der Zufallstreffer* lässt sich aus der Multiplikation der Randhäufigkeiten des Screenings dividiert durch die Gesamtzahl der untersuchten Personen berechnen (nach Tröster 2009, S. 136f., vgl. Formel 1).

Formel 1: Anteil der Zufallstreffer - E (aus Tröster 2009, S. 136/137)

$$E\,(a) = \frac{(a+c)\cdot(a+b)}{N}$$

$$E\,(b) = \frac{(b+d)\cdot(c+d)}{N}$$

$$E\,(c) = \frac{(a+c)\cdot(c+d)}{N}$$

$$E\ (d) = \frac{(b+d)\cdot(a+b)}{N}$$

Anmerkung: a = Anzahl richtig-positiver Screening-Befunde, b = Anzahl falsch-positiver Befunde, c = Anzahl falsch-negativer Befunde, d = Anzahl richtig-negativer Befunde

Aus den Zufallstreffern kann wiederum die *Selektionsquote* berechnet werden, die auch über diese Formel bestimmt werden kann:

Formel 2: Selektionsquote - SQ (aus Tröster 2009, S. 138)

$$SQ = \frac{a+b}{N}$$

Anmerkung: a = Anzahl richtig-positiver Screening-Befunde, b = Anzahl falsch-positiver Befunde

Die Sensitivität der Zufallsvorhersage ist die Selektionsquote. Wenn das Screening also per Zufall eine Vorhersage trifft, dann würde die Sensitivität der Selektionsquote entsprechen. Die Selektionsquote ist somit eine Grenze, die nur bei einem negativen Zusammenhang zwischen Prädiktor und Kriterium besteht. Die Sensitivität sollte möglichst größer sein als die Selektionsrate. Auch die positive Korrektheit sollte größer sein als die *Grundquote*. Die Grundquote entspricht wiederum der positiven Korrektheit der Zufallsvorhersage:

Formel 3: Grundquote - GQ (aus Tröster 2009, S. 138)

$$GQ = \frac{a+c}{N}$$

Anmerkung: a = Anzahl richtig-positiver Screening-Befunde, c = Anzahl falsch-negativer Befunde

Screenings lassen sich nicht aufgrund ihrer Güteindizes vergleichen, wenn sie sich in ihrer Selektionsquote unterscheiden oder wenn die Stichproben eine unterschiedliche Grundquote aufweisen. Um die Leistungsfähigkeit zu vergleichen, müssen also Grund- und Selektionsquote eine Berücksichtigung finden.

Im Folgenden werden Vergleichsmöglichkeiten über die relative Sensitivität, die relative Spezifität, die relative positive Korrektheit, die relative negative Korrektheit und den RATZ-Index geschaffen.

Die *Relative Sensitivität* misst, inwieweit die Sensitivität des Screenings die Sensitivität einer zufälligen Vorhersage übertrifft, es wird also die Sensitivität des Tests bzgl. der Selektionsquote relativiert. Sie beschreibt den Anteil der Steigerung der Sensitivität an der maximalen Sensitivitätssteigerung.

Formel 4: Relative Sensitivität - RSE (aus Tröster 2009, S. 140)

$$RSE = \frac{SE - SQ}{100 - SQ} \cdot 100$$

Anmerkung: SE = Sensitivität, SQ = Selektionsquote

Die Relative Sensitivität geht gegen Null, wenn das Screening kaum eine bessere Vorhersage als der Zufall treffen kann. Es nähert sich also die Sensitivität an die Selektionsquote an. Sie geht gegen 100 %, wenn alle Störungen identifiziert worden sind. Anhand der relativen Sensitivität lassen sich Screening-Verfahren mit unterschiedlicher Selektionsquote vergleichen (vgl. Tröster 2009, S. 140).

Die *Relative Spezifität* ist der Anteil der Steigerung der Spezifität an der maximalen Spezifitätssteigerung, wenn die Selektionsquote gegeben ist.

Formel 5: Relative Spezifizität - RSP (aus Tröster 2009, S. 141)

$$RSP = \frac{SP + SQ - 100}{SQ} \cdot 100$$

Anmerkung: SP = Spezifizität, SQ = Selektionsquote

Durch die Relative Spezifität lassen sich Screenings vergleichen, die eine unterschiedliche Selektionsquote haben. Es wird verglichen, inwieweit Nicht-Risikokinder auch richtig erkannt werden. Wenn die Vorhersagesicherheit nicht noch unter die Zufallssicherheit fällt, dann kann die Relative Spezifität Werte zwischen 0 und 100 % annehmen. Wenn sie sich 0% nähert, ist sie der Zufallswahrvorhersage sehr nah, sie erreicht 100 %, wenn alle Kinder, die keine Störung aufweisen, richtig klassifiziert worden sind (vgl. a. a. O., S. 141).

Die Relative Positive Korrektheit und die Relative Negative Korrektheit sind beide von der Grundquote abhängig.

Die *Relative Positive Korrektheit* kann verglichen werden, wenn sie bzgl. der Grundquote relativiert wird:

Formel 6: Relative Positive Korrektheit bzw. positiver prädiktiver Wert - RPPW (aus Tröster 2009, S. 141)

$$RPPW = \frac{PPW - GQ}{100 - GQ} \cdot 100$$

Anmerkung: PPW = Positive Korrektheit (positiver prädiktiver Wert), GQ = Gesamttrefferquote

Sie gibt den Anteil der Steigerung der Positiven Korrektheit gegenüber der einer Zufallsvorhersage im Verhältnis zur maximalen Steigerung wieder. Damit können Screenings mit unterschiedlichen Populationen, also unterschiedlichen Grundquoten, verglichen werden. Es ist derjenige Test leistungsfähiger, der bzgl. der Grundquote eine höhere Trefferwahrscheinlichkeit hat.

Die *Relative Negative Korrektheit* kann verglichen werden, wenn sie bzgl. der Grundquote relativiert wird:

Formel 7: Relative Negative Korrektheit bzw. negativer prädiktiver Wert - RNPW (aus Tröster 2009, S. 142)

$$RNPW = \frac{NPW + GQ - 100}{GQ} \cdot 100$$

Anmerkung: NPW = Negative Korrektheit (negativer prädiktiver Wert), GQ = Gesamttrefferquote

Sie gibt den Anteil der Steigerung der Negativen Korrektheit gegenüber der einer Zufallsvorhersage im Verhältnis zur maximalen Steigerung wieder. Screenings mit unterschiedlichen Populationen, also unterschiedlichen Grundquoten, können damit verglichen werden. Die Relative Negative Korrektheit misst die Wahrscheinlichkeit, inwieweit ein negativer Befund korrekt ist, über den Anteil der Zufallstreffer (vgl. Tröster 2009, S. 142f.).

Der *RATZ-Index* schafft ebenfalls eine Vorhersage mit Berücksichtigung des Zufalls bzgl. der Vorhersagegenauigkeit eines Screenings. Der RATZ-Index bestimmt den Anteil der richtigen Klassifikationen an den maximal möglichen richtigen Vorhersagen. RATZ bedeutet „Relativer Anstieg der Treffer gegenüber der Zufallstrefferquote“ (vgl. a. a. O., S. 143).

Formel 8: RATZ-Index (aus Tröster 2009, S. 143)

$$RATZ = \frac{GQ - ZW}{MQ - ZQ} \cdot 100$$

Anmerkung: GQ = Gesamttrefferquote, ZQ = Zufallstrefferquote, MQ = Maximaltrefferquote

Beim RATZ-Index werden auch die Selektionsquote und die Grundquote berücksichtigt, so dass ein Vergleich der Screening-Verfahren hinsichtlich ihrer Vorhersagequalität möglich wird.

Jansen (1992 zit. nach Jansen, Mannhaupt, Marx und Skowronek 2002, S. 47) bewerten den RATZ-Index wie in Tab. 5 dargestellt.

Tab. 5: Interpretation der RATZ-Werte nach Jansen (1992, in Jansen, Mannhaupt, Marx & Skowronek 2002, S. 47), Angaben in Prozent

Wert	Interpretation
> 66	sehr gute Klassifikation
34 - 66	gute, aber noch unspezifische Klassifikation
< 34	ungenügende Klassifikation

2.1.4 Weitere Gütekriterien

Zu den sogenannten „Nebengütekriterien“ gehören außerdem (vgl. Ingenkamp & Lissemann 2008, S. 60f.):

- *Normierung*: Die Ergebnisse des Tests werden in ein Bezugssystem eingeordnet.
- *Vergleichbarkeit*: Der Test liegt in Parallelformen vor oder in einer validen ähnlichen Fassung, sodass Vergleiche angestellt werden können.
- *Ökonomie*: Der Test benötigt wenig Zeit und Material für die Durchführung und Auswertung und ist bestenfalls auch in Gruppen anwendbar.
- *Nützlichkeit*: Die Ergebnisse des Tests sind besonders nützlich bzw. von hoher praktischer Relevanz.

Außerdem wurden von der Föderation deutscher Psychologenvereinigungen noch folgende Kriterien hinzugefügt (vgl. Moosbrugger & Rauch 2004, S. 195ff.; Ingenkamp & Lissmann 2008, S. 60f.):

- *Zumutbarkeit*: Das Ausmaß der zeitlichen und psychologischen Beanspruchung wird bewertet.
- *Unverfälschbarkeit*: Der Proband sollte möglichst nicht bei der Bearbeitung des Tests (z.B. durch Suggestivfragen) beeinflusst werden. Hierbei sollte insbesondere die soziale Erwünschtheit als Kriterium betrachtet werden.
- *Testfairness*: Geschlechtsspezifische, ethnische und kulturelle Diskriminierungen sollten möglichst ausgeschlossen werden.

Die im Folgenden vorgestellten Diagnostikverfahren für die Früherkennung der Lese-Rechtschreibschwäche in der Schule werden insbesondere bezüglich der hier vorgestellten Gütekriterien und der Kennwerte der prognostischen Validität interpretiert und speziell ihrer Vorhersagekraft entsprechend beurteilt, ob diese Risikokinder mit möglichst hoher Wahrscheinlichkeit frühzeitig identifizieren können.

2.2 Ausgewählte Testverfahren zur Vorhersage von Lese- und Rechtschreibschwierigkeiten im Vorschulalter

Zur Vorhersage von schulischen Lese- und Rechtschreibproblemen gibt es bereits verschiedene diagnostische Verfahren, die in der Praxis Anwendung finden. Siegel (1989a, S. 201ff.) postuliert eine Legastheniediagnostik anhand der phonologischen Fertigkeiten. Er befürwortet ebenfalls auf der Grundlage von Untersuchungen, die zeigen konnten, dass auch Kinder mit einem IQ von 80 Lesen und Schreiben lernen konnten, dass die Intelligenz bei der Diagnostik möglichst ignoriert bzw. allenfalls ein IQ-Richtwert von über 80 verwendet wird. Stanovich (1988, S. 590ff.) unterstützt diese These, dass die Hauptschwierigkeit der Kinder mit Lese-Rechtschreibschwäche unabhängig von dem Intelligenzwert einen Mangel in der phonologischen Informationsverarbeitung als Ursache hat. Ebenfalls in den nun im Folgenden vorgestellten Screening-Verfahren lässt sich ein Fokus auf die phonologische Bewusstheit oder auch die phonologische Informationsverarbeitung verzeichnen.

In diesem Kapitel werden fünf der gängigen Verfahren zur Prognose von Lese- und Schreibproblemen beschreiben, analysiert und miteinander in Beziehung gesetzt. Hierzu

dienen unter anderem insbesondere die Güteindizes für Screening-Verfahren für die prognostische Validität (vgl. Kap. I.2.1.3.2).

Für die Erfassung der Leistungen im Vorschulbereich dienen das Bielefelder Screening (BISC) von Jansen, Mannhaupt, Marx und Skowronek (2002), die Differenzierungsprobe von Breuer und Weuffen (2006), der Gruppentest zur Früherkennung von Lese- und Rechtschreibschwierigkeiten (PB-LRS) von Barth und Gomm (2008a) und das Heidelberger Auditive Screening in der Einschulungsuntersuchung (HASE) von Schöler und Brunner (2008). Der Rundgang durch Hörhausen (Martschinke, Kirschhock & Frank 2011) wird üblicherweise Anfang oder Mitte der ersten Klasse verwendet und überprüft ebenso die Vorläuferfertigkeiten zum späteren Lesen und Rechtschreiben.

Die Auswahl der hier im Folgenden vorgestellten fünf Verfahren wurde anhand dieser Kriterien getroffen.
Die Verfahren ...

(a) sind deutschsprachig.
(b) dienen der Vorhersage von Schwierigkeiten im Schriftspracherwerb.
(c) wurden empirisch evaluiert.
(d) können im Kindergarten und/oder zu Beginn der ersten Klasse eingesetzt werden.

Das Bielefelder Screening (BISC) wird in diesem Kapitel aufgrund umfassender Studien und Replikationsstudien ausgesucht, außerdem dient es in der vorliegenden Untersuchung als Testinstrumentarium. Die Differenzierungsprobe wurde gewählt, weil sie im Vergleich zu den anderen Verfahren auf die Wahrnehmungsbereiche wie z.B. die optisch-graphomotorische Entwicklung eingeht. Die PB-LRS ist einer der ersten wissenschaftlich fundierten Gruppentests, die für die phonologische Bewusstheit zur Verfügung stehen. HASE erfasst zusätzlich zur Leistung des phonologischen Arbeitsgedächtnisses auch Bereiche des Sprachentwicklungsstands der Kinder. Der Rundgang durch Hörhausen wird im Folgenden vorgestellt, da dieses Verfahren einen besonderen spielerischen Charakter aufweist und außerdem mit einem Trainingsverfahren zur phonologischen Bewusstheit im weiteren und im engeren Sinne kombiniert wird.

2.2.1 Bielefelder Screening (BISC)

Das Bielefelder Screening (BISC) wurde im Rahmen einer Längsschnittstudie in Bielefeld entwickelt (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002) und dient der frühzeitigen

Identifikation von Kindern, die aufgrund von spezifischen Vorläuferfertigkeiten ein höheres Risiko für eine Lese-Rechtschreibschwäche in der Schulzeit tragen.

Die Autoren achteten bei der Konzeption des Screening-Verfahrens und bei der Überprüfung durch die Validierungsstudie auf drei wichtige Aspekte (vgl. Marx, Jansen & Skowronek 2000, S. 11):

- Um den Kindern mit einer gezielten Diagnostik helfen zu können, wird statt einer korrelativen Vorhersage eine klassifikatorische Vorhersage getroffen.
- Die Prädiktoren wurden theoriegeleitet aus vorliegenden Befunden und der Entwicklung des Lesens und Rechtschreibens abgeleitet.
- Es wurde eine Einordnung zwischen spezifischen, unspezifischen und konfundierten Prädiktoren und Kriterien vorgenommen (vgl. Tab. 6).

Tab. 6: Einteilung von Prädiktoren und Kriterien in der Bielefelder Längsschnittstudie (nach Marx, Jansen & Skowronek 2000, S. 17)

Merk-mals-bereiche	Unterteilung der Prädiktoren oder Kriterien		
	spezifisch	unspezifisch	konfundiert
Prädik-toren	ausschließlich oder hauptsächlich an vorherzusagenden Kriterienfertigkeiten beteiligt	an vorherzusagenden und anderen Kriterienfertigkeiten gleichermaßen oder weniger beteiligt	durch vorherzusagende Kriterienfertigkeiten entwickelt oder verändert
Tests und Frage-bögen in der Biele-felder Studie	Testaufgaben zu Phonologischer Bewusstheit und Aufmerksamkeit und Gedächtnis (BISC)	• Grundintelligenztest • (CFT 1) • Erzieherinnenfrage-bogen (BBK) • Elternfragebogen	• Aufgaben zur Erfass-ung von Schrift-sprachkenntnissen vor Schuleintritt • Schriftsymbol-kenntnis
Kriterien	enthält nur vorhergesagte Fertigkeiten	enthält nicht nur vorhergesagte, sondern überwiegend andere Fertigkeiten	baut auf früheren Kriterienfertigkeiten auf
Tests in der Biele-felder Studie	• Lesetests (Knuspels Leseaufgaben) • Rechtschreibtests (DRT 2, Wortdiktat)	• Mathematiktest • Intelligenztest (CMM)	Testaufgaben zu Phonologischer Bewusstheit, Aufmerksamkeit und Gedächtnis und Schriftsprachleistungen nach Schuleintritt

Das „Bielefelder Screening zur Früherkennung von Lese- und Rechtschreibschwierigkeiten (BISC)“ ist als Einzeltest vier oder zehn Monate vor der Einschulung einsetzbar. Das Screening eignet sich weniger für Kinder, die schon schriftsprachliche Leistungen zeigen, da Deckeneffekte auftreten können. Bestandteile des Tests, die durch acht Aufgaben repräsentiert werden, sind die phonologische Bewusstheit, das Arbeitsgedächtnis und die visuellen Aufmerksamkeitsleistungen, die alle miteinander der phonologischen Informationsverarbeitung zugeordnet werden können (vgl. auch Kap. II.1.2). Die Untersuchungsaufgaben sind in einer Reihe von Vortests hinsichtlich ihrer Anwendung und ihrer Differenziertheit geprüft worden.

Dem Untersuchungsinstrument liegt ein kompensatorisches Modell zugrunde, demzufolge erst dann Schwierigkeiten in den schriftsprachlichen Fähigkeiten erwartet werden, wenn mehrere Vorläuferfertigkeiten betroffen sind. Die Autoren vermuten, dass Kinder mit Schwierigkeiten in einer der Vorläuferfertigkeiten diese Schwierigkeiten durch Stärken in anderen Bereichen kompensatorisch wieder ausgleichen können (vgl. Mannhaupt 2008, S. 142f.). Wenn eine besonders niedrige Leistung erzielt wurde, werden Risikopunkte vermerkt. Die Gefahr der Entwicklung einer Lese-Rechtschreibschwäche in der Schule wird anhand der Anzahl der Risikopunkte des Kindes bewertet. Die Aufgaben des Bielefelder Screenings wurden theoriegeleitet konzipiert (vgl. Tab. 7).

Tab. 7: Aufgaben des Bielefelder Screenings von Jansen, Mannhaupt, Marx & Skowronek (2002, S. 16ff.)

Phonologische Bewusstheit	Reimen Silben-Segmentieren Laut-zu-Wort Laute assoziieren
Visuelle Diskrimination	Wort-Vergleich-Suchaufgabe
Kapazität des Arbeitsgedächtnisses	Pseudowörter Nachsprechen
Zugriffsgeschwindigkeit auf das semantische Lexikon im Langzeitgedächtnis	Schnelles Benennen von schwarz/weiß Objekten von farbig inkongruenten Objekten

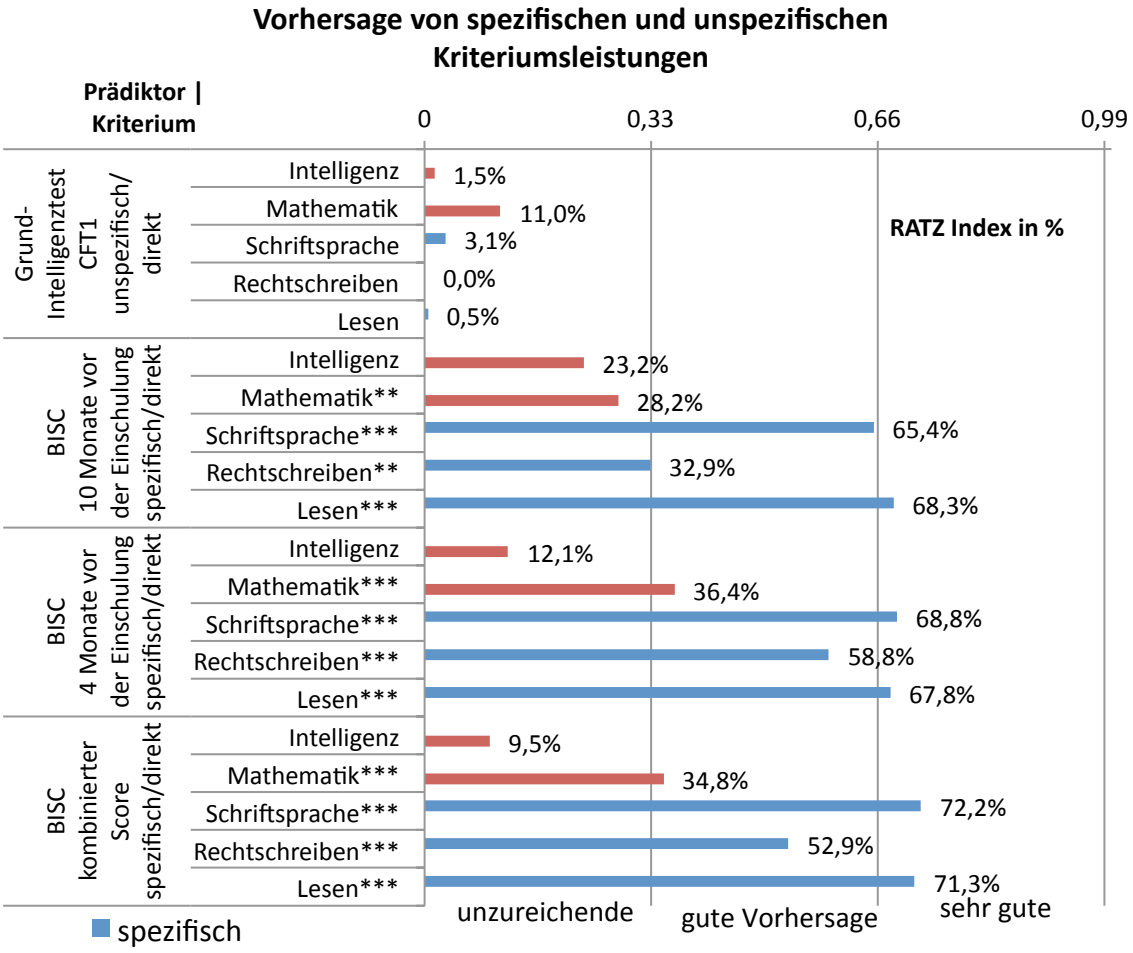
Auf der Grundlage der Bielefelder Längsschnittstudie wurde das Verfahren evaluiert und die Vorhersagegüte bis ins neunte Schuljahr hin erfasst. Für das Kriterium der Rechtschreibleistungen wurde ein standardisierter Rechtschreibtest (DRT 2, Diagnostischer Rechtschreibtest 2, Müller 1983) und für die Überprüfung der Leseleistung wurden Knuspels Leseaufgaben (Knuspel-L, Marx 1998) eingesetzt. Außerdem wurde ein Rechentest zum

Ende des zweiten Schuljahrs durchgeführt, um die spezifische Prädiktionsleistung für den Schriftspracherwerb zu prüfen (vgl. Tab. 6).

Das Ziel der Autoren bei der Erstellung des Screening-Verfahrens war, die prognostische Validität zu erhöhen und hauptsächlich spezifische Aufgaben auszuwählen. Sie wollten im Vergleich zu unspezifischen Verfahren damit bessere Klassifikationen (konkurrente Validität) und eine bessere Vorhersage der spezifischen Leistungen (differentielle Validität) erreichen (vgl. Skowronek & Marx 1989b, S. 39).

In der ersten Studie zum Bielefelder Screening konnte gezeigt werden, dass das Screening-Verfahren, durch klassische Gütekriterien bewertet, standhalten kann. Am ersten Testzeitpunkt waren 1 120 Vorschulkinder an der Erhebung beteiligt. Beim zweiten Testzeitpunkt wurden zwei Teilstichproben dieser Kinder nochmals überprüft. Das Screening-Verfahren wurde zweimal eingesetzt, um die Retest-Reliabilität und die Abhängigkeit der Prognose vom Testzeitpunkt zu überprüfen. Die erste randomisierte Teilstichprobe entsprach ca. 15 % der Gesamtstichprobe mit 171 Kindern und stellte damit eine repräsentative Stichprobe dar, die 35 Risikokinder umfasste. Die zweite Teilstichprobe wurde aus Risikokindern, die den unteren 15 % der restlichen Gesamtstichprobe angehörten, zusammengestellt. Die Risikostichprobe umfasste 169 Kinder (vgl. Skowronek & Marx 1989a, S. 287f.). Insgesamt lagen zum Ende des zweiten Testzeitpunktes dann die Daten von 136 Kindern der Repräsentativstichprobe und von 104 Kindern der Risikostichprobe vor. Die Ergebnisse zeigten, dass die klassischen Gütekriterien des Testes erfüllt werden konnten. Bei geschulten Testern ist die Durchführungs- und Auswertungsobjektivität des Tests gegeben, da dieser sowohl die genauen Instruktionen für die Kinder vorgibt als auch eine genaue Auswertung nach Punkten voraussetzt. An ökonomischen Maßstäben bemessen ist der BISC bei einer ungefähren Durchführungszeit von 30 Minuten in der Eins-zu-Eins-Situation als eher unökonomisch einzustufen. Die Retest-Reliabilität lag mit $r = 0.82$ ($p < .001$) bei einem noch guten Wert. Die Korrelation zwischen den kognitiven Leistungen (gemessen über den CFT 1, Weiß & Osterland 1980 und CMM 1-3, Schuck, Eggert & Raatz 1975) und dem Screening-Verfahren lag beim ersten Testzeitpunkt bei $r = .61$ ($p < .001$) und beim zweiten Testzeitpunkt bei $r = .52$ ($p < .001$). Zwischen dem Screening und dem Kriterium, den schriftsprachlichen Leistungen der Kinder, bestand eine Korrelation mit $r = .60$ ($p < .001$) beim ersten Testzeitpunkt und $r = .62$ beim zweiten Testzeitpunkt. Hier wird deutlich, dass die kognitiven Variablen einen sehr engen Zusammenhang zu den Fähigkeiten in der phonologischen Informationsverarbeitung aufzeigen.

Der BISC zeigte außerdem einen prädiktiven Wert auf die schulischen Lese-Rechtschreibleistungen. Risikokinder konnten damit bereits im Vorschulalter identifiziert werden. Dieser Wert wirkte sich auf die Lese-und Rechtschreibleistungen höher aus als die prädiktive Vorhersage durch allgemein kognitive Fähigkeiten. Der RATZ-Wert (vgl. Kap. I.2.1.3.2) betrug für die Rechtschreibleistung 52,9 % und fürs Lesen 71,3 %. Die gesamten schriftsprachlichen Leistungen wiesen eine Trefferrate von 90,2 % auf, die mit einem RATZ-Wert von 72,2 % weit über den Werten für eine zufällige Vorhersage liegt. Die Vorhersage von kognitiven oder mathematischen Leistungen entsprach dagegen den Zufallswerten, was die spezifische Prädiktionsleistung des Bielefelder Screenings unterstreicht (vgl. zusammenfassend Roth 1999, S. 124; Marx, Jansen & Skowronek 2000, S. 22ff.).



Anmerkung: * = p < .05, ** = p <.01, *** = p < .001, ohne Angabe = n.s.

Abb. 5: RATZ-Indizes und Signifikanzniveaus für die Vorhersage von spezifischen und unspezifischen Kriteriumsleistungen am Ende des 2. Schuljahres mit spezifischen, unspezifischen und konfundierten Prädiktoren (verändert nach Marx, Jansen & Skowronek 2000, S. 25)

Die prognostische Validität des Verfahrens schien durch die erste Studie erwiesen. Die Korrelationen zu anderen Testverfahren sind akzeptabel. Die Risikoeinschätzung durch den BISC ergab sich in 92,2 % der Fälle als richtig. Die im Vergleich zu anderen Studien sehr guten Ergebnisse dieser ersten Studie zum Bielefelder Screening zeigen jedoch trotz der guten Werte auch die Problematik von Fehlklassifikationen auf. Nicht alle als Risikokinder identifizierten Kindergartenkinder hatten später auch Schwierigkeiten im Lesen und Schreiben, und auf der anderen Seite entwickeln andere Kinder, die als Nicht-Risikokinder klassifiziert worden sind, in der Schule eine Lese-Rechtschreibschwäche. Die Sensitivität des Verfahrens lag je nach Kriterium und Erhebungszeitpunkt zwischen 50 und 75 %. Das Verfahren wurde als eine Art „Grobsiebverfahren“ entwickelt und zeigt bei besonders guten Kindern auch schon mal Deckeneffekte.

Schnitzler (2008, S. 87) merkt auch an, dass es sein kann, dass die Stichprobe aus der ersten Studie zum BISC mehr Risikokinder identifiziert hat, da die Drop-Outs der Stichprobe zum Teil prophylaktisch mit Kindern gleicher Leistungsstärke vor allem im unteren Bereich ersetzt wurden (vgl. Jansen, Mannhaupt, Marx & Skowronek 1999, S. 65). Schnitzler (ebd.) vermutet, dass der Schwellenwert für ein Defizit in der Stichprobe tiefer liegen könnte als in der Gesamtpopulation.

Die Autoren Schneider und Näslund (1993, S. 273ff.) replizierten in der *LOGIK-Studie* die Vorhersagegenauigkeit des BISC. Sie ermittelten für die Vorhersage von Rechtschreibschwierigkeiten zum Ende des dritten Schuljahres einen geringeren RATZ-Index als in der ersten Studie von 57 %. Etwas weniger als die Hälfte der Kinder in der Studie konnten richtigerweise als Risikokinder identifiziert werden (Sensitivität: 47,6 %). Die Prädiktortrefferquote lag am Ende der dritten Klasse bei 62,7 % bzgl. der Rechtschreibleistungen.

Brügelmann (2003a, S. 1) reanalysierte die Daten und bestätigte:

> *„Unsere Reanalyse von Daten aus der Münchener LOGIK-Studie bestätigt die Überlegenheit des BISC über alle anderen Prädiktoren für Prognosen bis zum Alter von acht Jahren (in der Regel Ende der 2. Klasse). Bezogen auf die Rechtschreibleistungen von 10jährigen und erst recht von 17jährigen sinkt die Klassifikationsleistung allerdings erheblich ab.“*

In einer Replikationsstudie im Jahr 2006 konnten die hohen RATZ-Indizes der ersten Untersuchung nicht repliziert werden (vgl. Marx & Weber 2006, S. 254ff.). An der Studie im

Jahr 2001 nahmen 176 Kinder im Großraum Würzburg teil. Bis zum vierten Schuljahr reduzierte sich die Stichprobe auf 154 Kinder. Jeweils am Ende des ersten bis vierten Schuljahres wurden eine Reihe von Tests zum Messen von Rechtschreib- und Leseleistung durchgeführt. Zusätzlich wurde ein Mathematiktest und zu Anfang der ersten Klasse ein nonverbaler Intelligenztest durchgeführt. Die Größe der Stichprobe wurde vergleichbar zur Erststudie gewählt. Insgesamt schnitt die Stichprobe von Marx und Weber (2006, S. 254) bezüglich ihrer kognitiven Leistungen etwas besser ab. Es konnten 33% der Problemkinder im Bereich Rechtschreibung aufgedeckt werden, wobei die Grenzfälle mit eingerechnet wurden. Der niedrigste RATZ-Index von 16 % am Ende des ersten Schuljahres konnte im Bereich Rechtschreibung berechnet werden (vgl. Tab. 8).

Tab. 8: RATZ-Indizes für die BISC-Messung 4 Monate vor der Einschulung und den Kriterien Lesen und Rechtschreiben (BISC-Werte aus Marx & Weber 2004, S. 199).

	Studie 1 (Marx & Skowronek 1989a)		Studie 2 (Marx & Weber 2006)	
Messzeitpunkt	Ende 1. Klasse	Ende 2. Klasse	Ende 1. Klasse	Ende 2. Klasse
Lesen/BISC	29,6 %	55,8 %	33 % / 25 %	19 % / 14 %
Rechtschreiben/BISC	35,8 %	49,6 %	16 %	25 %
Lesen + Rechtschreiben/BISC	35,5 %	77,1 %	44 %	69 %

Anmerkung: Unter „Lesen" bei Marx und Weber ist jeweils erst die Lesegeschwindigkeit und dahinter das Leseverständnis vermerkt. Die Grenzfälle wurden in diesen Zahlen als Risiko gewertet.

Aus Tab. 9 ist ersichtlich, dass alle RATZ-Werte bis auf die Werte zu den Rechtschreibleistungen im dritten und vierten Schuljahr nicht angemessen sind und deutlich unter den erhobenen Werten der ersten Evaluierungsstudie liegen.

Die Sensitivitätswerte legen dar, dass kaum mehr als die Hälfte der Kinder, die später Lese-Rechtschreibschwierigkeiten entwickelt haben, identifiziert worden sind (mit Ausnahme der Messungen mit dem DRT 3 und DRT 4), obwohl Grenzfälle in den Berechnungen mit aufgenommen worden sind. Die Intelligenzwerte (erhoben am Anfang der ersten Klasse) zeigen fast ähnliche Werte der Vorhersage wie die Lese-Rechtschreibschwierigkeiten, so dass die Autoren die diskriminante Validität in Frage stellen (vgl. a. a. O., S. 257). Durch die zeitliche Nähe des CFT und des BISC lassen sich jedoch keine genauen Aussagen treffen.

Tab. 9: RATZ-Indizes zur Vorhersagegüte des BISC, vier Monate vor Einschulung (nach Marx & Weber 2006, S. 256)

Erfasste Variable	Klasse	Test	RATZ
Rechtschreiben	1	DRT 1	16 %
	2	WRT 2+	25 %
	3	DRT 3	66 %
	4	DRT 4	53 %
Lesegeschwindigkeit	1	WLLP	33 %
	2	WLLP	19 %
	3	WLLP	14 %
	4	WLLP	16 %
Leseverständnis	1	LuV	25 %
	2	Näslund	14 %
	3	ELFE-Satz	17 %
	3	ELFE-Text	26 %
	4	ELFE-Satz	17 %
	4	ELFE-Text	44 %
Nonverbale Intelligenz	1	CFT 1	35 %
Mathematik	2	DEMAT 2	9 %
Evaluierungsstudie	1	Kombiniert	44 %
Evaluierungsstudie	2	Kombiniert	69 %

Anmerkung: DRT 1-4 = Deutscher Rechtschreibtest 1-4, WRT 2+ = Weingartener Grundwortschatz Rechtschreib-Test für 2. und 3. Klassen, LuV = Lesen und Verstehen, ELFE = Leseverständnistest für Erst- bis Sechstklässler, CFT 1 = Grundintelligenztest Skala 1, DEMAT 2 = Deutscher Mathematiktest für 2. Klassen

Die Vorhersagewerte der mathematischen Leistungen sind dagegen eher schlechter als die Vorhersagen der Lese-Rechtschreibleistungen. Für die Vorhersagewerte der Rechtschreibleistungen zum DRT 3 und DRT 4 berechneten die Autoren die Werte nochmals. Dieses Mal integrierten sie die Grenzfälle jedoch bei den Nicht-Risikokindern. Unter diesen Voraussetzungen ergaben sich nur noch ein RATZ-Index von 11 % für den DRT 3 und ein RATZ-Index von 32 % für den DRT 4. Der BISC wurde nur einmal durchgeführt, da die Autoren der Ansicht sind, dass dies in der pädagogischen Praxis auch eher realistisch sei (vgl. Marx & Weber 2006, S. 257). Gründe für die schlechteren

Vorhersagewerte könnten zum einen die unterschiedlichen Vorgehensweisen im Erstunterricht sein, der nicht genauer evaluiert worden ist, oder auch die unterschiedlichen Tests. In der ersten Evaluierungsstudie zum BISC wurden Knuspels Leseaufgaben (Knuspel-L, Marx 1998) genutzt, die einen größeren phonologischen Anteil integrieren.

2.2.2 Die Differenzierungsprobe

Die Differenzierungsprobe von Breuer und Weuffen (2006) besteht aus Einzeltests, die im Kindergartenalter oder zu Schulbeginn bei Kindern mit länger anhaltenden Sprachwahrnehmungsproblemen eingesetzt werden können. Breuer und Weuffen (1993, S. 21) gehen davon aus, dass die Sprachwahrnehmung eine grundlegende Fähigkeit zum Erlernen des Sprechens, Schreibens und Lesens ist. Sie unterscheiden in drei verschiedene Differenzierungsproben (vgl. Breuer & Weuffen 2006, S. 52ff.):

- DP 0: für Vier- bis Fünfjährige
- DP I: für Fünf- bis Sechsjährige oder für Schüler mit Lernschwierigkeiten im Anfangsunterricht
- DP II: für Sechs- bis Siebenjährige bzw. Schüler mit beständigen Lernschwierigkeiten

Die Aufgaben der Differenzierungsprobe umfassen unterschiedliche Wahrnehmungsbereiche (vgl. Tab. 10, Breuer 1998, S. 34ff.).

Einige Aufgaben ähneln den Aufgaben des Bielefelder Screenings, es ist jedoch zu bezweifeln, ob die Aufgaben, die vornehmlich auf der Wahrnehmungsebene angesiedelt sind, wirklich spätere Lese- und Rechtschreibschwierigkeiten vorhersagen können (vgl. auch Martschinke, Kirschhock & Frank 2005, S. 14). Die Wahrnehmungsleistung ist nach Martschinke, Kirschhock und Frank (a. a. O., S. 14f.) zwar eine relevante, jedoch keine spezifische Fähigkeit für den Erwerb schriftsprachlicher Fähigkeiten. Das Zusammenfassen der Aufgaben zu fünf verschiedenen Wahrnehmungsbereichen beschreiben sie als eine Art „Schrotschussproblematik“, die mehrere Lernvoraussetzungen für die Vorhersage von Leistungen in der Schriftsprache heranzieht, jede einzelne jedoch nur mittelmäßig berücksichtigt.

Tab. 10: Aufgaben der Differenzierungsprobe (nach Martschinke, Kirschhock & Frank 2005, S. 14)

Sprachwahrnehmungsleistungen	
Wahrnehmungsbereiche	Aufgaben
Optisch-graphomotorische Differenzierung	Den Buchstaben ähnliche Zeichen müssen abgemalt werden.
Akustisch-phonematische Differenzierung	Aus zwei klangähnlichen Begriffen muss der vorgesprochene Begriff dem richtigen Bild zugeordnet werden.
Kinästhetisch-artikulatorische Differenzierung	Dem Kind nicht geläufige Wörter müssen richtig nachgesprochen werden.
Melodische Differenzierung	Das Kinderlied „Alle meine Entchen“ soll ohne Melodieabweichungen und Rhythmusfehler vorgesungen werden.
Rhythmische Differenzierung	Kinder müssen einen vorgeklatschten Takt nachklatschen.

Steinbrink, Schwanda, Klatte und Lachmann (2010, S. 188ff.) führten erstmals eine Studie unter Berücksichtigung der Gütekriterien von Testverfahren durch. Sie konnten zeigen, dass die Differenzierungsprobe 3 bis 9 % der Varianz der späteren Lese- und Rechtschreibleistung aufklärt. Die Güteindizes Prädiktortrefferquote, Sensitivität (< 28 % für die Vorhersage in Klasse 1) oder auch der RATZ-Index (< 28 % für die Vorhersage in Klasse 1) sind größtenteils als inakzeptabel zu bewerten. Somit kann insgesamt konstatiert werden, dass die prognostische Validität des Verfahrens unzureichend ist und damit auch aufgrund der fehlenden theoretischen Basis nicht als Diagnostikinstrument zur Prognose von Lese-Rechtschreibschwierigkeiten empfohlen werden kann.

2.2.3 Phonologische Bewusstheit bei Kindergartenkindern und Schulanfängern (PB-LRS)

Barth und Gomm (2008a) entwickelten erstmals einen Gruppentest zur Früherkennung von Lese-Rechtschreibschwierigkeiten im Schulalter. Anhand dieses Tests werden Leistungen in der phonologischen Bewusstheit zeitökonomisch untersucht und die Kinder aufgrund von Risikogefährdung eingestuft. Der Test ist für Vorschulkinder (ein halbes Jahr vor der Einschulung) und Schulanfänger (in den ersten zwei bis vier Wochen) konzipiert und besteht je nach Untersuchungszeitraum aus vier bis sechs Untertests zur Reimerkennung, Silbensegmentierung, Anlautanalyse, Lautsynthese, Erfassung der Wortlänge oder zur

Identifikation des Anlautes. Die letzten beiden Untertests sind nur für den Einsatz bei Schulanfängern geeignet. Der Test dauert ca. 60 Minuten im ersten Schuljahr und 45 Minuten ein halbes Jahr vor Schulbeginn und kann in Kleingruppen mit bis zu zwölf Kindern durchgeführt werden. Jeder Untertest beinhaltet drei Instruktionsaufgaben, gefolgt von zehn Testitems. Zur Einschätzung des Risikos gibt es für jeden Untertest Cut-Off-Werte.

Der Test wurde anhand von 105 Kindergartenkindern und 474 Schulanfängern normiert. Sie stellten eine Sensitivität des Verfahrens von 63 % und eine Spezifität von 87,1 % fest (vgl. Tab. 11). In den Testkennwerten des PB-LRS ist jedoch kein RATZ-Wert angegeben. Marx und Lenhard (2010, S. 12) berechneten diesen anhand der gegebenen Daten nach. Der RATZ-Wert zum Kriterium Rechtschreibung beträgt 55 % (bei N = 450).

Tab. 11: Kennwerte der klassifikatorischen Validität des PB-LRS (vgl. Barth & Gomm 2008b, S. 35)

Selektionsrate	18,0 %
Gesamttrefferquote	84,7 %
Sensitivität	63,0 %
Spezifität	87,1 %
Prädiktortrefferquote	35,8 %

Die Evaluation des Verfahrens zeigte, dass nicht jedes für eine Lese-Rechtschreibschwäche gefährdete klassifizierte Kind später auch Probleme im Lesen und Schreiben entwickelte. Die Autoren empfehlen jedoch bei Unsicherheit eine Intervention zur frühzeitigen Förderung von phonologischer Bewusstheit einzuleiten (vgl. Barth & Gomm 2008a, S. 15). Das Verfahren ist aufgrund der Umsetzung als Gruppentest ökonomischer als die anderen hier dargestellten Verfahren.

2.2.4 Heidelberger Auditives Screening in der Einschulungsuntersuchung (HASE)

Das Heidelberger Auditive Screening in der Einschulungsuntersuchung (HASE) von Schöler und Brunner (2008) erfasst die Leistungsfähigkeit des phonologischen Arbeitsgedächtnisses und den Sprachentwicklungsstand der Kinder. Das Verfahren hat zum Ziel vier- bis sechsjährige Kinder mit Sprach- und Schriftspracherwerbsstörungen zu identifizieren.

Der inhaltliche Unterschied zu den anderen Messverfahren ist insofern erheblich, da der Fokus nicht ausschließlich auf die phonologische Bewusstheit gelegt wird. HASE enthält folgende Teilbereiche:

- NS – Nachsprechen von Sätzen
- WZ– Wiedergeben von Zahlenfolgen
- EW – Erkennen von Wortfamilien
- NK – Nachsprechen von Kunstwörtern sowie eines Zauberwortes

Die Aufgabenbereiche sollen reliable und valide Indikatoren für die allgemeine Sprachleistungsfähigkeit sowie für Bereiche der phonologischen Schleife des Arbeitsgedächtnisses darstellen. Es liegen Altersnormen und kritische Risikowerte vor. Die Durchführungs- und Auswertungszeit des Screening-Verfahrens beträgt ungefähr zehn Minuten mit der PC-Version.

HASE wurde anhand dreier Studien mit N = 2 940 (2003 in Mannheim), N = 1 467 (2003 in Biberach) und N = 2694 (2003 in Münster) Kindern evaluiert. Die Objektivität ist durch die genaue Vorgabe bei Durchführung und Auswertung gegeben. Cronbach's Alpha für die interne Konsistenz liegt zwischen $\alpha = .71$ und .83. Die kriterienbezogene Validität konnte über Korrelationen mit einem deutlichen Zusammenhang zwischen dem HASE-Screening und den Lese- und Rechtschreibleistungen im Grundschulalter bis zum Ende der dritten Klasse dokumentiert werden.

Die prognostische Validität konnte mit den Indizes für die Güte mit zufriedenstellenden bis guten Ergebnissen nachgewiesen werden (vgl. Schöler & Brunner 2008, S. 42). Aufgrund der Tatsache, dass einige Mitarbeiter der durchführenden Gesundheitsämter die Aufgaben ohne CD durchführten, vermuten Schöler und Brunner (ebd.) auf der Grundlage zu Berechnungen an einer bereinigten Stichprobe, dass die Ergebnisse für die Güteindizes noch besser ausfallen würden. Die RATZ-Indizes liegen in Abhängigkeit zum Kriterium und der jeweiligen Klassenstufe zwischen 39 und 72 % (vgl. Tab. 12).

Marx und Lenhard (2010, S. 12) vermuten jedoch, dass die hohen Werte zustande gekommen sind, weil die Autoren die Risikokriterien erst nach der Feststellung der Kritieriumsleistung festgelegt haben. Somit haben sie die Vorhersagegenauigkeit durch dieselben Datensätze gewonnen, mit denen auch die Festlegung der Cut-Offs stattgefunden hat. Sie fordern deshalb eine Replikation der Studie. Auch Mängel in der Durchführung

(keine Gewährleistung der Durchführungsobjektivität, da nicht alle Testleiter die beiliegende CD genutzt haben) machen eine erneute Evaluation notwendig.

Tab. 12: Güteindizes (in Prozent) für die klassifikatorische Validität von HASE (nach Tab. 14 aus Treutlein, Roos & Schöler 2007, S. 18)

Klasse	Kriteriumsleistung	SE	SP	PPW	NPW	RATZ
1	Lesenote	75	66	28	94	58
	Rechtschreibnote	69	64	23	93	49
	Rating Lesen	75	60	4	99	58
	Rating Rechtschreiben	76	60	6	99	59
2	Deutschnote	78	65	22	96	64
	Lesenote	77	65	19	96	61
	Rechtschreibnote	68	65	22	93	47
	Rating Lesen	50	61	0	100	18
	Rating Rechtschreiben	80	61	3	100	67
3	Deutschnote	75	67	25	95	59
	Lesenote	76	66	19	96	61
	Rechtschreibnote	65	70	38	87	43
	Rechtschreibleistung (DRT 3)	65	66	25	92	43
	Lesegeschwindigkeit (WLLP)	58	66	26	89	33
	Vorläuferfertigkeiten (Knuspel-L-Score 1)	73	70	37	92	56
	Leseverständnis (Knuspel-L-Score 2)	78	68	29	95	64

Anmerkung: SE = Sensitivität, SP = Spezifität, PPW = Positive Korrektheit, NPW = Negative Korrektheit, RATZ = Relativer Steigerungsbetrag gegenüber dem Zufall

2.2.5 *Der Rundgang durch Hörhausen*

Das Nürnberger Diagnose- und Förderprogramm besteht aus zwei Teilen (vgl. Abb. 6), die miteinander in Beziehung stehen:

- Das Diagnoseinstrument „Der Rundgang durch Hörhausen“ (Martschinke, Kirschhock & Frank 2011)
- Das Trainingsprogramm „Leichter lesen und Schreiben mit der Hexe Susi“ (Forster & Martschinke 2008)



Abb. 6: Übersicht zum Nürnberger Diagnose- und Förderprogramm (verändert nach Einsiedler 2003, S. 54)

Der Rundgang durch Hörhausen wurde als Einzeltest im Rahmen der Nürnberger Längsschnittstudien konzipiert. Der Zeitpunkt wurde zu Beginn des ersten Schuljahres gewählt. Die Autoren sind bei der Konzeption theoriegeleitet vorgegangen und gehen davon aus, dass die phonologische Bewusstheit eine wichtige Grundvoraussetzung für einen erfolgreichen Schriftspracherwerb darstellt (vgl. Martschinke, Kirschhock & Frank 2001,

S. 16). Es werden sieben Aufgaben zur phonologischen Bewusstheit im weiteren und im engeren Sinne abgefragt.

Aufgaben zu phonologischen Bewusstheit im weiteren Sinne:

- Silben segmentieren
- Silben zusammensetzen
- Endreime erkennen

Aufgaben zu phonologischen Bewusstheit im engeren Sinne:

- Anlaute erkennen
- Endlaute erkennen
- Phonemanalyse
- Lautanalyse mit Umkehraufgabe

Zusätzlich werden drei weitere Aufgaben gestellt, die schon frühe Kenntnisse der Schriftsprache abfragen. Diese Aufgaben zum Schreiben des eigenen Namens oder anderer erster Wörter sowie zur Abfrage der Buchstabenkenntnis gehen jedoch nicht in die Wertung mit ein. Alle Aufgaben werden in eine Geschichte, den Rundgang durch Hörhausen, eingebettet. Sie werden spielerisch inszeniert und die Kinder entwickeln beim Rundgang durch den Ort Hörhausen meistens eine Motivation an den einzelnen Stationen teilzunehmen.

Für die Auswertung werden die Werte der Leistungen für den Bereich der phonologischen Bewusstheit im engeren Sinn und für die phonologische Bewusstheit im weiteren Sinn einzeln zusammengefasst. Der Gesamtwert wird mit den Normen verglichen, um ein individuelles Leistungsniveau festzustellen. Kinder mit Schwierigkeiten im Bereich Sprache wurden innerhalb der Stichprobe nicht beachtet. Ein Risiko für Lese-Rechtschreibschwäche wird einem Kind zugeschrieben, wenn es in den Leistungsbereich der 20 % schwächsten Kinder der Normierungsstichprobe fällt. Der Test ist als Einzeltestverfahren konzipiert und dauert ca. 30 - 40 Minuten (vgl. Frank, Kirschhock & Martschinke 2003, S. 59ff.). Damit ist er ähnlich wie auch das Bielefelder Screening in der Anwendung als eher unökonomisch zu bewerten.

Die Validität wurde unter anderem mit Korrelationen überprüft, die als zufriedenstellend bewertet wurden. Am Ende des ersten Schuljahres wurde mit $N = 375$ Kindern aus 15 Klassen eine Korrelation zum Lesetest von $r = .49$ ($p < .001$) und eine Korrelation zum

Rechtschreibtest mit r = .40 (p < .001) festgestellt. Die Korrelation zum Ende des zweiten Schuljahres blieb zum Rechtschreibtest gleich, die Korrelation zum Lesetest betrug r = .39 (p < .001). Die Korrelation zum Leseverständnis betrug r = .44 (p < .001) zum Ende der ersten und zweiten Klasse. Für die Validitätsprüfung wurde auch die Vorhersagekraft des Verfahrens überprüft. Zusammenfassend liegen die prognostischen Werte für das Ende der ersten Klasse für das Leseverständnis und fürs Rechtschreiben im guten Bereich. Die RATZ-Werte zum Ende des zweiten Schuljahres sind jedoch nur noch für die Vorhersage des Leseverständnisses im guten Bereich (vgl. Tab. 13). Die Autoren empfehlen das Verfahren nicht für die Früherkennung von Lese-Rechtschreibschwäche, jedoch zur Früherkennung von Risikokindern bei phonologischen Schwierigkeiten.

Tab. 13: RATZ-Indizes vom „Rundgang durch Hörhausen“ zu zwei Messzeitpunkten (zusammengestellt aus Martschinke, Kirschhock & Frank 2005, S. 38-39)

Kriterium	RATZ-Index Ende Klasse 1	RATZ-Index Ende Klasse 2
Lesefertigkeit	25,0 %	29,2 %
Leseverständnis	76,9 %	39,4 %
Rechtschreiben („Graphemtreffer“)	36,6 %	29,2 %

Anmerkung: RATZ = Relativer Steigerungsbetrag gegenüber dem Zufall

Schnitzler (2008, S. 94) bemängelt die Relevanz der Aufgabe zum Silben zusammensetzen hinsichtlich des Schriftspracherwerbs und vermutet, dass diese zu schwer ist. Sie konnte auch nur von 30 % der Schulanfänger richtig beantwortet werden, während die anderen Aufgaben im Mittel eine Quote von 48 % aufwiesen (vgl. Martschinke, Kirschhock & Frank 2001, S. 34). Frank, Kirschhock und Martschinke (2003, S. 59) betonen, dass mit dem Verfahren zu Schulbeginn Schüler mit erhöhtem Förderbedarf ermittelt werden können, der jedoch erst durch die Betrachtung aller Aufgaben vom „Rundgang durch Hörhausen“ ausreichend abgesichert festgestellt werden kann.

Insgesamt scheint es sich um ein schülernahes, motivierendes Konzept zu handeln, das sich bzgl. der RATZ-Indizes jedoch nur sinnvoll zur Vorhersage von schriftsprachlichen Leistungen im ersten Schuljahr eignet. Ebenso wie bei HASE (vgl. Kap. I.2.2.4) wird zu überprüfen sein, ob sich in einer weiteren Studie diese Befunde replizieren lassen.

Im nächsten Abschnitt werden die hier vorgestellten Verfahren in Bezug auf ihr Einsatzgebiet, ihre Zielsetzungen und ihre erfassten Prädiktoren verglichen. Außerdem werden die errechneten RATZ-Werte auf der Grundlage der zugehörigen empirischen Untersuchungen der Verfahren zusammengestellt und somit die Verfahren aufgrund ihrer Vorhersageleistung vergleichsweise beurteilt.

2.3 Vergleich der Verfahren

Für den Schulbeginn lassen sich der PB-LRS sowie der Rundgang durch Hörhausen von Martschinke, Kirschhock und Frank (2001) einsetzen, alle anderen Verfahren werden ausschließlich im Vorschulbereich angewendet. Der Einsatzzeitpunkt, das Ziel der Untersuchung und die Prädiktoren, die bei den einzelnen Tests erfasst werden, sind in Tabelle 14 dargestellt.

Vergleicht man die Screening-Verfahren zur Prognose von Schwierigkeiten im Lesen und Rechtschreiben, so zeigt sich auf inhaltlicher Ebene eine deutliche Dominanz der Überprüfung der phonologischen Fähigkeiten. Nur die Differenzierungsprobe beleuchtet dagegen etwas andere Inhalte und auch HASE bezieht sich vorrangig auf den Sprachentwicklungsstand und auf Arbeitsgedächtnisprozesse.

Die Studien zu den einzelnen Verfahren konnten jedoch im Überblick keine überzeugenden Ergebnisse liefern, wie die RATZ-Werte der Studien darlegen (vgl. Tab. 15).

Im Bereich der Rechtschreibung zeigen die RATZ-Werte nur für die PB-LRS, HASE und den Rundgang durch Hörhausen, im Bereich Lesen nur HASE und der Rundgang durch Hörhausen (nur im Bereich Leseverständnis) eine gute, jedoch unspezifische Klassifikation an. Das Bielefelder Screening und die Differenzierungsprobe können bezüglich ihrer Prognoseleistung durch die vorliegenden Studien als inakzeptabel interpretiert werden. Die etwas positiveren Befunde zu dem Screening-Verfahren HASE sind nach Marx und Lenhard (2010, S. 14) auf die hohe Selektionsrate von über 40 % zurückzuführen. Sie fordern eine Replikation der Studie, da die Vorhersagegenauigkeit an den gleichen Werten ermittelt wurde, mit denen auch die Grenzwerte festgelegt worden sind.

Tab. 14: Verfahren zur Früherkennung von schulischen Lese- und Rechtschreibschwierigkeiten (erweitert nach Marx & Lenhard 2010, S. 290)

Verfahren	Zeitpunkt	Ziel	erfasst
Bielefelder Screening (BISC; Jansen, Mannhaupt, H. Marx & Skowronek 2002)	10 bzw. 4 Mon. vor Einschul-ung	Identifikation von Risikokindern für spätere LRS	• phonologische Bewusstheit im weiteren Sinne • phonologisches Arbeitsgedächtnis • schnelles Benennen • visuelle Aufmerksamkeit
Differenzierungs-probe (Breuer & Weuffen 2006)	5 -7 Jahre	Feststellung von Förderbedarf, Identifikation von Risikokindern für spätere LRS	• rhythmische Differenzierung • melodische Differenzierung • kinästhetische Differenzierung • phonematische Differenzierung • optische Differenzierung
Gruppentest zur Früherkennung von Lese- und Rechtschreib-schwierigkeiten (PB-LRS; Barth & Gomm 2008a)	vor oder kurz nach Schul-eintritt	Identifikation von Risikokindern für spätere LRS	• phonologische Bewusstheit im weiteren Sinne • phonologische Bewusstheit im engeren Sinne
Heidelberger Auditives Screening in der Einschulungs-untersuchung (HASE; Schöler & Brunner 2008)	4 - 6 Jahre	Erfassung von auditiven Infor-mationsver-arbeitungsstörungen, Prognose von Schriftspracherwerbs störungen	• Leistungsfähigkeit des phonologischen Arbeitsgedächtnisses • Sprachentwicklungsstand
Rundgang durch Hörhausen (Martschinke, Kirschhock & Frank 2011)	Anfang oder Mitte 1. Klasse	Auffinden von Förderansätzen, Identifikation von Risikokindern für spätere LRS	• phonologische Bewusstheit im weiteren Sinne • phonologische Bewusstheit im engeren Sinne

Keines der Screening-Verfahren kann, anhand der RATZ-Werte beurteilt, eine sehr gute Klassifikation sicherstellen. Es ist nicht eindeutig nachgewiesen, auf welche Ursachen diese teils mangelhafte Vorhersageleistung vorhandener Verfahren zurückzuführen ist. Im folgenden Abschnitt werden daher zunächst die Probleme einer Klassifikation von Risiko und Nicht-Risikokindern des Weiteren erörtert.

Tab. 15: Vergleich der RATZ-Werte anderer Studien zur Früherkennung von Lese- und Rechtschreibschwierigkeiten am Ende der ersten Klasse

Verfahren	Studie	N	RATZ (Lesen)	RATZ (Rechtschreiben)
BISC	Marx & Weber (2006)	176	33 % / 25 %*	16 %
DP	Steinbrink, Schwanda, Klatte & Lachmann (2010)	664	6 – 28 % (je nach Kriterium und Cut-Off)	12 – 18 % (je nach Kriterium und Cut-Off)
PB-LRS	Gütekriterien von Marx & Lenhard (2010, S. 12) berechnet auf der Basis von Barth & Gomm, 2004, Tab. 18	450	-	55 %
HASE	Treutlein, Roos & Schöler (2007, S. 18ff.)	1 562	58 %	49 – 59 % (je nach Kriterium)
Rundgang durch Hörhausen	Martschinke, Kirschhock & Frank, 2005, S. 38 f.)	375	25 % / 77 %*	37 %

Anmerkung: *erster Wert ist die Lesegeschwindigkeit, zweiter Wert das Leseverständnis

3 Probleme der Klassifikation von Risikokindern durch bisherige Testinstrumentarien

Die Fehlklassifikationen eines Diagnostikinstruments stehen in Abhängigkeit zu den jeweiligen Prädiktor- und Kriteriumsvariablen, die im Messverfahren eingesetzt wurden sowie zur Reliabilität dieser Verfahren. Außerdem ist die Differenz der Selektions- und der Prävalenzrate von Belang. Zur Erfassung von Vorläuferfertigkeiten sollten die prädiktive Validität der Messinstrumente nach Marx (1992a, S. 250) bedeutend und möglichst groß und die Trefferwahrscheinlichkeit größer oder gleich der Zufallswahrscheinlichkeit sein, sodass möglichst viele richtig-positive und richtig-negative Kinder klassifiziert werden können.

Es gibt somit nach Marx (1992a, S. 255) zwei unterschiedliche Fehlerarten, die beide in engem Zusammenhang mit der jeweiligen Grundrate oder Selektionsrate stehen. Es handelt sich dabei um die falsch-negativ klassifizierten Kinder, denen fälschlicherweise kein Risiko

vorhergesagt wurde und um die falsch-positiv klassifizierten Kinder, denen fälschlicherweise ein Risiko vorhergesagt wurde, obwohl sich dies später nach der Kriteriumsmessung nicht bestätigen konnte. Wenn die Grundrate im Wesentlichen mit der Selektionsrate übereinstimmt, dann gibt es wenige Fehlklassifizierungen. Wenn die Grundrate jedoch nun 15 % statt 7 % beträgt, also sie in der Klassifikation unterschätzt worden ist, dann ist die Rate der falsch klassifizierten Kinder recht hoch. Tröster (2009, S. 86) beschreibt mögliche Schäden der Falschklassifizierungen (vgl. Tab. 16).

Marx (1992a, S. 260) beschreibt, dass in der Literatur eine Ausgeglichenheit zwischen den beiden Fehlerarten gefordert wird. Dieser Ansatz wird aus den verschiedenen Perspektiven jedoch unterschiedlich betrachtet.

Tab. 16: Potenzielle Schäden durch falsche Screening-Befunde in der Früherkennung (nach Tröster 2009, S. 86)

Potenzieller Schaden infolge eines falsch-negativen Befundes	Potenzieller Schaden infolge eines falsch-positiven Befundes
• Chancen der frühen Intervention werden nicht genutzt • die Störung tritt in vollem Umfang auf • Folgeprobleme der Störung treten auf	• Belastung durch nachfolgende diagnostische und therapeutische Maßnahmen • Beunruhigung des Kindes und der Eltern • Stigmatisierung durch das soziale Umfeld des Kindes

Aus *forschungsethischer Richtung* wird die Bevorteilung einer der Fehlerarten sehr differenziert wahrgenommen. Einige Forscher glauben, dass ein Labelingprozess eintritt. Diese Aussage wird von Marx (1992a, S. 261) relativiert, da dieser Effekt schließlich auch positiv auf die Kinder wirken kann, die fälschlicherweise als Nicht-Risikokinder klassifiziert worden sind, sich aber aufgrund der self-fulfilling-prophecy wirklich zum Kind ohne Probleme entwickelt haben. Somit kann dieser Effekt auch positive forschungsethische Effekte aufweisen. Im Einzelfall ist immer wieder zu bewerten, ob der Widerstand gegen die Etikettierung oder der Nutzen der Interventionsmaßnahme vorteilhafter ist.

Aus *entwicklungspsychologischer Sicht* sind die beiden Fehlerklassifikationen unvermeidbar und nicht vorhersagbar, da bestimmte Variablen wie Erstunterricht, Schulatmosphäre oder Motivation erst nach der Prädiktormessung in Kraft treten.

Aus *ökonomischer Sichtweise* steht die Reduzierung der falsch-positiv klassifizierten Kinder im Vordergrund, da der Aufwand für die Förderung ansteigen würde.

Aus *pädagogisch-administrativer Sicht* ist die Beurteilung der beiden Fehlklassifikationen auch abhängig von der administrativen Durchlässigkeit. In Deutschland sind die Frühidentifikationen größtenteils unabhängig von der Schule. In einigen Bundesländern werden diese Daten nun zur besseren Verbindung von Schule und Kindergarten weitergereicht, ein Kind wird aber nicht vom gleichen pädagogischen Personal weiterbetreut, wie es bspw. in Amerika der Fall ist. In Deutschland ist die Gefahr einer Etikettierung damit nicht so groß.

Tröster (2009, S. 87) nimmt an, dass bei einem Vorschulscreening zur Früherkennung von Lese-Rechtschreibschwierigkeiten möglichst viele Kinder identifiziert werden sollen, die einer Förderung bedürfen. Somit würde es eher darauf ankommen, falsch-negative Befunde zu vermeiden. Falsch-positive Klassifizierungen könnten eher in Kauf genommen werden, da dann Kinder eine zusätzliche Förderung erhielten, die diese nicht unbedingt benötigt hätten, umgekehrt würden jedoch Kinder nicht gefördert werden, die Förderbedarf gehabt hätten. Diese Meinung kann auch Barth (2006, S. 183) vertreten. Der Schaden sei bedeutend geringer, als wenn Kinder weitreichendere Lernprobleme aufgrund unzureichender Förderung entwickeln. Außerdem sollte immer berücksichtigt werden, dass eine frühe Identifikation von Risikokindern nur dann sinnvoll ist, wenn auch sinnvolle Interventionsmaßnahmen zur präventiven Förderung der Kinder möglich sind (vgl. Kap. I.1.2). Auch ökonomische Aspekte sollten Berücksichtigung finden. Marx (1992a, S. 250) beschreibt, dass der Aufwand für die Früherkennung immer geringer sein sollte als für die anschließende Behandlung bzw. Interventionsmaßnahme.

Marx (1992a, S. 249) unterscheidet bei der vorschulischen Identifikation von Risikokindern in unvermeidbare Unzulänglichkeiten und vermeidbare Unzulänglichkeiten. Als unvermeidbare Unzulänglichkeiten bei der Früherkennung konstatiert er die inhaltliche Unvollständigkeit und methodische Unvollkommenheit der Messinstrumente, sowie den unvorhersehbaren Einfluss der Umwelt auf die Entwicklung der Kinder nach der Messung im Kindergarten. Außerdem ließe sich die Größe der Risikostichprobe vor der Messung des Kriteriums nur relativ bestimmen. Auch das Verhalten des Kindes zwischen der Prädiktor- und der Kriteriumsmessung und seine Reaktion auf die Umwelt sind unvorhersehbar. Vermeidbare Unzulänglichkeiten sind dagegen eine Ignoranz der methodischen Teststandards (z. B. zu kleine Stichproben, mangelnde theoretische Fundierung, inadäquate oder nicht vorhandene Kontrollgruppen, …).
Für die Reduzierung der Fehlklassifikation in der Früherfassung regt Marx (1992a, S. 256) eine dreischrittige Klassifikation an, in der zunächst eine Grobauslese durch ein Screening

vorgenommen wird, im Anschluss werden die identifizierten Kinder dann nochmals differenzierter untersucht, um dann im dritten Schritt eine Interventionsmaßnahme zu beginnen.

Die Identifizierung von Risikokindern wird von vielen Forschern kritisch betrachtet. Marx (1992a, S. 262) zeichnet Alternativvorschläge auf. Es wird vorgeschlagen, alle Kinder als potenziell hilfsbedürftig zu betrachten (vgl. Leach 1983 in Marx 1992a, S. 263) und die Dichotomisierung von Risikokindern bzw. Nicht-Risikokindern komplett aufzugeben, was jedoch bezüglich knapper Ressourcen eine neue Problematik aufwerfen könnte. Ein weiterer Ansatz ist, dass die Kinder einer späten Früherfassung am Anfang der Schule unterzogen werden, um gleich zu Beginn des Lernprozesses Schwierigkeiten im Schriftspracherwerb aufzudecken (vgl. Dummer 1989 in Marx 1992a, S. 263). Nach Marx (1992a, S. 263) verschiebt sich die Problematik der Klassifizierung dadurch nur zeitlich oder ist auf Extremfälle beschränkt. Die Alternativen scheinen dadurch nur verkappte Alternativen zu sein, die auf dieselbe Problematik der Klassifikation zurückführen.

Marx (1992a, S. 263f.) fasst die Vorschläge zur Verbesserung der Früherkennung von verschiedenen Autoren zusammen, die sich auf Mängel beziehen, die Studien möglichst vermeiden sollten. Zu den Aufgaben der Prädiktoren konstatiert er, dass diese als Basis für Folgebeobachtungen verwendet werden sollten, nicht nur zur einfachen Klassifikation. Außerdem sollten sie mindestens zwei verschiedene Bereiche umfassen und spezifische Aspekte prüfen. Eine kurze Zusammenfassung weiterer Aspekte folgt:

- Screening-Verfahren sollten in Hinblick auf eine spezifische Förderung ausgerichtet sein und nur durchgeführt werden, wenn eine Fördermaßnahme geplant ist.
- Screening-Verfahren sollten ohne Normierung und wissenschaftliche Fundierung im pädagogischen Einsatz nicht gebraucht werden.
- Es sollte die Möglichkeit einer „Grenzfalldiagnostik“ geben, also eine dritte Kategorie von Kindern, die ggf. das Screening wiederholen oder einer weiteren Untersuchung unterzogen werden.
- Zur Verminderung der Fehlklassifizierungen sollten Screening-Verfahren Angaben zu Kreuzvalidierungen über das gesamte Verfahren machen.
- Fehlklassifizierte Kinder sollten zur Korrektur der Prädiktorvariablen bzgl. ihrer Charakteristika den richtig klassifizierten Kindern gegenübergestellt werden.
- Vorhersagestudien sollten aufgrund der Einschätzung der Güte des Screenings nicht als Korrelationsstudien angelegt werden.

Eine hundertprozentige Identifikation aller Risikokinder ist nicht möglich. Dies wurde auch in der Replikationsstudie von Schneider und Näslund (1993, S. 273ff.) in der *LOGIK-Studie* gezeigt, die die Vorhersagegenauigkeit des BISC in einer zweiten Studie überprüfte. Weniger als die Hälfte der Kinder in der Studie konnten als Risikokinder identifiziert werden (Sensitivität: 47,6 %). Ein möglicher Grund könnte sein, dass die Klassifikationsleistung des BISC nur für den Beginn des Schriftspracherwerbs bedeutend ist und Kinder mit einem Rückstand im phonologischen Wissen diesen im Anfangsunterricht relativ schnell aufholen können. Brügelmann (2003a, S. 19ff.) vermutet, dass die hohe Trefferquote auch im Alter der Kinder in der LOGIK-Studie begründet ist. Viele der Risikokinder mussten ein Jahr später eingeschult werden, wurden jedoch regulär mit den anderen Kindern getestet, hatten dann aber ein Jahr weniger Beschulung hinter sich. Marx und Weber (2006, S. 253) vermuten zudem, dass der BISC auch allgemeine Leistungen vorhersagen könnte, da ein sehr großer Anteil der Risikokinder vom Schulbesuch zurückgestellt worden ist. Dies wurde in der ersten Studie wiederlegt, Marx und Weber (2006, S. 253) erscheint die Replikation dieses Sachverhalts als sinnvoll. Bei einem Ausschluss der wiederholenden Kinder ergab sich lt. Brügelmann (2003a, S. 19ff.) eine Prädiktortrefferquote von 18 % bzgl. der Rechtschreibleistungen am Ende der zweiten Klasse.

Brügelmann (a. a. O., S. 8) konnte in einer eigenen Reanalyse feststellen, dass die Prädiktortrefferquoten am Ende der ersten, zweiten und dritten Klasse jeweils unter 50 % lagen. Brügelmann (a. a. O., S. 30) konstatiert, dass man aufgrund der hohen Schwankungen in den RATZ-Werten zu verschiedenen Messzeitpunkten zur Vorsicht mahnen sollte, der phonologischen Bewusstheit eine „zentrale Rolle“ für die Entwicklung bzw. Prävention von Lese-Rechtschreibschwierigkeiten zuzusprechen. Außerdem betont Brügelmann (a. a. O., S. 1), dass die Repräsentativstudien an ihre Grenzen stoßen und Besonderheiten der individuellen Entwicklung nicht berücksichtigen können. Dies zeigt er anhand einiger Fallbeispiele, die trotz gleich gemessener Lernausgangslage sehr unterschiedliche Entwicklungsverläufe aufzeigen.

Brügelmann (2003a, S. 31) fasst folgende Ergebnisse zusammen:

- Die hohen linearen Korrelationen in der LOGIK-Studie zwischen Prädiktor und Kriterium müssen bei einer qualitativen Klassifikation relativiert werden.
- Die Klassifikationswerte der ersten Bielefelder Studie lassen sich nicht ohne weiteres auf andere Stichproben übertragen.
- Die langfristige Vorhersagekraft wird in der LOGIK- und in der ersten Evaluationsstudie in Bielefeld durch lernökologische Bedingungen überlagert.

Die schlechte prognostische Vorhersage konnte wie oben bereits ausgeführt (vgl. Kap. 2.2.1) auch in der Replikationsstudie zum Bielefelder Screening von Marx und Weber (2006, S. 245ff.) nochmals beeindruckend gezeigt werden, wobei die geringe Anzahl der Probanden als relativierend beachtet werden sollte. Jedes zweite Kind, das als Risikokind identifiziert worden war, hatte nach der Einschulung keine Schwierigkeiten beim Schriftspracherwerb. Die Studie wurde in Bayern durchgeführt, wo vor kurzer Zeit im Lehrplan verankert wurde, dass die Schüler der ersten Klassen zu Beginn des Erstleseunterrichts die Graphem-Phonem-Korrespondenzen und die phonologische Bewusstheit üben. Insgesamt handelt es sich um einen synthetisch orientierten Anfangsunterricht, der die Prognose beeinflussen kann. Marx und Weber (a. a. O., S. 254) vermuten, dass der Unterricht heute in allen Bundesländern mehrheitlich synthetisch erfolgt, eher seltener ganzheitlich. Auch in der Studie von Schabmann, Schmidt, Klicpera, Gasteiger-Klicpera und Klingebiel (2009, S. 315ff.) konnte nur eine sehr geringe prognostische Validität bzgl. der phonologischen Bewusstheit nachgewiesen werden. Für Kinder mit einem systematischen Unterricht war diese insgesamt höher einzuschätzen. Wimmer, Mayringer und Landerl (2000, S. 668) konnten zeigen, dass ein synthetisch-orientierter Unterricht die vorherigen Defizite in der phonologischen Bewusstheit ausgleichen kann. Dies könnte eine mögliche Begründung für die Untersuchungsergebnisse sein.

Die Autoren schlussfolgern, dass die individuellen Ergebnisse des BISC mit Vorsicht genossen werden sollten und unter anderem auch die Art des Erstleseunterrichts mit einbezogen werden muss. Sie würden das BISC als erstes grobes Screening noch empfehlen, es würden sich daraus jedoch keine Fördergruppen ableiten lassen, da mindestens die Hälfte der Risikokinder vermutlich übersehen wird. Marx und Weber (2006, S. 258) konstatieren, dass für ein vorschulisches Verfahren im Allgemeinen eine Sensitivität oder Prädiktortrefferquote über 50 % nicht realistisch ist, da schon bei der Klassifikation der Rechtschreibleistung über ein Jahr hinweg kaum mehr als die Hälfte des untersten Quartils stabil bleibt (vgl. Schneider, Stefanek & Dotzler 1997, S. 113ff.).

In der ersten Studie zum Bielefelder Screening wird zudem auch nicht deutlich gemacht, ob in der Stichprobe auch Kinder mit besonderen Risikofaktoren wie Sprachentwicklungsstörungen, mangelnden Deutschkenntnisse usw. enthalten waren. Schnitzler (2008, S. 89) vermutet, dass sich diese zusätzlichen Faktoren auch auf die Leistungen der Kinder auswirken können und das BISC vielleicht nicht ausreichend Teilbereiche der Vorläuferfertigkeiten des Lesens und Rechtschreibens erfasst. Waligora (2003, S. 747) vermisst ebenfalls Hinweise auf den Einfluss von sozioökonomischem Hintergrund und geschlechtsspezifischen Daten der Stichprobe.

Zusammenfassend müssen auch Marx und Weber (2006, S. 256) feststellen, dass das Bielefelder Screening trotz der einzelnen Kritikpunkte das bisher am besten evaluierte Screening zur Vorhersage von Lese-Rechtschreibschwierigkeiten im deutschen Sprachraum ist.

Brügelmann (2005, S. 160) zweifelt dagegen die Konstruktvalidität des Bielefelder Screenings an:

> *„'Risikokinder' haben meist nicht nur einen hohen BISC-Index, sondern auch andere Schwächen. Damit stellt sich erneut die Frage, ob der BISC die 'Ursache' von Lernschwierigkeiten erfasst oder lediglich Indikator für Schwierigkeiten ist."*

Dennoch vermutet er, dass die phonologischen Leistungen gute Prädiktoren für LRS sind, sich der Einfluss jedoch auf die Anfangsphase beschränkt und andere Voraussetzungen ähnlich wichtig sein könnten (vgl. Brügelmann 2005, S. 165). Der Frage nach den Ursachen der Schwierigkeiten bei der richtigen Klassifikation von Risiko- und Nicht-Risikokindern als Prognose von Lese- und Rechtschreibfähigkeiten soll im nächsten Kapitel nachgegangen werden.

II. Ursachen der Klassifikationsproblematik

Im ersten Kapitel konnte gezeigt werden, dass die Prognose von Schwierigkeiten im Lesen und Rechtschreiben bis zum gegenwärtigen Zeitpunkt erhebliche Probleme bereitet und offenkundig kein Früherkennungsverfahren existiert, das eine sichere Vorhersage leisten kann. Damit rückt die Frage nach der Ursache für diesen Sachverhalt in den Mittelpunkt der Betrachtung.

Für die Suche nach den Gründen dieser defizitären Situation können folgende drei Aspekte in Betracht gezogen werden:

- Die Inhalte der Verfahren wirken nicht oder nur teilweise prädiktiv zur Prognose von Schwierigkeiten im Schriftspracherwerb.
- Vom Zeitpunkt der Erhebung durch das Screening bis zur Erhebung der Kriterien werden die Leistungen des Kindes durch weitere Einflussfaktoren so stark abgewandelt, dass diese die Ergebnisse der prognostischen Validität des Verfahrens verändern.
- Die methodische Vorgehensweise der Screening-Verfahren wirkt sich negativ auf die Vorhersage der schriftsprachlichen Fähigkeiten der Kinder aus.

Im zweiten Kapitel wird zunächst erörtert, welche Inhalte und Prädiktoren derzeit bei Screening-Verfahren verwendet werden und es wird ein Ausblick gegeben, welche weiteren vielleicht zu wenig Berücksichtigung finden (Kap. II.1). Im Anschluss daran wird dargestellt, welche Faktoren die Ergebnisse zwischen dem Erhebungszeitpunkt des Prädiktors und dem Erhebungszeitpunkt des Kriteriums des Weiteren beeinflussen können (Kap. II.2), um dann schließlich die Methodik der existierenden Verfahren als mögliche Ursache für die Klassifikationsproblematik zu beleuchten (Kap. II.3).

1 Inhalte der Screening-Verfahren

Der Ansatz, präventive Maßnahmen zu forcieren und eine Förderung am Entwicklungsstand des Kindes zu orientieren (vgl. Kap. I.1.2), bedingt die Notwendigkeit einer dezidierten, möglichst frühzeitigen Diagnostik auch von Lese- und Rechtschreibproblemen. Der Erfolg im Lesen und Rechtschreiben ist davon abhängig, ob das Kind eine Einsicht in die alphabetische Struktur der Sprache erlangt.

Bereits in der Forschung zur Vorhersage von Schulleistungen konnte festgestellt werden, dass keine der Variablen, die bisher untersucht wurden, eine hohe Korrelation zur Schulleistung aufweisen. Die höchsten Korrelationen konnten bei der Intelligenz, dem Vorwissen oder der Quantität oder Qualität des Unterrichts erzielt werden: $.30 \leq r \leq .50$ (vgl. Hartke & Koch 2008, S. 40). Es konnte jedoch noch keine einzelne Variable gefunden werden, mit Hilfe derer die Schulleistungen gut vorhergesagt werden konnten. Schulleistungen und damit auch Lese-Rechtschreibleistungen scheinen multifaktoriell bedingt zu sein.

Für die Ermittlung verschiedener Prädiktoren für Lese-Rechtschreibleistungen gab es bereits seit den 60er-Jahren diverse Untersuchungen. Horn und Packard (1985, S. 597ff.) stellten in einer Meta-Analyse 58 verschiedene korrelative Untersuchungen zusammen, die die Vorhersage von schriftsprachlichen Leistungen geprüft haben. Es wurden umfassende Prädiktoren erfasst, wie Grob- und Feinmotorik, emotionale Befindlichkeit, Verhalten, Sprachstand, Hirndominanz, kognitive Leistung, Aufmerksamkeit, Sensorik und Lehrerbeurteilung. Die Autoren konnten feststellen, dass fast alle Variablen einen signifikanten Zusammenhang zum Lesen aufwiesen. Die kognitiven Leistungen, die Sprache und das Verhalten wiesen den höchsten Zusammenhang auf (vgl. auch Küspert 1998, S. 62f.).

Schneider (1989, S. 161) merkt zu dieser Studie an:

> *„Das Gesamtbild dieser Analyse ist im Hinblick auf die Selektion von geeigneten Treatment-Variablen verwirrend. Es hat den Anschein, daß die meisten erfaßten Prädiktormerkmale aus unterschiedlichen Bereichen mit späteren Leseproblemen in Beziehung stehen.“*

Schneider (1989, S. 160f.) wirft der Methodik, reine korrelative Zusammenhänge zu bilden, vor, dass diese korrelativen Beziehungen zu den Lese-Rechtschreibleistungen auch genauso gut auf andere Schulleistungen zutreffen könnten. Die Ableitung von kausalen Beziehungen sei unter diesen Umständen nicht möglich. Tramontana, Hooper und Selzer (1988, S. 89ff.) beanstandeten die Meta-Analyse aufgrund einer zu großen Homogenität der Einzeluntersuchungen. Sie stellten selbst eine Analyse aus 74 Studien zusammen. Die Gemeinsamkeit der Studien bestand darin, dass die Erhebung der Maße der Prädiktoren bereits vor der Einschulung stattgefunden hatte. Sie konnten trotz einer anderen Zusammenstellung ebenfalls wie Horn und Packard (1985, S. 597ff.) nachweisen, dass die Intelligenz, die Sprache, die Buchstabenkenntnis und visuomotorische Fähigkeiten die größte

Aussagekraft für spätere Leseleistungen hatten. Küspert (1998, S. 63) merkt dazu kritisch an, dass die Autoren der beiden genannten Meta-Analysen die differentielle Validität der Prädiktoren kaum geprüft haben und wenn doch, konnte verzeichnet werden, dass die Vorhersage für das Lesen ebenso wie bspw. für das Fach Mathematik galt. Marx (1992c, S. 250f.) erläutert, dass es bei der Vielzahl an Prädiktoren statistisch zu erwarten gewesen wäre Merkmale zu finden, die mit der Leseleistung korrelieren. Er zieht es vor, theoriegeleitet Zusammenhänge zu untersuchen und auch Zusammenhänge zwischen Prädiktor und Kriterium genauer zu beleuchten, anstatt wahllos Prädiktoren zu untersuchen, die ggf. die Möglichkeit besäßen Zusammenhänge aufzuweisen. Dieser Fehler sei schon bei der Ursachensuche in der Legasthenieforschung gemacht worden.

Küspert (1998, S. 65) glaubt, dass durch oben genannte Untersuchungen die Zusammenhänge zwischen den Prädiktoren und dem Schriftspracherwerb überschätzt wurden, da die Prädiktoren häufig nur isoliert betrachtet wurden und keine Überprüfung stattfand, inwiefern die Kriterien nur spezifisch für den Schriftspracherwerb wirken oder auch noch andere schulische Kompetenzen betreffen. Zudem wird in den Untersuchungen nicht immer deutlich gemacht, inwieweit die korrelativen Beziehungen auch kausal interpretiert werden dürfen. Sie schlägt ebenso wie Marx (s. o.) eine theoriegeleitete Vorgehensweise vor. In der LOGIK-Studie überprüften die Autoren beispielsweise anhand ausgewählter Prädiktoren theoriegeleitet die Vorhersagekraft bzgl. der Lese- und Rechtschreibleistungen der ca. 170 Kinder. Die Prädiktoren wurden vor der Einschulung bzw. zu Beginn der ersten Klasse gemessen, die Kriteriumsleistungen wurden innerhalb der zweiten bis vierten Jahrgangsstufe überprüft. Schneider und Näslund (1993, S. 273ff.) überprüften die Relevanz der Prädiktorvariablen über schrittweise Regressionsanalysen und zeigten, dass der Einfluss der Prädiktoren phonologische Bewusstheit, frühe Schriftkenntnisse, Gedächtniskapazität und nonverbale kognitive Leistungen auf die späteren schriftsprachlichen Leistungen signifikant war. Prädiktiv für die Rechtschreibleistungen wirkten insbesondere die kognitiven Fähigkeiten und die sprachliche Informationsverarbeitungsgeschwindigkeit.

Anhand theoriegeleiteter Untersuchungen ließ sich also darstellen, dass bestimmte prädiktive Variablen vorrangig im Zentrum der Untersuchungen zu Prädiktoren für Lese- und Rechtschreibleistungen stehen. In den 80er Jahren konnte wiederholt nachgewiesen werden, dass die Faktoren Intelligenz, frühe Schriftsprachfertigkeiten, Gedächtnis und metaphonologische Fähigkeiten scheinbar eine gewisse Vorhersagekraft besitzen (vgl. Küspert 1998, S. 65).

Das Ziel der Forschung im Bereich von Prognostik und Prävention von Schwierigkeiten im Schriftspracherwerb ist die Identifikation von Prädiktoren für die jeweiligen Störungen (vgl. Hartke & Koch 2008, S. 45). Diese Prädiktoren bilden die Basis, auf der Präventionsmaßnahmen aufgebaut werden, sodass Lese- und Rechtschreibschwierigkeiten nicht auftreten oder deren negative Auswirkungen vermindert werden. In der Forschung wird bei der Suche nach Variablen zur Vorhersage von Lese- und Rechtschreibschwierigkeiten häufig zwischen spezifischen und unspezifischen Prädiktoren unterschieden.

Spezifische Prädiktoren sind Faktoren, die mit der abhängigen Variablen in einem kausal interpretierbaren näheren Zusammenhang stehen. Dies können verschiedene Prädiktoren sein. Bezogen auf die Lese-Rechtschreibleistung sollte für einen spezifischen Prädiktor ausgeschlossen werden, dass dieser auch andere Kriterien, wie z.B. mathematische Leistungen, erklärt.

Marx, Jansen und Skowronek (2000, S. 17) definieren spezifische Prädiktoren für die Lese-Rechtschreibleistung:

> *„Als spezifische Prädiktoren werden solche Merkmalsbereiche bezeichnet, die erstens noch nicht durch Lese- und Schreibfertigkeiten beeinflusst sind und zweitens Aufgabenstellungen umfassen, die überwiegend beim Erwerb der Schriftsprache von Bedeutung sind und weniger beim Erwerb anderer Fertigkeiten zum Einsatz kommen.“*

In der neueren Forschung wird Lesen und Schreiben im Allgemeinen als Entwicklungsprozess verstanden. Der Beginn der Entwicklung von schriftsprachspezifischen Fähigkeiten und Fertigkeiten liegt weit vor dem Schulbeginn. Grundlagen werden von den Kindern bereits vor der Einschulung erlernt, die somit nicht die „Stunde Null“ des Schriftspracherwerbs darstellt.
Es gibt unterschiedliche Prozesse, die im Zusammenhang mit dem Lesen und Schreiben vorkommen oder während des Lernprozesses erworben werden. Dies sind spezifische Prozesse, die nur im Zusammenhang mit dem Schriftspracherwerb auftreten und die größtenteils auch nur für eine bestimmte Schrift oder Sprache spezifisch sind. Der Schriftspracherwerb hat jedoch auch einen Einfluss auf andere kognitive Prozesse, sowie diese auch die Entwicklung des Schriftspracherwerbs fördern können.

Es werden Variablen gesucht, die als spezifisch-qualitative Aussagen eine Ursache für Lese-Rechtschreibschwierigkeiten sein können, oder Variablen, die zumindest in einem

spezifischen Zusammenhang mit dem Auftreten dieser Schwierigkeit stehen (vgl. Hartke & Koch 2008, S. 46f.). Das Vorhandensein kausaler Aussagen ist jedoch eine wichtige Grundlage, dem bloßen Auftreten von Korrelationen mangelt es an relevanter Aussagekraft (vgl. Aye 1982, S. 335). Ebenso wie sich die Korrelation zwischen der Geburtenrate und dem Aufkommen von Störchen in einem kleinen Dorf in Schweden nicht in Zusammenhang bringen lässt, sollten die Kausalitätsmodelle der Lese-Rechtschreibleistung auch bei der Forschung zur Lese-Rechtschreibschwäche mit einbezogen werden. Zur Identifikation von spezifischen Vorhersagevariablen fordert Schneider (1989, S. 161) speziell für Korrelationsstudien folgende drei Aspekte:

- Die Prädiktormerkmale sollten theoriegeleitet ausgewählt werden.
- Es sollte methodisch geklärt werden, ob es sich um kausale oder bloß korrelative Verknüpfungen handelt.
- Bzgl. der Spezifität zu späteren Lese-Rechtschreibleistungen sollten zusätzlich auch andere Variablen (z.B. mathematische Leistungen) kontrolliert werden.

Unspezifische Prädiktoren dagegen sind Fähigkeiten oder Merkmalsbereiche, die sich im Vergleich zu spezifischen Prädiktoren gleichermaßen auf andere Kriterien, z. B. andere Schulleistungen auswirken können (vgl. Abb. 7, die Auswirkungen von unspezifischen Prädiktoren auf Rechenleistung und Schriftsprachkompetenz).

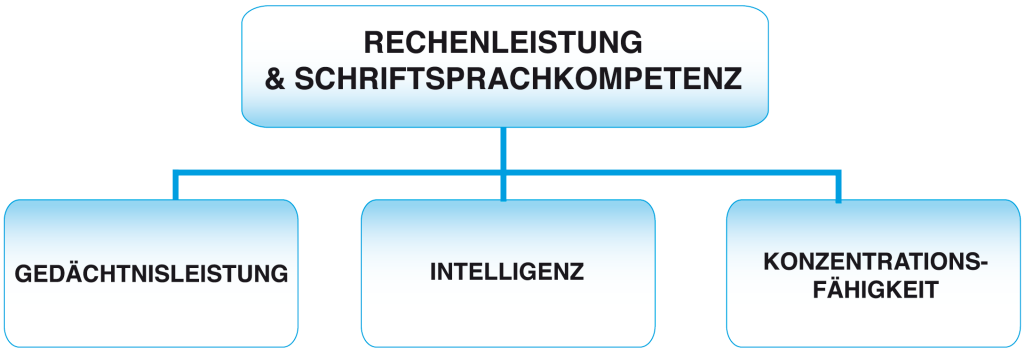


Abb. 7: Gemeinsame kognitive Prädiktoren von Schriftsprachkompetenz und Rechenleistung (verändert nach Petermann 2003, S. 195)

Weinert und Schneider (1987, S. 15ff.) erfassen in ihrer Studie verschiedene Merkmale, die sowohl für die Leseleistung als auch für mathematische Leistungen relevant sind (vgl. Tab. 17). Die kausale Interpretation dieser Studie ist jedoch nicht eindeutig zu tätigen, da die Zeitintervalle zwischen der Erhebung des Prädiktors und des Kriteriums in den zusammengefassten Studien zum Teil zu kurz waren und die Korrelationen aus dieser Perspektive betrachtet als erwartet angesehen werden können (vgl. Schneider 1989, S. 161).

Tab. 17: Vorhersage von Schulleistungen nach Weinert & Schneider 1987 (angelehnt an Schneider 1989, S. 161)

	Anzahl der Studien	Vorhersagequalität			
		Lesen		Mathematik	
		signifikant	nicht signifikant	signifikant	nicht signifikant
Allgemeine Intelligenztests	20	16	4	13	3
Schulreife- bzw. Screening-Tests	53	45	8	26	5
Piaget-Aufgaben (z. B. Seriation, Zahlinvarianz)	5	4	-	5	-
Kognitive Aufgaben (z. B. phonetische Bewusstheit, schnelles Benennen)	19	19	-	10	-
Kindergärtnerinnenurteil	21	19	2	13	2
Selbstkonzept-/ Leistungsmotivation	7	6	1	6	1

Psychomotorisches Ungeschick und Schwierigkeiten in der visuellen Wahrnehmung bzw. in der visuomotorischen Koordination sowie die hyperkinetische Störung sind unspezifische Prädiktoren, die laut einiger Autoren (vgl. Warnke, Hemminger & Plume 2004, S. 7) auf eine Lese-Rechtschreib-Problematik hinweisen können. Warnke, Hemminger und Plume (ebd.) betonen jedoch, dass eine Prognose in Bezug auf eine spätere Lese-Rechtschreibstörung nicht zulässig ist, da viele dieser Kinder keine Schwierigkeiten im Schriftspracherwerb haben werden.

Ein unspezifischer Faktor ist bspw. das Hörverständnis. Bei der Diskussion über die Diskrepanzmessung (vgl. Kap. I.1.3.1) rückte Elbro (1998, S. 129ff.) immer mehr die Betrachtung des Potenzials der Kinder in den Vordergrund. Aufgrund herkömmlicher Vorgehensweise wurde mit „Potenzial“ häufig der Unterschied zwischen Lese-Rechtschreibleistungen und kognitiven Leistungen betrachtet. Elbro dagegen betrachtete auch die Diskrepanz zwischen Hör- und Leseverständnis. Auch Stanovich (1991, S. 7ff.) unterstützt diese These im Vergleich zur kognitiven Diskrepanz aufgrund einer größeren Bedeutsamkeit

bzgl. Fördermaßnahmen. Somit sollten Kinder mit geringem Hör- und Leseverständnis sowie einem zumindest durchschnittlichen IQ nicht als lese-rechtschreibschwach eingestuft werden. Elbro definiert für die Messung des Hörverständnisses drei Ebenen, die bei einer Diagnostik betrachtet werden sollten:

- Texte
- Kurze Passagen oder Sätze
- Wörter

Die verschiedenen Untersuchungen der Überprüfung der drei Ebenen führte jedoch zu uneinheitlichen Ergebnissen, sodass Elbro (vgl. Weber 2003, S. 132) bzgl. der Sinnhaftigkeit einer Diskrepanzdefinition auch auf Grundlage des Hörverständnisses skeptisch ist. Ausschließlich die Diagnostik auf Wortebene zum phonologischen Rekodieren von Wörtern (vgl. Kap. II.1.2.2) erscheint ein vielversprechender Indikator für die Leseleistung zu sein und ist wiederum den spezifischen Prädiktoren zuzuordnen.

Wendeler (1986, S. 10ff.) konnte bei einer Längsschnittstudie mit 294 Vorschulkindern feststellen, dass Tests zu unspezifischen Prädiktoren wie der visuellen Wahrnehmung, Visuomotorik, der auditiven Wahrnehmung und des Gedächtnisses nur schwach positive Korrelationen zu den Lese- und Rechtschreibfähigkeiten zwei Jahre später aufwiesen. Die Studie wurde mit 11 unterschiedlichen Prädiktoren zu den oben genannten Themen durchgeführt. Das Ergebnis der Klassifikation konnte zeigen, dass 64 % der Schüler beim Lesen und 66 % der Schüler beim Schreiben, die als Risikokinder identifiziert worden sind, später keine schwachen Leistungen zeigten. Betrachtet man die Korrelationen der elf Prädiktoren im Einzelnen, so scheinen die Prädiktoren mit näherem inhaltlichen Zusammenhang mit der Lese- und Schreibfertigkeit (z. B. Zerlegen in Silben: Schreiben $r = .46$ und Lesen $r = .37$) eine höhere Korrelation aufzuweisen (vgl. a. a. O., S. 13). Auch in einer zweiten Untersuchung mit 255 Kindern und ähnlichen Prädiktoren (Tests zu visuellen oder visuell-motorischen Fähigkeiten, Tests zu auditiven Fähigkeiten und Tests zum Gedächtnis) konnten keine positiveren Ergebnisse erzielt werden (vgl. Wendeler 1988, S. 770).

Dies lässt wiederum vermuten, wie bereits oben angedeutet, dass spezifische Prädiktoren eine bessere Vorhersagekraft haben und unspezifische Prädiktoren nur eine relativ unwichtige Voraussetzung für das Lesen und Schreiben darstellen. Dies könnte auch der Grund für die relativ gute Quote bei den richtig vorhergesagten Kindern mit einer Leserechtschreibproblematik (richtig-positiv) sein, bei denen die Grundvoraussetzung nicht erfüllt ist, jedoch bei den nicht erkannten (falsch-positiven) Kindern mit einer

Leserechtschreibproblematik die Quote zu hoch ist, da die Prädiktoren nicht spezifisch und genau genug identifiziert waren.

Aus dem Grund der möglichen besseren Prognoseleistungen von spezifischen Prädiktoren sollen diese im Folgenden im Zentrum der Diskussion stehen. Es werden ausgewählte mögliche Vorläuferfertigkeiten für das Lesen und Schreiben in Bezug auf die Aussichten der Prognose einer Lese- und Rechtschreibschwäche erörtert. Verschiedene Studien konnten zeigen, dass die phonologische Informationsverarbeitung bedeutsam zur Vorhersage von Schwierigkeiten im Lesen und Schreiben im Schulalter beiträgt (vgl. bspw. Schneider & Näslund 1999, S. 268; Wagner & Torgesen 1987, S. 202ff.; Gathercole & Baddeley 1993a, S. 131ff.; Jansen, Mannhaupt, Marx & Skowronek 2002, S. 30ff.). Insbesondere die phonologische Bewusstheit, die Arbeitsgedächtniskapazität sowie die Abrufgeschwindigkeit von Informationen aus dem semantischen Lexikon haben sich als bedeutende Prädiktoren für die Lese- und Rechtschreibleistungen herausgestellt (vgl. Barth & Gomm 2008b, S. 7ff.). Auf die phonologische Informationsverarbeitung und insbesondere die phonologische Bewusstheit soll im Rahmen der spezifischen Prädiktoren näher eingegangen werden. Es werden einige exemplarische Studien aus der Fülle der vorliegenden Untersuchungen herausgegriffen, um einen Überblick zu vermitteln. Im Anschluss werden als spezifische Prädiktoren noch die Buchstabenkenntnis sowie die Sprachkompetenz vorgestellt. Aufgrund der Verwendung vieler englischsprachiger Studien wird vorab die mögliche Übertragbarkeit von englischsprachigen Studien in einen deutschsprachigen Kontext erörtert, da sich die unterschiedlichen Sprachen speziell auf die Untersuchungsergebnisse auswirken könnten und dies bei einer Übernahme der Befunde Berücksichtigung finden sollte.

1.1 Zur Übertragbarkeit englischer Studien auf den deutschen Sprachraum

In vielen hier angeführten Studien wird häufig auf Literatur aus dem englischen, skandinavischen oder anglo-amerikanischen Raum zurückgegriffen (vgl. bspw. Ehri 1986, S. 121ff.; Goswami 1993, S. 443ff.; Stanovich 1986, S. 360ff.). Die Entwicklung der schriftsprachlichen Fähigkeiten orientiert sich am jeweiligen orthografischen System und am linguistischen Umfeld der Schüler. Die deutsche Sprache ist im Gegensatz zum Englischen regelhafter. Die Graphem-Phonem-Korrespondenzen sind eindeutiger als im Englischen, auch wenn es sich nicht um eine Eins-zu-Eins-Zuordnung handelt (vgl. Küspert 1998, S. 108f.). Somit bedarf es im Folgenden einer Überprüfung der Übertragbarkeit der Ergebnisse englischsprachiger Untersuchungen auf den deutschen Sprachraum.

Die Autoren Wimmer, Hartl und Moser (1990, S. 136) nehmen an, dass bei einer regulären Sprache wie der deutschen die logographische Stufe wahrscheinlich direkt mit der alphabetischen Strategie ersetzt wird. Die Autoren vermuten, dass eine logographische Phase in der deutschen Sprache somit nicht zwingend auftreten muss. Der Schriftspracherwerb der englischen Sprache ist nicht direkt mit dem Erwerb der Schriftsprache im deutschen Sprachraum vergleichbar. Die irreguläre englische Orthografie erschwert den Erstklässlern das Lesen. Die basale Lesekompetenz stellt sich dadurch ungefähr ein Jahr später ein als bei deutschen Schülern (vgl. Hasselhorn, Marx & Schneider 2008, S. 2; Landerl & Wimmer 1998, S. 82ff.). Es ist möglich, dass die Schwierigkeit des Schriftspracherwerbs in Abhängigkeit von der orthografischen Konsistenz des zu erlernenden orthografischen Systems steht. Seymour, Aro und Erskine (2003, S. 143ff.) konnten in ihrer Studie herausfinden, dass der Erwerb der Dekodierungsfähigkeit und der allgemeinen Worterkennung im Englischen im Vergleich zum Deutschen verzögert auftritt. Dies scheint zum einen allgemein an der silbischen Komplexität germanischer Sprachen zum anderen an dem intransparenten orthografischen System zu liegen. Die Autoren vermuten, dass Dekodierstrategien und phonologische Komponenten in transparenteren Systemen von etwas geringerer Bedeutung sind.

Wimmer und Goswami (1994, S. 91ff.) konnten vermerken, dass sich sieben-, acht- und neunjährige englische und österreichische Schüler beim Erlernen von Ziffern und Zahlwörtern zwar nicht in der Geschwindigkeit oder in der Fehleranzahl unterschieden, jedoch waren die deutschsprachigen Schüler den englischsprachigen Schülern beim Lesen von Pseudowörtern, die von Zahlwörtern abgeleitet waren, deutlich überlegen. Die Autoren vermuten, dass deutschsprachige Kinder die Graphem-Phonem-Korrespondenzen entschlüsselten, während sich die englischsprachigen Schüler auf die Strategie der direkten Worterkennung verließen. Aro und Wimmer (2003, S. 621ff.) konnten dies in einer weiteren Studie mit Erst- bis Viertklässlern bestätigen. Wimmer, Landerl und Schneider (1994, S. 469ff.) erhielten im deutschsprachigen Raum den Befund, dass die orthografischen Einheiten Onset und Rime für den Anfang des Schriftspracherwerbs nicht relevant sind, sondern erst später mit den Leseleistungen im neunten oder zehnten Lebensjahr korrelieren. Goswami (1993, S. 443ff.) konnte dagegen für englischsprachige Kinder herausfinden, dass die Einheiten Onset und Rime bzgl. der Analogiebildung bereits zu Beginn des Schriftspracherwerbs bedeutsam sind. In der englischen Sprache scheinen aufgrund der Unregelmäßigkeiten in den Graphem-Phonem-Korrespondenzen die größeren Einheiten bereits zu Schulbeginn wichtig zu sein, während deutsche Kinder mit der alphabetischen Strategie zu Schulbeginn größtenteils auskommen. Des Weiteren konnte in einer Reihe von Studien (vgl. Kap. I.1.2) sowohl mit deutsch- als auch mit englischsprachigen Kindern

gezeigt werden, dass die phonologische Informationsverarbeitung in beiden Sprachräumen von Relevanz für den Schriftspracherwerb zeugt. Phonologische Schwierigkeiten scheinen sich aufgrund der Transparenz der deutschen Sprache jedoch im deutschsprachigen Raum nicht so stark auszuwirken wie im Englischen (vgl. Wimmer, Landerl, Linortner & Hummer 1991, S. 244).

Auch das Einschulungsalter ist gerade in Bezug auf Studien zur phonologischen Bewusstheit im Kindergarten und im Anfangsunterricht der ersten Klasse von Relevanz. Schon in Deutschland ist es in den einzelnen Bundesländern sehr unterschiedlich. In Schleswig-Holstein werden die Kinder mit sechs oder sieben Jahren eingeschult, während in Berlin für die Kinder in dem Jahr, in dem sie sechs Jahre alt werden, also mit fünf oder sechs Jahren die Schulpflicht beginnt. In England dagegen liegt das Einschulungsalter der Schüler bei fünf Jahren (vgl. Küspert 1998, S. 109). Vermutlich verfügen die älteren Kinder bei der Einschulung über ein größeres phonologisches Wissen und über ein Arbeitsgedächtnis mit größeren Kapazitäten.

Insgesamt lässt sich zusammenfassen, dass sich die Ergebnisse aus englischsprachigen Studien größtenteils auf die deutsche Sprache übertragen lassen, oder diese zumindest Hinweise für die Forschung liefern können, die dann im deutschsprachigen Raum nochmals bestätigt werden müssen. Die Bewusstheit der sprachlichen Unterschiede und auch der unterschiedlichen strukturellen Voraussetzungen im Schulsystem (bspw. bzgl. des Einschulungsalters) sollte jedoch vorhanden sein, die Studien können aus dem Englischen keinesfalls direkt auf den deutschen Sprachraum übertragen werden.
Berücksichtigt werden sollte, dass die logographemische Phase im deutschen Sprachraum keine hohe Relevanz besitzt und auch die Zeit und die Methodik der Beschulung eine Rolle spielen könnte.

Graf (1994, S. 225) fasst anhand ihrer Analyse der Literatur zusammen:

> *„Insgesamt gilt, dass die früher ausführlich beschriebenen Ansätze im englischen Sprachraum, die m.E. mit der Lese-Rechtschreib-Schwäche verbunden sind, weitgehend auf den deutschen Sprachraum übertragen werden können. Dies trifft insbesondere auf die theoretischen Ansätze, wie jenen der kognitiven Neuropsychologie, der Entwicklungsmodelle des Lesens und Schreibens und des Arbeitsgedächtnisses zu.“*

1.2 Phonologische Informationsverarbeitung

Zum Erlernen der Schriftsprache muss ein Kind zunächst Einsicht in die Struktur unserer Sprache und unserer Schriftsprache erlangen. Dazu wird die gesprochene Sprache abstrahiert und die formalen Aspekte der Sprache rücken in den Vordergrund. Die wissenschaftlichen Untersuchungen der letzten Jahre konnten zeigen, dass das Erlernen des Lesens und Schreibens immer mehr von Sprachwahrnehmungs- und Sprachverarbeitungsprozessen abhängig ist (vgl. Barth & Gomm 2008a, S. 4f.). Die Interaktivität beschreibt das Zusammenspiel der einzelnen Prozesse, aber auch die Interaktion zwischen dem Lesen und Schreiben und dem erlernten Wissen, der Sprachkompetenz oder anderen Fähigkeiten, die Kinder in den Erwerb der Schriftsprache mit einbeziehen (vgl. Scheerer-Neumann 1998a, S. 32f.). Daher werden in diesem Kapitel umfänglich Erkenntnisse zur phonologischen Informationsverarbeitung beschrieben.

Vor Beginn des Schriftspracherwerbs können Kinder lediglich mit den groben lautlichen Strukturen unserer Sprache umgehen. Es scheint eine Wechselwirkung zwischen dem Erlernen von lautlichen Strukturen und der Weiterentwicklung der Sprache des Kindes zu geben. Lesen und Schreiben benötigen einige Fähigkeiten der gesprochenen Sprache, diese Prozesse werden andererseits aber auch im Verlaufe des Schriftspracherwerbs in hohem Maße entwickelt oder optimiert (vgl. Mannhaupt 2006, S. 101f.).
Adams (1994, S. 293ff.) konnte herausfinden, dass ungefähr ein Drittel der Erstklässler die phonologische Struktur von Wörtern noch nicht verstanden haben und damit die alphabetische Strategie noch nicht anwenden können. Fowler und Scarborough (1993) konnten zeigen, dass Schüler mit schwachen Leistungen im Lesen und Rechtschreiben ähnliche Schwierigkeiten im Bereich der metalinguistischen Fähigkeiten zeigen wie die Schüler im Erstunterricht.

Die psychologische Forschung hat herausgefunden, dass neben den unspezifischen Vorläufermerkmalen wie Sprache, Kognition und den auditiven- oder visuellen Fähigkeiten auch spezifische Fähigkeiten eine Rolle spielen. Dies hat zur Identifikation der sogenannten Vorläuferfertigkeiten geführt. Diese spezifischen Merkmale sind neben den meta-linguistischen Fähigkeiten (Syntax, Grammatik, Semantik), insbesondere in dem Begriff der phonologischen Informationsverarbeitung enthalten (vgl. Barth & Gomm 2008b, S. 10).

Die phonologische Informationsverarbeitung umfasst folgende Komponenten (vgl. Marx 1997, S. 86ff.; Wagner & Torgesen 1987, S. 192f.):

- die phonologische Bewusstheit (phonological awareness),
- das phonologische Rekodieren beim Zugriff auf das semantische Lexikon (phonological recoding in lexical access),
- das phonologische Rekodieren im Arbeitsgedächtnis (phonetic recoding in working memory)

Wagner und Torgesen (1987, S. 192f.) führten den Begriff „phonologische Informationsverarbeitung (phonological processing)“ als Sammelbegriff für die Anwendung von metalinguistischen Informationen der Lautstruktur zur Analyse der gesprochenen und geschriebenen Sprache ein.

Die *phonologische Bewusstheit* ist eine spezifische kognitive Fähigkeit. Sie bezieht sich auf die Fähigkeiten der Kinder phonologisch-strukturelle Informationen der gesprochenen Sprache zu entnehmen und ohne Rücksicht auf die Bedeutung des Gesprochenen eine Einsicht in die Lautstruktur der gesprochenen Sprache zu erhalten, diese zu durchgliedern und mit ihr umzugehen (vgl. i. F. Kap. II.1.2.1).

Das *phonologische Rekodieren beim Zugriff auf das semantische Lexikon* umfasst die sprachgebundene Informationsverarbeitungsgeschwindigkeit, in der die relevanten Informationen zu einem soeben rekodierten Wort aus dem semantischen Lexikon abgerufen werden können. Über die Rekodierung lautlicher Symbole werden Graphem-Phonem-Korrespondenzen ermittelt, die einen Zugang zum semantischen Lexikon verschaffen (vgl. Stanovich 1982, S. 488ff.; Barth & Gomm 2008b, S. 12f.). Dabei werden die Buchstaben mit den entsprechenden Lauten verknüpft (vgl. i. F. Kap. II.1.2.2).

Das *phonetische Rekodieren im Arbeitsgedächtnis* als weitere Vorläuferfertigkeit für einen erfolgreichen Schriftspracherwerb beschreibt die Fähigkeit phonetische Informationen im Arbeitsgedächtnis so lange präsent zu halten, bis sie wieder abgerufen werden. Beim Lesen unbekannter Wörter werden beispielsweise die zu den einzelnen Buchstaben gehörenden Laute im Kurzzeitgedächtnis präsent gehalten, bis die einzelnen Informationen beim Zusammenziehen der Einzellaute wieder abgerufen werden (Rekodieren) – (vgl. i. F. Kap. II.1.2.3).

Die Erfassungsmethoden für die einzelnen Komponenten der phonologischen Informationsverarbeitung sind in Tab. 18 mit Beispielen belegt.

Tab. 18: Komponenten phonologischer Informationsverarbeitung und deren mögliche Erfassungsmethoden (nach Walter 2007a, S. 479, Schneider 1989, S. 162)

Komponenten phonologischer Informationsverarbeitung	Mögliche Erfassungsmethoden
Phonologische Bewusstheit	Reimaufgaben, Alliterationsaufgaben, Silbenklatschen, Laut-zu-Wort-Zuordnungsaufgaben
Phonologische Rekodierung im lexikalischen Zugriff	Schnelles Benennen von Worten, Farben, Objekten, Lesen von Pseudowörtern
Phonetische Rekodierung im Arbeitsgedächtnis	Artikulationsgeschwindigkeit, Gedächtnisspanne für Bilder, Gedächtnisspanne für Wörter

Für alle drei Komponenten der phonologischen Informationsverarbeitung ließen sich signifikante Zusammenhänge zu den Lese- und Rechtschreibfähigkeiten nachweisen (vgl. Gathercole & Baddeley 1993a, S. 131ff.; Skowronek & Marx 1989b, S. 45ff.).

Im Folgenden werden die einzelnen Aspekte der phonologischen Informationsverarbeitung und ihr Einfluss auf den Erwerb der Schriftsprache dargestellt und analysiert. Es wird auch in Augenschein genommen, inwieweit die phonologischen Fähigkeiten eine Voraussetzung oder eine Konsequenz des Schriftspracherwerbs darstellen. Dies könnte Schlüsse auf die Gründe der Klassifikationsproblematik der Früherkennungsverfahren von Lese-Rechtschreibschwäche zulassen, da ein Großteil der Screening-Verfahren phonologische Fähigkeiten als zentrale Komponente berücksichtigt.

1.2.1 Phonologische Bewusstheit

Die phonologische Bewusstheit ist das am häufigsten untersuchte Konstrukt der phonologischen Informationsverarbeitung. Die Definition in den einzelnen Studien differiert (vgl. zusammenfassend Küspert 1998, S. 68).

Mattingly (zit. nach Mannhaupt & Jansen 1989, S. 50) definiert beispielsweise die phonologische Bewusstheit als

> *„[...] das Phänomen, daß Sprache aus distinkten lautlichen Einheiten bestehend wahrgenommen und mit diesen Segmenten analytisch/synthetisch umgegangen werden kann."*

In den 90er-Jahren wurde das Konzept der phonologischen Bewusstheit als Konstrukt der Forschung im Bereich des Schriftspracherwerbs insbesondere von Skowronek und Marx (1989a, S. 268ff.) publik gemacht.

Skowronek und Marx (1989b, S. 43f.) unterscheiden dabei in zwei Fähigkeiten:

- die phonologische Bewusstheit im weiteren Sinn,
- die phonologische Bewusstheit im engeren Sinn.

Die Unterscheidung bringt den zeitlichen Abstand in der Entwicklung, den Schriftsprachfähigkeiten und den natürlichen Zusammenhang der beiden Fähigkeiten in Verbindung. Die Definition von Skowronek und Marx (a. a. O., S. 40) wird im Folgenden als Grundlage gewählt.

Die *phonologische Bewusstheit im weiteren Sinn* ist die Fähigkeit Reime zu erfassen und zu bilden, Silben zu erkennen und den Lautfluss der Sprache in größere Einheiten (Sätze, Wörter, …) zu segmentieren. Diese Fähigkeit wird normalerweise in der Vorschulzeit entwickelt.

Die *phonologische Bewusstheit im engeren Sinn* ist die Fähigkeit Laute in Wörtern oder Silben zu charakterisieren und zu unterscheiden. Diese Fähigkeit dagegen entwickelt sich normalerweise erst im Laufe des Schriftspracherwerbs und wird hauptsächlich im ersten Schuljahr ausgebildet.

In der Zusammenstellung von Meta- und Meganalysen kann Walter (2002a, S. 40f.) feststellen, dass sich das Training der phonologischen Bewusstheit auf einer „Hitliste von Interventionsformen im sonderpädagogischen Feld“ mit einer hohen Effektstärke von .86 auf Platz 4 befinden. Die Befunde basieren auf der Meta-Analyse von Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh und Shanahan (2001). Barth und Gomm (2008a, S. 27f.) konnten anhand einer Stichprobe von 450 Erstklässlern zeigen, dass die phonologische Bewusstheit im engeren und im weiteren Sinne eine mittlere Korrelation von $r = .585$ und eine hohe statistische Signifikanz aufweist. Bezüglich der bewussten Wahrnehmung der phonologischen Bewusstheit kann derzeit keine Aussage getroffen werden. Die Bewusstheit der metalinguistischen Prozesse ist umstritten (vgl. Schnitzler 2008, S. 6f.). Stackhouse und Wells (1997, zit. nach Schnitzler 2008, S. 15ff.) sehen in ihrem konnektionistischen Modell die phonologische Bewusstheit als ein Teil der zentralen Sprachverarbeitung, die mit Aufmerksamkeits- und Gedächtniskomponenten in Verbindung steht. Die phonologische

Bewusstheit steht in direkter Verbindung zur gesprochenen wie auch zur geschriebenen Sprache (vgl. Abb. 8). Durch dieses Modell wird anschaulich gezeigt, dass Erwachsene Schwierigkeiten in der Schriftsprache aufgrund phonologischer Defizite mit anderen Komponenten wie z. B. der Grammatik oder des Wortschatzes scheinbar kompensieren können.

In der Untersuchung von Svensson und Jacobson (2005, S.9) konnte die Stabilität der phonologischen Defizite gezeigt werden. Die 30 Schüler mit Lese- und Rechtschreibschwierigkeiten wiesen von der zweiten Klasse an bis zu zehn Jahre später immer noch phonologische Defizite auf.

Zur methodischen Überprüfung des Zusammenhangs zwischen schriftsprachlichen Fähigkeiten und der phonologischen Bewusstheit definieren Wagner und Torgesen (1987, S. 193) drei unterschiedliche Ansätze:

- Korrelationsstudien (vgl. i. F. Kap. II.1.2.1.1): In Längsschnittstudien werden korrelative Zusammenhänge erfasst, die anhand von Kausalmodellen geprüft werden können. In den Kausalmodellen müssen alle in der Untersuchung genutzten relevanten Variablen erfasst werden. Dies ist gleichzeitig auch die Schwierigkeit der Methodik, da es sein könnte, dass eine Variable nicht erfasst worden ist.
- Experimentelle Trainingsstudien (vgl. i. F. Kap. II.1.2.1.2): Die phonologische Bewusstheit oder das Lesen und Rechtschreiben wird trainiert und die Entwicklung der nicht geförderten Fähigkeit wird beobachtet. Derlei Studien bieten einen zuverlässigeren Nachweis über die Kausalität der Komponenten, andererseits ist bei Laborexperimenten tendenziell eine geringere ökologische Validität vorhanden (vgl. Küspert 1998, S. 71).
- Gruppenvergleiche (vgl. i. F. Kap. II.1.2.1.3): In Gruppenvergleichen werden i. d. R. Erwachsene und/oder Kinder mit einer Lese-Rechtschreibschwäche untersucht und mit einer durchschnittlichen Gruppe verglichen. In diesen Studien ist jedoch nicht zu erkennen, ob die Befunde eine Ursache oder die Folge der Lese-Rechtschreibschwäche sind.

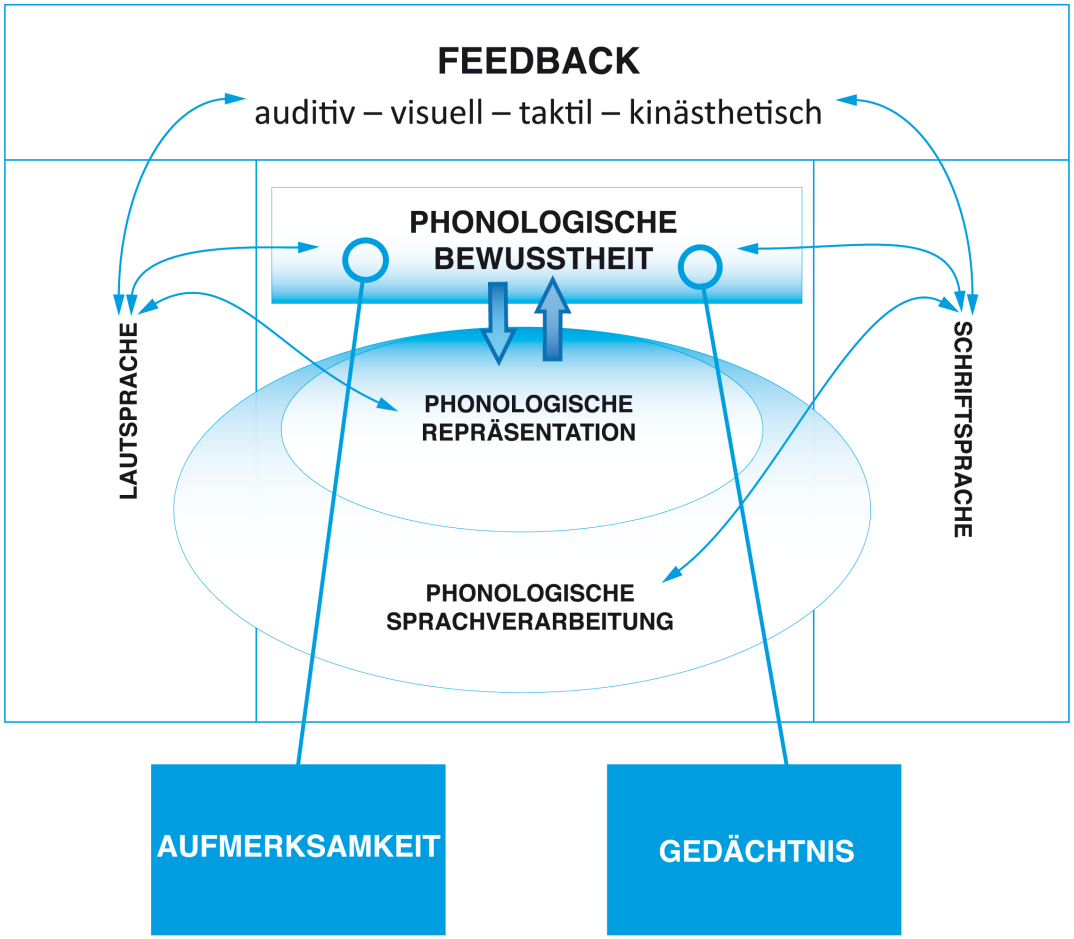


Abb. 8: Konnektionistisches Modell der phonologischen Bewusstheit von Stackhouse und Wells 1997 (nach Schnitzler 2008, S. 15)

Eine isolierte Anwendung einer der Methoden zur Interpretation der Forschungslage ist nicht ausreichend. Küspert (1998, S. 71) betont bzgl. der Korrelations- und Trainingsstudien:

> *„Eine Kombination beider Methoden, deren Nachteile teilweise komplementär sind, kann zur Klärung der angesprochenen Fragestellung schon wesentlich mehr beitragen.“*

Im Folgenden werden Ergebnisse aus Korrelations- und Trainingsstudien sowie aus Gruppenvergleichen vorgestellt. Im Anschluss folgt eine Übersicht über den Stand der Forschung bzgl. der Frage: Sind ausgeprägte Fähigkeiten in der phonologischen Bewusstheit eine Folge oder eine Voraussetzung für einen erfolgreichen Schriftspracherwerb?

1.2.1.1 Korrelationsstudien zur phonologischen Bewusstheit

In Korrelationsstudien wurde der Zusammenhang zwischen der phonologischen Bewusstheit und dem Schriftspracherwerb überprüft, ohne dass ein vorheriges Training der Schriftsprache oder der phonologischen Bewusstheit stattgefunden hat. Bei Korrelationsstudien muss bezüglich der Interpretation darauf geachtet werden, dass die Zusammenhänge nicht per se kausal interpretiert werden, sondern ihnen erst nach theoretischer Fundierung ein inhaltlicher Wert zugemessen wird. Die Zusammenhänge müssen also anhand von Kausalmodellen überprüft werden (vgl. Weber 2003, S. 53). Diese Vorgehensweise ist stark von der Auswahl der überprüften Variablen geprägt und es besteht die Gefahr, dass Zusammenhänge als relevant beurteilt werden, die eigentlich aus einer nicht erfassten dritten Variable resultieren.

Zunächst werden einige Studien zur phonologischen Bewusstheit vorgestellt, die vorrangig aus skandinavischer oder US-amerikanischer Forschung stammen, im Anschluss daran werden dann vergleichbare und weiterführende Studien aus dem deutschen Sprachraum thematisiert.

Ergebnisse von Korrelationsstudien aus dem nicht-deutschsprachigen Raum

Vellutino und Scanlon (1987, S. 321ff.) konnten in ihrer Studie mit 295 Kindern im Vorschulalter nachweisen, dass die phonologische Bewusstheit den Erfolg im Lesen vorhersagen kann. Diese signifikanten Korrelationen blieben auch bei Kontrolle der Intelligenz als Variable bestehen (vgl. Tab. 19).

Bradley und Bryant (1983, S. 419) konnten als eine der ersten in einer Längsschnittstudie mit 368 Kindergartenkindern nachweisen, dass Onset-Reim-Fähigkeiten im Alter von vier bis fünf Jahren und schriftsprachliche Fähigkeiten unabhängig von Gedächtnis und Intelligenz signifikant nach vier Jahren noch in einem korrelativen Zusammenhang standen. Die Lautkategorisierungsaufgabe der Autoren korrelierte mit den Tests zur Lesefähigkeit mit $r = .52$ und $r = .57$, zu der Rechtschreibleistung mit $r = .48$ und zum Mathematiktest mit $r = .33$. Als Prädiktor (unter Konstanthaltung von IQ, Alter und Gedächtnisspanne) klärte damit die Lautkategorisierung als ein Faktor der phonologischen Bewusstheit beim Lesen 4 bis 10 % der Varianz ($p < .01$), bei der Rechtschreibung 6 bis 8 % der Varianz ($p < .01$) und bei den mathematischen Fähigkeiten 1 bis 4 % der Varianz (n. s.) auf. Somit konnte hier bereits die phonologische Bewusstheit als ein spezifischer Faktor zur Vorhersage der schriftsprachlichen Leistungen identifiziert werden, da die mathematischen Fähigkeiten davon nicht beeinflusst wurden. Küspert (1998, S. 72) merkt jedoch an, dass die prädiktive

Valenz der Lautkategorisierung an Relevanz verloren hätte, wenn die Autoren anstatt nur einer Variable noch weitere Variablen der phonologischen Bewusstheit mit berücksichtigt hätten. Auch MacLean, Bryant und Bradley (1987, S. 255ff.) in Bezug auf die Onset-Reim-Ebene und MacDonald und Cornwell (1995, S. 523ff.) in Bezug auf die Phonemebene konnten diese Zusammenhänge zu schriftsprachlichen Fähigkeiten in ihren Studien bestätigen.

Tab. 19: Einfache und Partialkorrelationen zwischen Prädiktormaßen zur phonologischen Sprachentwicklung und Leseleistungen gegen Ende der ersten und zweiten Klassenstufe nach Vellutino & Scanlon 1987 (angelehnt an Schneider 1989, S. 164)

Prädiktor	Lesetest 1.Klasse		Lesetest 2.Klasse	
	Einfache Korrelation	Partial-Korrelation	Einfache Korrelation	Partial-Korrelation
IQ	.34	-	.39	-
Phonologische Maße				
Buchstabennamen	.27	.20	.38	.30
Reimaufgabe	.27	.15	.30	.19
Laut-Buchstaben-Zuordnung	.46	.36	.48	.37
Buchstaben-Laut-Zuordnung (Konsonanten)	.50	.42	.44	.35
Buchstaben-Laut-Zuordnung (Vokale)	.39	.34	.37	.35
Substitution des Anfangskonsonanten	.53	.46	.49	.41
Sichtwörter	.60	.57	.42	.37

Anmerkung: IQ wurde kontrolliert.

Lundberg, Olofsson und Wall (1980, S. 159ff.) maßen in ihrer Studie mit 143 Vorschulkindern viele verschiedene Prädiktoren, sowohl der phonologischen Bewusstheit im weiteren Sinne (Reimen, Silben segmentieren, Silben synthetisieren), als auch der phonologischen Bewusstheit im engeren Sinne (Phoneme segmentieren, Phoneme synthetisieren, Phoneme vertauschen, Stand des Phonems im Wort erkennen). Im Kindergarten wurde ein

Screening zum Lesen durchgeführt, zu Ende des ersten und zu Beginn des zweiten Schuljahres wurden zusätzlich noch die kognitiven Leistungen und die Lese- und Rechtschreibleistungen der Kinder erfasst. Die Leistungen zur phonologischen Bewusstheit im engeren Sinne korrelierten höher mit der späteren Leseleistung als die Leistungen zur phonologischen Bewusstheit im weiteren Sinne. Die Autoren konnten damit die hohe Relevanz der phonologischen Bewusstheit für den Schriftspracherwerb aufdecken, überprüften jedoch nicht die alternative Kausalrichtung. Wagner und Torgesen (1987, S. 199f.) reanalysierten daraufhin die Daten und kamen zu dem Schluss, dass bei Konstanthaltung der anfänglichen Lesefähigkeit der Kinder die Korrelationen zwischen phonologischer Bewusstheit und der Leseleistung nicht signifikant waren (vgl. zusammenfassend Küspert 1998, S. 73).

In den Studien von Mann (1984, S. 117ff.; 1986, S. 69ff.; 1991, S. 194ff.; 1993, S. 259ff.; Mann & Libermann 1984, S. 592) konnte ebenfalls gezeigt werden, dass die phonologische Bewusstheit in Bezug auf sprachliche Einheiten in verschiedenen Größen (Phonem, Reim, Silbe) im Kindergarten die späteren Fähigkeiten im Lesen vorhersagt. Analoges vermerkten Muter und Snowling (1998, S. 332f.), die anhand einer Stichprobe von N = 34 Kindern in ihrer Längsschnittstudie feststellen konnten, dass die phonologische Bewusstheit im engeren Sinne sowohl als kurzfristiger Prädiktor (erstes Schuljahr) als auch als langfristiger spezifischer Prädiktor (mit neun Jahren) sehr gute Vorhersagen für die Lesegenauigkeit bietet. Die Stichprobe ist jedoch nicht als repräsentativ anzusehen. Die Fähigkeit Reime zu identifizieren stellte sich dagegen nicht als Prädiktor für die Lesefähigkeit im Alter von neun Jahren heraus.

Ergebnisse von Korrelationsstudien aus dem deutschsprachigen Raum

Die Studien aus dem angloamerikanischen bzw. nicht-deutschsprachigen Bereich sind nicht ohne Weiteres auf den deutschen Sprachraum zu übertragen, da die deutsche Sprache wesentlich lautgetreuer ist als die englische Sprache und deutsche Kinder daher die alphabetische Strategie schneller und erfolgreicher lernen können (vgl. Aro & Wimmer 2003, S. 621ff. und Kap. II.1.1). Umfassende deutsche Studien konnten die Ergebnisse englischsprachiger Studien jedoch bezüglich der Relevanz der phonologischen Bewusstheit bestätigen: Mehrere Langzeitstudien wie beispielsweise die LOGIK-Studie (vgl. Näslund & Schneider 1993, S. 295ff.; 1996, S. 30ff.; Schneider 1989, S. 157ff.; Schneider & Näslund 1992, S. 256ff.), die Untersuchungen im deutsch-österreichischen Raum (vgl. Landerl, Linortner & Wimmer 1992; Landerl & Wimmer 1994; Wimmer, Landerl & Schneider 1994), die Studien der Bielefelder Forschungsgruppe zur Früherkennung von Risiken der Lese-Rechtschreibschwäche (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002; Skowronek

& Marx 1989a) oder die Studien der Nürnberger Gruppe (vgl. Einsiedler, Frank, Kirschhock, Martschinke & Treinies 2002; Kirschhock 2003) konnten bestätigen, dass sich die phonologische Bewusstheit im weiteren und im engeren Sinne zur Vorhersage von Lese- und Rechtschreibleistungen im Grundschulalter eignet. Steinbrink, Schwanda und Vogt (2008, S. 649) konnten feststellen, dass die phonologische Bewusstheit immerhin 16 % der Varianz im Rechtschreiben und 19 % der Varianz im Lesen aufklärt.

In der LOGIK-Studie wurden die Leistungen der Kinder vom vierten Lebensjahr an verfolgt. Die Kenntnisse zur phonologischen Bewusstheit wurden am Ende des Kindergartens (mit sechs Jahren), die schriftsprachlichen Fähigkeiten wurden zu Anfang des zweiten Schuljahrs sowie zum Ende der Schuljahre zwei bis fünf erfasst. Aufgaben auf der Basis von Phonemen zur phonologischen Bewusstheit wurden nicht durchgeführt, man konzentriere sich ausschließlich auf größere Einheiten (Wort, Silbe). Es konnten erst zum Ende des zweiten Schuljahrs signifikante Korrelationen (unter Kontrolle der nonverbalen Intelligenz) zwischen der phonologischen Bewusstheit und dem Erfolg in der Schriftsprache festgestellt werden. Zu Beginn der zweiten Klasse scheint noch die alphabetische Strategie zu überwiegen, so dass Fähigkeiten der phonologischen Bewusstheit im weiteren Sinne keine Relevanz zu besitzen scheinen. Die Studie zeigt, dass schriftsprachliche Leistungen am Ende des zweiten Schuljahrs unabhängig von der Befähigung des Einsatzes des alphabetischen Prinzips durch die phonologische Bewusstheit vorhergesagt werden können (vgl. Näslund & Schneider 1993, S. 295ff.; Schneider & Näslund 1992, S. 260ff.; 1997, S. 139).

Auch Landerl, Linortner und Wimmer (1992, S. 17ff.) konnten feststellen, dass die Phonemebene der phonologischen Bewusstheit besonders am Ende des ersten Schuljahres relevant ist. Sie kann der alphabetischen Strategie zugeordnet werden. Zum Ende des zweiten Schuljahres überwiegt jedoch wieder die Onset-Reimebene, die vermuten lässt, dass in der Phase vermehrt größere phonologische Einheiten (Silben, …) genutzt werden, die für die orthografische Ebene notwendig sind. Zusammenfassend können gute schriftsprachliche Fähigkeiten am Ende der ersten bis dritten Klasse durch die Fähigkeiten bei Aufgaben zur phonologischen Bewusstheit auf der Phonemebene vorhergesagt werden, während die Reimerkennungsfähigkeit erst am Ende des zweiten Schuljahres einen prognostischen Wert aufwies. Es zeigt sich also auch hier, dass die phonologische Bewusstheit differenziert mit ihren einzelnen Faktoren betrachtet werden muss (vgl. Wimmer, Landerl, Linortner & Hummer 1991, S. 219ff.; Landerl & Wimmer 1994, S. 153ff.; Landerl, Linortner & Wimmer 1992, S. 17ff.).

Mayringer, Wimmer und Landerl (1998, S. 57f.) erreichten dagegen in einer Studie mit 560 Jungen nur sehr schwache korrelative Zusammenhänge zwischen phonologischen Fähigkeiten und der Lese-Rechtschreibleistung. Unabhängig von den Korrelationen bei den Einzelprädiktoren konnten jedoch ca. 40 bis 50 % der Kinder mit Lese-Rechtschreibschwierigkeiten durch die Gesamtprädiktion auch richtig identifiziert werden. Dies ist im Vergleich zu anderen Studien jedoch ein eher geringer Wert und bedeutet auf der anderen Seite auch, dass ein Großteil der Kinder mit schwachen Leistungen in der phonologischen Bewusstheit unauffällig bleibt. Die Lese-Rechtschreibleistungen wurden zum Ende des ersten Schuljahres erhoben. So ist nicht auszuschließen, dass der langsam voranschreitende phonemische Anfangsunterricht bis zum Ende der ersten Klasse noch keine bessere Vorhersage zulässt, da selbst lese-rechtschreibschwache Kinder vielleicht noch die Anforderungen der ersten Klasse erfüllen können, sie jedoch bei komplexeren Leistungen (sinnverstehendes Lesen, orthographisch korrekte Schreibung, …) an ihre Grenzen stoßen und erst zu einem späteren Zeitpunkt ihrer Schullaufbahn auffällig werden.

In der Bielefelder Längsschnittstudie wurde das Bielefelder Screening (BISC – vgl. auch Kap. I.2.2.1) entwickelt, das Kinder mit besonderen Schwächen in der phonologischen Informationsverarbeitung erfasst. Die Autoren konnten belegen, dass die im Vorschulalter erfassten Werte der phonologischen Bewusstheit relevante Variablen im Schriftspracherwerb sind (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002, S. 48ff.). Marx und Jansen (1999, S. 17ff.) konnten herausfinden, dass die Prognose schriftsprachlicher Fähigkeiten durch die phonologische Bewusstheit bei 86 % der Rechtschreibleistung und 90 % der Leseleistung lag. Die Prognose fiel umso günstiger aus, je mehr die Schüler die Wortstrukturen analysieren konnten (vgl. Schneider & Näslund 1999, S. 126ff.; Landerl & Wimmer 1994, S. 153ff.). Zu den Studien gibt es jedoch wie bereits erwähnt schon gegenläufige Studien (vgl. Marx & Weber 2006; Schabmann 2009). Marx und Weber (2006, S. 251ff.) zeigten in einer Replikationsstudie weniger verheißende Ergebnisse auf.

Die Differenziertheit der phonologischen Bewusstheit stellte auch Schnitzler (2008, S. 70) anhand dreier Diplomarbeiten (Guntram 2004, Berendes 2006 und Rosebrock 2006) fest. Sie überprüften die Silben-, Onset-Reim- und die Phonemebene in ersten und dritten bzw. vierten Klassen. Die Schüler (N = 42) der ersten Klasse befanden sich größtenteils noch in der alphabetischen Phase, die Schüler der dritten bzw. vierten Klasse (N = 57) befanden sich größtenteils in der orthografischen Phase. Sowohl beim Wortlesen, Wortschreiben als auch beim Pseudowortlesen und -schreiben korrelierte die Phonemebene der phonologischen Bewusstheit mehr mit den abgefragten Fähigkeiten der Schüler in der alphabetischen Phase, während bei der orthografischen Phase die Silbenebene beim Wortlesen, die Onset-

Reimebene beim Wortschreiben und die Onset-Reim-Ebene beim Pseudowortlesen und -schreiben im Vordergrund stand (vgl. Tab. 20 und Tab. 21).

Tab. 20 Korrelationen zwischen Schriftsprachefähigkeiten und phonologischer Bewusstheit für Schüler der ersten Klasse (alphabetische Strategie), N = 42 (nach Schnitzler 2008, S. 70)

	Silben	**Onset-Reim**	**Phonem**
Wortlesen	0,479***	0,553***	0,589***
Wortschreiben	0,312*	0,248 n. s.	0,569***
Pseudowortlesen	0,294 n. s.	0,313*	0,499***
Pseudowortschreiben	0,137 n. s.	0,125 n. s.	0,507***

*Anmerkung: *** bedeutet p ≤ 0,001; ** bedeutet p ≤ 0,001; * bedeutet p < 0,05; n. s. = nicht signifikant*

Tab. 21 Korrelationen zwischen Schriftsprachefähigkeiten und phonologischer Bewusstheit für Schüler der dritten/vierten Klasse (orthografische Strategie), N = 57 (nach Schnitzler 2008, S. 70)

	Silben	**Onset-Reim**	**Phonem**
Wortlesen	0,334*	0,287*	0,234 n. s.
Wortschreiben	0,432***	0,537***	0,394**
Pseudowortlesen	0,152 n. s.	0,409**	0,131 n. s.
Pseudowortschreiben	0,214 n. s.	0,610***	0,421***

Anmerkung: *** bedeutet p ≤ 0,001; ** bedeutet p ≤ 0,001; * bedeutet p < 0,05; n. s. = nicht signifikant

Sie schließt daraus, dass zu vermuten ist, dass die phonologische Bewusstheit auf Phonemebene offensichtlich ein Prädiktor für die alphabetische Strategie zu sein scheint, während die Silben- und Onset-Reimebene für die orthografische Strategie prädiktiv ist. Mit Hilfe einer Varianzanalyse der Daten konnte Schnitzler (2008, S. 71f.) zeigen, dass jedoch nur in etwa die Hälfte der Varianz der Schriftsprachfähigkeiten durch die kontrollierten Variablen (phonologische Bewusstheit, nonverbale kognitive Fähigkeiten) aufzuklären ist. Sie vermutet, dass sich die ungeklärte Varianz durch zusätzliche Faktoren (psychosoziale Situation, Motivation, Interesse, schulische Instruktion) aufklären lässt. Aufgrund der sehr geringen Stichprobengröße dürfen die Ergebnisse nicht überinterpretiert werden, sie unterstützen jedoch bisherige Befunde.

Die Längsschnittstudie von Klicpera und Gasteiger-Klicpera (1993, S. 180ff.) konnte zeigen, dass die Vorhersage durch die phonologische Bewusstheit und das phonologische

Rekodieren beim Zugriff auf das semantische Lexikon bis ins achte Schuljahr hinein eine bedeutsame Relevanz hat. Außerdem bestätigten sie, dass die phonologische Bewusstheit auf der Phonemebene die Basis zum Erlernen der alphabetischen Strategie bildet (vgl. Klicpera & Gasteiger-Klicpera 1993, S. 180ff.). Eine Studie von Ennemoser, Marx, Weber und Schneider (2012, S. 53ff.) konnte bezüglich der phonologischen Bewusstheit durch zwei längsschnittliche Untersuchungen über vier Jahre bis zum Ende der Grundschulzeit mit N = 165 und N = 175 Kindern feststellen, dass diese neben weiteren Faktoren prädiktiv auf Lesegeschwindigkeit, Leseverständnis und auf das Rechtschreiben wirkt. Die Korrelation zwischen phonologischer Bewusstheit und dem Rechtschreiben lag bei $r = .47$ ($p < .01$ in Studie 1) bzw. $r = .64$ ($p < .01$ in Studie 2) und für die Lesegeschwindigkeit bei $r = .31$ ($p < .01$ in Studie 1) bzw. $r = .57$ ($p < .01$ in Studie 2).

Sowohl die englischen als auch die Korrelationsstudien aus dem deutschsprachigen Raum bestätigen die Ansicht, dass die implizite phonologische Bewusstheit eher als Voraussetzung des Schriftspracherwerbs betrachtet werden kann, es gibt jedoch Hinweise, dass eine Wechselwirkung vorliegt und die beiden Faktoren sich auch gegenseitig verstärken (vgl. zusammenfassend Schnitzler 2008, S. 56ff.). Außerdem scheinen schlechte Leistungen in der phonologischen Bewusstheit nicht zwangsläufig (vgl. Studie von Mayringer, Wimmer & Landerl 1998) einen Misserfolg im Lesen und Schreiben mit sich zu bringen, da sich die phonologische Bewusstheit auch noch im Laufe des Schulbesuchs herausbildet.

Die vorgestellten Korrelationsstudien stellen nur einen kleinen Ausschnitt der sehr umfangreichen Forschung zur phonologischen Bewusstheit dar. Sie konnten einen signifikanten Zusammenhang zwischen phonologischer Bewusstheit und dem Schriftspracherwerb herstellen. Die kausalen Beziehungen sollten jedoch zusätzlich über experimentelle Studien aufgestellt werden. Ein Ausschnitt dieser Trainingsstudien wird im Folgenden vorgestellt.

1.2.1.2 Trainingsstudien zur phonologischen Bewusstheit

In Trainingsstudien wurden Kinder im Bereich der phonologischen Bewusstheit gefördert und deren Leistungen im Schriftspracherwerb mit einer Kontrollgruppe verglichen. Auch hier ist zu berücksichtigen, dass beeinflussende Faktoren mit ausreichendem Umfang kontrolliert werden. In den meisten Studien wurde die phonologische Bewusstheit im letzten Kindergartenjahr trainiert und die Auswirkungen auf die schriftsprachlichen Fähigkeiten in der Grund- oder auch weiterführenden Schule überprüft. Es werden hier zunächst Studien

aus dem nicht-deutschen Sprachraum präsentiert, die dann durch deutsche Studien ergänzt werden.

Ergebnisse von Korrelationsstudien aus dem nicht-deutschsprachigen Raum

Die bereits bei den Korrelationsstudien erwähnten Autoren Bradley und Bryant (1983, S. 219) konnten in einer Untersuchung mit 65 Kindern nachweisen, dass Kinder, die ein Training im Bereich der phonologischen Bewusstheit erhielten, den Kindern ohne Training am Ende des zweiten Schuljahrs sowohl in der trainierten Fähigkeit als auch in schriftsprachlichen Fähigkeiten überlegen waren:

> *"Put together our longitudinal and training results provide strong support for the hypothesis that the awareness of rhyme and alliteration which children acquire before they go to school, possibly as a result of their experiences at home, has a powerful influence on their eventual success in learning to read and to spell." (a. a. O., S. 421)*

Die Autoren (1991, S. 37ff.) überprüften insgesamt sieben Jahre nach dem Training erneut 67 der Kinder und konnten weiterhin einen Vorteil gegenüber den Kontrollgruppen aufzeigen. Die fehlende Signifikanz der Werte kann, ihres Erachtens nach, durch Deckeneffekte und starke Varianz erklärt werden. Kritisch anzumerken ist, dass Bradley und Bryant nicht die vorherigen Schriftsprachfähigkeiten der Kinder im Kindergarten erfassten und damit eine gegenläufige Richtung ausschlossen (vgl. Küspert 1998, S. 79). Außerdem sei eine kausale Interpretation nicht sichergestellt, da die Kinder in der Untersuchung im Verlauf der zwei Jahre neben dem Lautkategorisierungstraining parallel mit dem Lese- und Rechtschreibunterricht der Schule konfrontiert waren (vgl. Schneider 1989, S. 157ff.).

Lundberg, Frost und Peterson (1988, S. 263ff.) dagegen kontrollierten frühe Schriftsprachfähigkeiten. Sie erzielten im skandinavischen Raum große Erfolge mit einem Trainingsprogramm der phonologischen Bewusstheit vor dem Schuleintritt. Die Trainingseffekte wurden anhand einer Gruppe (N = 235), die mit dem Programm täglich über acht Monate hinweg gefördert wurde, und einer Kontrollgruppe (N = 155) belegt. Alle 390 teilnehmenden Vorschulkinder waren normal entwickelt, es wurde kein Ausschuss gebildet. Anschließend wurde mit einem metaphonologischen Transfertest nochmals überprüft, ob die Schüler auch langfristig Effekte des Trainings in der ersten und zweiten Klasse erzielen konnten. Es zeigte sich, dass die Leistungen der phonologischen Bewusstheit auch über längere Zeit bei den geförderten Kindern stabil blieben. Die Trainingsgruppe erzielte auch am Ende des zweiten Schuljahrs noch signifikant bessere Ergebnisse als die

Kontrollgruppe. Das Trainingsprogramm hatte damit einen spezifischen Effekt. „Thus, the level of phonemic ability in pre-school is a powerful predictor of reading and spelling performance in school“ (Lundberg, Frost & Peterson 1988, S. 282).

Die Experimentalgruppe konnte für das Lesen signifikante Langzeiteffekte am Ende des ersten Schuljahres nachweisen, beim Rechtschreiben blieben die Effekte geringer (vgl. Abb. 9 und 10). Am Ende der zweiten und dritten Klasse zeigten sich die Effekte sowohl beim Lesen als auch beim Rechtschreiben für die trainierte Gruppe als signifikant (vgl. zusammenfassend Küspert 1998, S. 79ff.). Die Spezifität konnte durch einen Mathematiktest nachgewiesen werden, in dem die Kontrollgruppe der Experimentalgruppe überlegen war. Die Studie konnte damit zeigen, dass die phonologische Bewusstheit ohne Einblick in das alphabetische Prinzip erfolgreich trainiert werden kann und diese den Schriftspracherwerb insbesondere in den ersten beiden Grundschuljahren unterstützt und eine kausale Wirkungsrichtung von der phonologischen Bewusstheit auf die Lese-Rechtschreibfähigkeit angenommen werden kann.

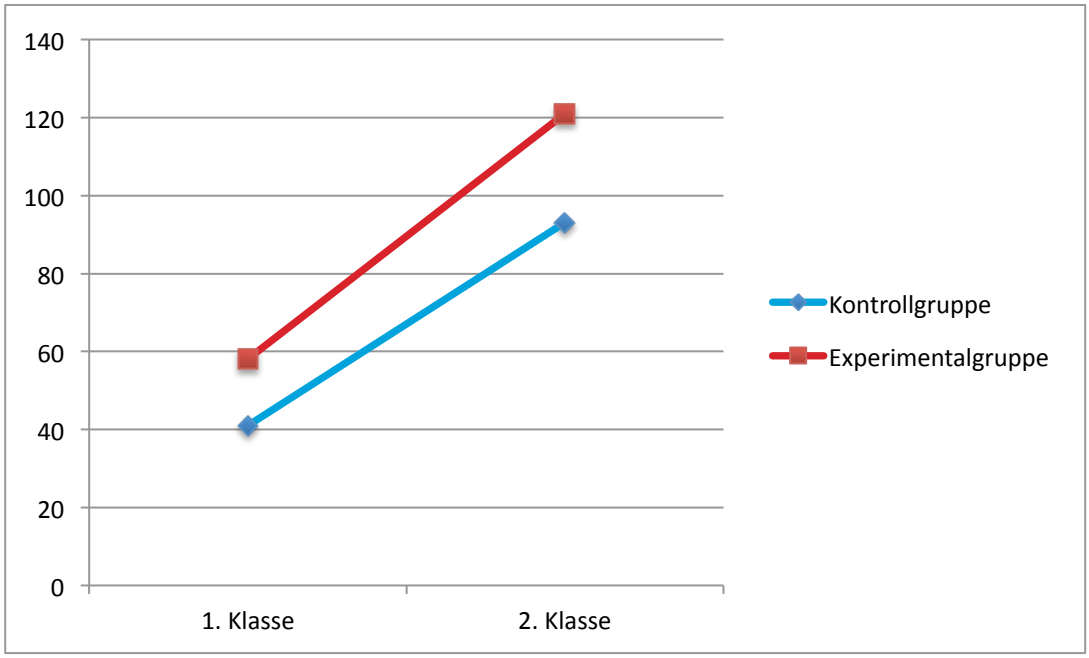


Abb. 9: Anzahl richtig gelesener Wörter in der ersten und zweiten Klasse in Experimental- und Kontrollgruppe (übersetzt aus dem Englischen, nach Lundberg, Frost & Peterson, S. 281)

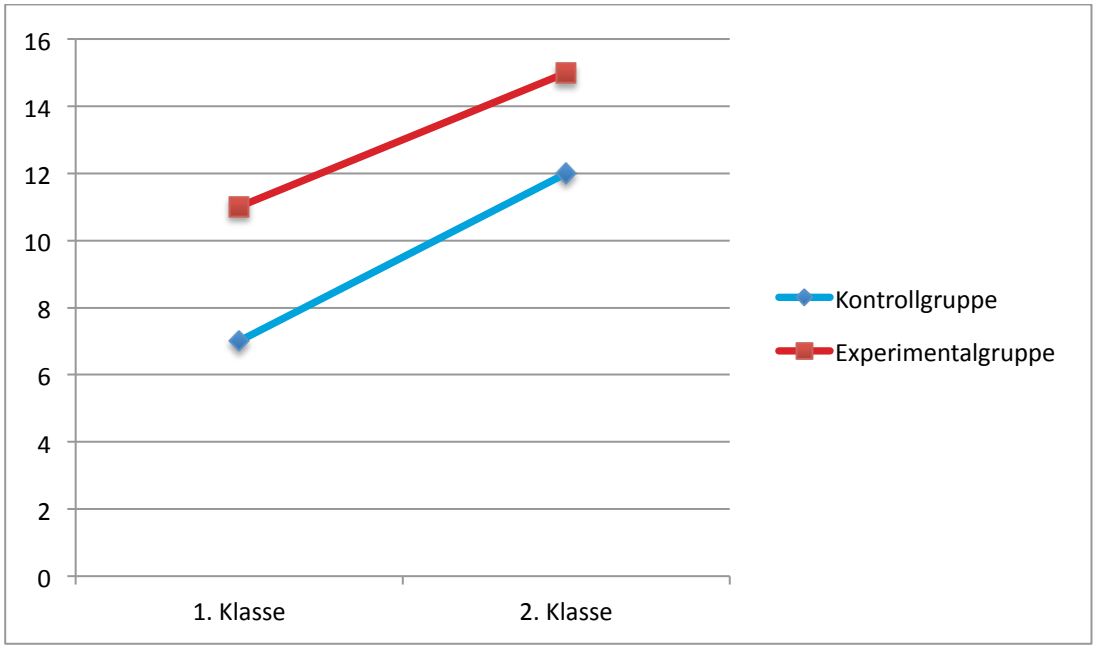


Abb. 10: Anzahl richtig buchstabierter Wörter in der ersten und zweiten Klasse in Experimental- und Kontrollgruppe (übersetzt aus dem Englischen, nach Lundberg, Frost & Peterson, S. 281)

Ähnliche Ergebnisse zeigt zusammenfassend die Metaanalyse von Bus und Ijzendoorn (1999, S. 406ff.) zu angloamerikanischen und europäischen Interventionsstudien. Die Analyse konnte nochmals verdeutlichen, dass die Effektstärke für das Training für die phonologische Bewusstheit bei etwa einer Standardabweichung, während die Effektstärke für die schriftsprachlichen Fähigkeiten nur bei einer halben Standardabweichung lag. Dies verdeutlicht, dass ein Training der phonologischen Bewusstheit in erster Linie die phonologische Bewusstheit trainiert und die Transfereffekte auf die Lesefertigkeit verringert wirken. Die Effektstärke beträgt d = .70, was nach Cohen (1988) noch als ein mittlerer bis hoher Wert eingeschätzt wird. Trotzdem konnte gezeigt werden, dass ein Training der phonologischen Bewusstheit 12 % der Varianz der Lesefähigkeit aufklärt (vgl. a. a. O., S. 403). “In sum, phonological awareness is an important but not a sufficient condition for early reading.” (a. a. O., S. 403).

In den Metaanalysen von Ehri et al. (2001, S. 250ff.) und Bus und van Ijzendoorn (1999, S. 406ff.) konnte auch gezeigt werden, dass sich ein Training der phonologischen Bewusstheit im Vorschulalter auf die schriftsprachlichen Fähigkeiten in der Schule auswirkt und dass hier Trainings, die zusätzlich auf Buchstabe-Laut-Korrespondenzen eingehen, einen Vorteil gegenüber rein phonologischen Trainings aufweisen. Dies konnte auch in den Würzburger Evaluationsstudien gezeigt werden (vgl. Weber, Marx & Schneider 2007, S. 91ff.), jedoch ließen sich die Ergebnisse dort statistisch nicht eindeutig absichern.

Byrne und Fielding-Barnsley (1989, S. 315ff.; 1991, S. 451ff.; 1993, S. 104ff.; 1995, S. 488f.) führten in einer australischen Studie mit Vorschulkindern eine Trainingsstudie mit dem Programm Sound-Foundations (Byrne & Fielding-Barnsley 1991) durch. Das Programm trainiert die Phoneminvarianz anhand gleicher Anfangs- oder Endlaute. Es wurden neun Phoneme bearbeitet. 64 Vorschulkinder wurden für zwölf Wochen trainiert. Die Kontrollgruppe (N = 62) wurde nur mit semantischen Kategorisierungskriterien trainiert. Durch einen Vortest wurden die Fähigkeiten vor Beginn des Trainings festgehalten. Die Experimentalgruppe verzeichnete direkt nach Ende der Einheit einen signifikanten Unterschied in den Leistungen der phonologischen Bewusstheit und im Lesetest im Vergleich zur Kontrollgruppe. Zum Ende des ersten Schuljahres konnten jedoch keine signifikanten Unterschiede in der phonologischen Bewusstheit und auch in den Leseleistungen festgestellt werden, nur ein Vorteil der Experimentalgruppe beim Pseudowortlesen war zu verzeichnen. Dieses Bild ergab sich ebenfalls nach dem zweiten Schuljahr. Im dritten Schuljahr konnte zusätzlich festgestellt werden, dass die Schüler der Experimentalgruppe einen signifikanten Unterschied im Leseverständnis zur Kontrollgruppe aufwiesen. Die Autoren interpretierten ihre Ergebnisse so, dass die phonologische Bewusstheit vor Schulbeginn sich vermutlich auf den späteren Schriftspracherwerb positiv auswirkt. Die Auswirkung auf die Rechtschreibleistungen konnte in dieser Studie nicht bestätigt werden. Es wird vermutet, dass durch zusätzliche Buchstaben-Laut-Instruktionen die Rechtschreibleistungen hätten besser vorbereitet werden können (vgl. Küspert 1998, S. 85).

Ergebnisse von Korrelationsstudien aus dem deutschsprachigen Raum

Im deutschsprachigen Raum konnte Blumenstock (1979, S. 178f.) bereits sehr früh mit Erstklässlern erfolgreich eine Präventionsmaßnahme mit der phonologischen Bewusstheit durchführen. Die Schüler der ersten Klassen wurden mit Aufgaben zur phonologischen Bewusstheit sowie mit Übungen zur Phonemdiskrimination und Graphem-Phonem-Korrespondenzen trainiert. Die Erstklässler verbesserten sich sowohl im Bereich der trainierten Fähigkeiten zur phonologischen Bewusstheit als auch indirekt in ihren schriftsprachlichen Fähigkeiten. Es wurden jedoch keine Langzeiteffekte überprüft.

Die Förderung phonologischer Fähigkeiten im deutschsprachigen Raum scheint möglich zu sein. Dies wurde beispielsweise auch bei der Evaluation des Trainingsprogramms „Lobo vom Globo“ (vgl. Fröhlich, Metz & Petermann 2009, S. 204ff.) gezeigt. Schneider und Küspert (2003, S. 117ff.) konnten ebenfalls Erfolge mit dem Förderprogramm zur phonologischen Bewusstheit „Hören, Lauschen, Lernen“ erzielen, was sich auf die schriftsprachlichen Fähigkeiten auswirkte. Die durch das Bielefelder Screening ausgelesenen

Kinder der Trainingsgruppe wurden über 20 Wochen hinweg täglich in Kleingruppen gefördert. Die Trainingsgruppe zeigte gegenüber der Kontrollgruppe sowohl beim ersten Nachtest als auch beim späteren metaphonologischen Transfertest einen signifikanten Unterschied in den Lese-Rechtschreib-Leistungen. Der Vorsprung der Kinder in der Trainingsgruppe ließ sich bis zum Ende der zweiten Klasse nachweisen (vgl. Abb. 11) und hatte im Vergleich zu einer unausgelesenen repräsentativen Stichprobe (8 % Kinder mit Schwierigkeiten im Lesen und Rechtschreiben) nur 20 % lese-rechtschreibschwache Kinder. Unspezifische Trainingseffekte waren nicht vorhanden, das Training hat sich ausschließlich auf den Bereich Lesen und Rechtschreiben sowie auf die Leistungen in der phonologischen Bewusstheit ausgewirkt.

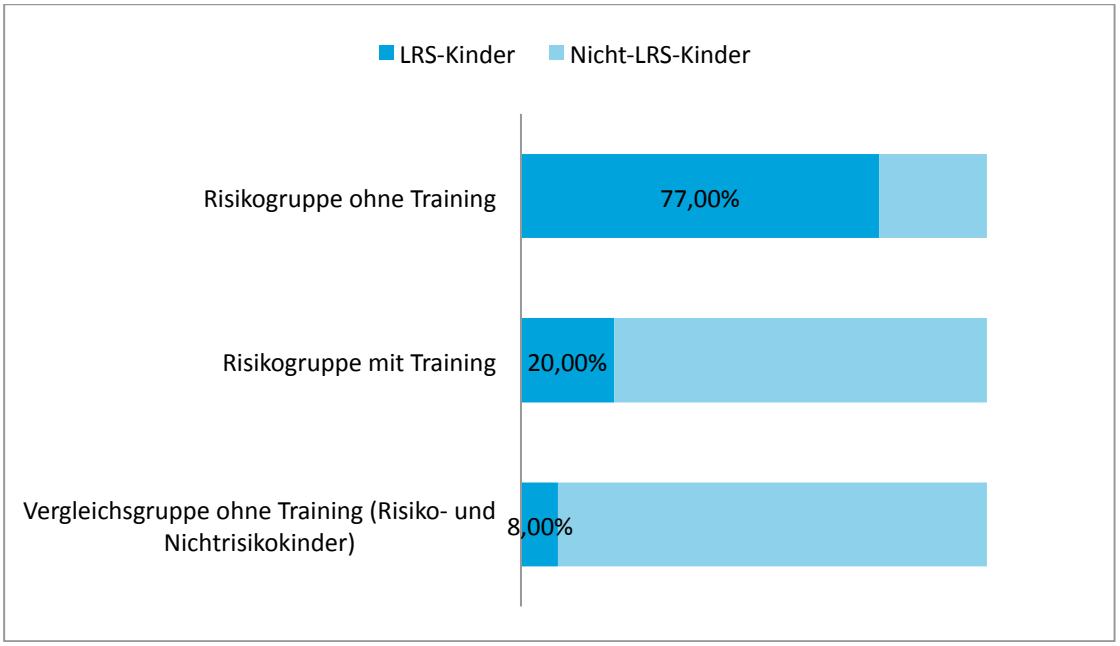


Abb. 11: Anteil von Kindern mit Lese-Rechtschreibschwäche am Ende der zweiten Klasse in drei verschiedenen Gruppen, (angelehnt an Mannhaupt 2003, S. 48)

Die Trainingsstudien verdeutlichen, dass die Förderung spezifischer Fertigkeiten der phonologischen Bewusstheit auch zu positiven Effekten für das Erlernen des Lesens und Schreibens der Kinder führt. In einer weiteren Studie konnten Küspert und Schneider (2006, S. 121ff.) mit einer Stichprobe von 150 durch den BISC als Risikokinder identifizierten Kindern nachweisen, dass eine Förderung der phonologischen Bewusstheit und der Buchstabenkenntnis den Risikokindern den Einstieg in das Lesen und Rechtschreiben erleichtert. Fischer und Pfost (2015, S. 40ff.) konnten in einer Meta-Analyse von Trainingsstudien zur phonologischen Bewusstheit im deutschsprachigen Raum einen mittleren statistisch bedeutsamen Trainingseffekt von $d = .36$ auf die phonologische

Bewusstheit sowie signifikante kurzfristige (d = 0.21) und langfristige Effekte (d = 0.14) auf das Lesen und Rechtschreiben feststellen.

In der Schweiz wurde ein kombiniertes Training zur phonologischen Bewusstheit und zu Graphem-Phonem-Korrespondenzen durchgeführt (vgl. Blaser, Preuss, M. Groner, L. Groner & Felder 2007, S. 276f.). Die Autoren zeigten anhand einer gepaarten Stichprobe von je 26 Kindern in Kontroll- und Experimentalgruppe, dass sich die Effekte des Trainings zum Ende des ersten Schuljahres noch feststellen ließen. Nach dem zweiten Schuljahr ließ sich noch eine deutlich verringerte Prävalenz der Trainingsgruppe zeigen, die sich jedoch nur auf die Rechtschreibproblematik bezog und sich nicht auf das Lesen übertragen ließ. Rechtschreibleistungen konnten somit durch ein Training deutlich positiver beeinflusst werden, was sich auch mit den Befunden von z. B. Schneider, Roth und Ennemoser (2000) deckt. Blaser, Preuss und Felder (2010, S. 183) konnten auch die Langzeiteffekte des Trainings bzgl. des Anteils an rechtschreibschwachen Kindern in der Trainingsgruppe in der dritten und vierten Klasse bestätigen. Die Trainingsgruppe wies eine deutlich geringere Zahl an Kindern mit Rechtschreibschwierigkeiten auf. Sie betonen jedoch, dass trotzdem ein kleiner Anteil an Kindern übrig bleibt, der trotz der Förderung eine Lese-Rechtschreibproblematik entwickelt hat.

Einsiedler, Frank, Kirschhock, Martschinke und Treinies (2002, S. 194ff.) führten mit 15 Klassen (unausgelesene Stichprobe) eine Studie zur Förderung der phonologischen Bewusstheit in der ersten Klasse durch. Fünf der Klassen nahmen an einem Training zur phonologischen Bewusstheit teil, die anderen fünf Klassen erhielten Fibelunterricht und weitere fünf Klassen setzten ein eher offenes Konzept zum Schriftspracherwerb ein. Die drei Untersuchungsgruppen wurden aufgrund ihrer Leistungen in der phonologischen Bewusstheit parallelisiert. Das Resultat war, dass die trainierten Klassen klare Vorteile gegenüber den Vergleichsgruppen hatten. Die Kinder mit normalen oder schlechten Lernvoraussetzungen profitierten von dem Training am meisten, wobei die Kinder mit den schlechtesten Lernvoraussetzungen eine zusätzliche Stunde Training pro Woche erhielten. Die Effekte unterschieden sich besonders hoch bei den Lesefähigkeiten und dem Leseverständnis am Ende der zweiten Klasse. Beim Rechtschreiben blieben signifikante Langzeiteffekte aus. Die Autoren vermuten an dieser Stelle, dass die Lehrer in der Kontrollgruppe beim Rechtschreiben explizit dem Trainingsprogramm ähnliche Aufgaben auf Basis der phonologischen Bewusstheit gestellt haben (Durchgliederung der Wörter usw.) und dadurch der Effekt verfälscht worden ist (vgl. a. a. O., S. 205; Einsiedler & Kirschhock 2003, S. 57).

Eine Förderung von lese-rechtschreibschwachen Kindern in der zweiten Klasse führte in der Studie von Wimmer und Hartl (1991, S. 74ff.) nicht zum Erfolg, da die Basiskenntnisse der phonologischen Bewusstheit selbst bei lese-rechtschreibschwachen Kindern soweit ausgeprägt waren, dass Deckeneffekte bei der Überprüfung der phonologischen Bewusstheit durch ein geeignetes Testinstrument auftraten. Es ist zu vermuten, dass die teilnehmenden Schüler bereits zu dem Zeitpunkt die alphabetische Strategie nutzten und sie deshalb Übungen auf einer anderen Ebene bedürfen. Es ist jedoch auch nicht auszuschließen, dass die gewählten Aufgaben und Übungen nicht sinnvoll waren, und dass diese sich beispielsweise mehr an der alphabetischen und orthografischen Strategie orientieren müssen.

Scheerer-Neumann (1981a, S. 183) konnte dagegen feststellen, dass sogar Schüler der dritten Klasse mit einer entfalteten alphabetischen Strategie noch von einem Silbentraining (auditive und visuelle Segmentierung) profitierten. Dies stützt die Ergebnisse der Korrelationsstudien der Diplomarbeiten von Schnitzler (2008, S. 71f., vgl. Kap. II.1.2.1.1), die davon ausgeht, dass das Wissen über größere phonologische Einheiten besonders beim Erwerb der orthografischen Strategie eine größere Bedeutung hat.
Die Studie von Lundberg, Frost und Peterson (1988, S. 263ff.) wurde im deutschsprachigen Raum von Schneider, Visé, Reimers und Blaesser (1994, S. 186ff.) repliziert. Die Ergebnisse konnten weitestgehend bestätigt werden. Die groß angelegte Trainingsstudie erfolgte von 1991 bis 1998. Als Trainingsgegenstand wurde über 20 Wochen täglich in den teilnehmenden Kindergärten das Würzburger Trainingsprogramm eingesetzt. In einer ersten Studie zeigte sich jedoch, dass nicht alle Kinder vom Training profitieren konnten und insbesondere die Langzeiteffekte von der Qualität der Trainingsdurchführung abzuhängen schienen. Bei einer zweiten Studie wurden diese Effekte durch eine Supervision der Erzieher kontrolliert, außerdem wurden einige redundante Einheiten (Reime) im Training ersetzt. Die 191 Kinder der unausgelesenen Stichprobe waren der Kontrollgruppe (N = 155) sowohl bei der phonologischen Bewusstheit als auch beim Lesen und Schreiben selbst bis zum Ende des zweiten Schuljahres überlegen (vgl. Schneider, Visé, Reimers & Blaesser 1994, S. 186ff.; Schneider 2001a, S. 215ff.; Schneider 2001b, S. 86f.). Besonders die Übungen auf der Phonemebene waren für die Effizienz des Trainings verantwortlich. Der Grund dafür kann die besondere Nähe an der alphabetischen Strategie sein. Die Effekte auf phonologischer Ebene konnte auch eine Studie von Noack und Brändel (2007, S. 326ff.) bestätigen.

Bei einer Sekundäranalyse der Daten (vgl. Schneider, Roth, Küspert & Ennemoser 1998, S. 30ff.) wurden die Kinder der Trainings- und der Kontrollgruppe aufgrund ihres Leistungsniveaus in drei Gruppen geteilt. Es war zu überprüfen, ob auch die schwächsten

Kinder eines Jahrgangs von einem Training der phonologischen Bewusstheit profitierten. Die Ergebnisse zeigten, dass die Förderung für Risikokinder effektiv ist, um sie auf einen erfolgreichen Schriftspracherwerb vorzubereiten. Während der Schulzeit erreichten die trainierten Risikokinder sogar das mittlere Leistungsniveau. Damit hatten sie gegenüber den Risikokindern der Kontrollgruppe einen erheblichen Vorsprung. Insgesamt wiesen nur noch ein Drittel der Risikokinder der Trainingsgruppe bei den Ergebnissen einen PR[14] < 15 auf, während zwei Drittel der Risikokinder, die kein Training erhalten haben, dieses Leistungsniveau aufzeigten. Auch Kinder des oberen Leistungsniveaus konnten von dem Training profitieren (vgl. Schnitzler 2008, S. 63).

Rothe (2008, S. 21f.) führte ebenfalls eine experimentelle Längsschnittstudie mit zwei Altersklassen von Kindern von 2001 bis 2004 im letzten bzw. vorletzten Jahr des Kindergartens durch. Die Schüler wurden zunächst aufgrund ihrer Grundvoraussetzungen in der phonologischen Bewusstheit überprüft. Im Anschluss fand in den Experimentalgruppen ein halbes Jahr lang das Würzburger Trainingsprogramm „Hören, Lauschen und Lernen“ (Küspert & Schneider 2006) statt. Die älteren Kinder nahmen einmal ein halbes Jahr und die jüngeren je zweimal ein halbes Jahr am Training teil. In der Kontrollgruppe fand keine explizite Förderung statt. Nach der Einschulung wurde das Training in der ersten Klasse fortgeführt, parallel dazu wurde eine weitere Trainingsgruppe erstellt, die ein Training der phonologischen Bewusstheit ab Klasse 1 erhielt. Das Resultat der Untersuchung war, dass die phonologische Bewusstheit der Schüler in den Experimentalgruppen sich im Vergleich zu den Kontrollgruppen signifikant verändert hat, jedoch der Transfer auf die schriftsprachlichen Leistungen bereits am Ende des ersten Schuljahres ausblieb. Die Vermutung der Autorin ist, dass zwar ein Zusammenhang bestehen kann, jedoch noch weitere relevante Variablen den Schriftspracherwerb beeinflussen.

Im Projekt EVES (Evaluation eines Vorschultrainings zur Prävention von Schriftsprach-erwerbsproblemen sowie Verlauf und Entwicklung des Schriftspracherwerbs in der Grundschule) konnten die Autoren Roos und Schöler (2007, S. 8) mit 232 Kindern eine Untersuchung im letzten Kindergartenjahr durchführen. Sie konnten im Vergleich zur Kontrollgruppe feststellen, dass sich die Effekte in der ersten Klasse nur für Mädchen signifikant auf das Lesen auswirkten. Für Jungen war eher eine Tendenz des Transfers zu erkennen. Auch in den Klassenstufen zwei bis vier zeigten nur die Werte der Mädchen im Rechtschreiben und in der Lesegeschwindigkeit einen signifikanten Unterschied. In der Lesefähigkeit und in den Vorläuferfertigkeiten des verstehenden Lesens waren keine

14 PR steht für Prozentrang.

signifikanten Unterschiede zwischen Experimental- und Kontrollgruppen zu vermerken. Die Autoren schlussfolgern, dass insbesondere Mädchen von einem vorschulischen phonologischen Training profitieren, der Verlauf der Leistungen (insbesondere ein Anstieg der Leistungen am Ende der vierten Klasse) beim Rechtschreiben und bei den Vorläuferfertigkeiten des verstehenden Lesens bei Jungen jedoch darauf hinweist, dass die Effekte des Trainings verzögert auftreten könnten. Das Training hat also für die Rechtschreibung einen verspäteten Effekt und kann die Lesegeschwindigkeit deutlich erhöhen. Es scheint auch Auswirkungen zu haben, welche spezifischen Teilfertigkeiten im Schriftspracherwerb betrachtet werden. Außerdem betonen die Autoren einen besonderen Vorteil für Kinder nichtdeutscher Muttersprache (vgl. a. a. O., S. 24). Auch Walter (2002a, S. 41f.) erklärte über die Meta-Analyse von Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh und Shanahan (2001), dass die mittleren Transfereffekte nach einem Training der phonologischen Bewusstheit im engeren Sinne auf die Leseleistung in Nachuntersuchungen immer weniger werden, bis sie im Verlaufe der weiteren Entwicklung der Kinder nicht mehr festzustellen sind. Er weist darauf hin, dass die phonologische Bewusstheit als Allheilmittel ausgedient habe (vgl. a. a. O., S. 48).

Anhand der Trainingsstudien wird deutlich, dass ein Training der phonologischen Bewusstheit bis in das erste Schuljahr hinein sinnvoll ist und dass sich bei diesem Training nicht nur die trainierten Fähigkeiten verbessern, sondern die Schüler auch beim Erwerb der Schriftsprache erfolgreicher sind. Risikokinder können ihre Rückstände sogar ausgleichen. Bei einem Training nach der zweiten Jahrgangsstufe scheinen keine Effekte hinsichtlich einer weiteren Förderung aufzutreten (vgl. Wimmer & Hartl 1991, S. 74), dies scheint aber auch von der Auswahl der Übungen zur phonologischen Bewusstheit abzuhängen. Es fällt jedoch ins Auge, dass ein prädiktiver Wert der phonologischen Bewusstheit auch scheinbar in Abhängigkeit mit dem schulischen Leselehrgang zu betrachten ist (vgl. Einsiedler, Frank, Kirschhock, Martschinke & Treinies 2002, S. 194ff.) und die Effektivität des phonologischen Trainings umso höher ist, je expliziter es im Zusammenhang mit dem Erwerb des Buchstabeninventars steht (vgl. May & Okwumo 1999, S. 14). Außerdem darf der Einfluss des schulischen Anfangsunterrichts nicht unterschätzt werden, was in Kapitel II.2.3 nochmals genauer beleuchtet wird.

1.2.1.3 Gruppenvergleich

In Gruppenvergleichen wurde der Zusammenhang von phonologischer Bewusstheit und dem Erfolg im Schriftspracherwerb hergestellt. Kinder mit einer Lese-Rechtschreibschwäche oder

erwachsene Analphabeten werden mit einer Gruppe von Kindern bzw. Erwachsenen mit ausreichenden schriftsprachlichen Fähigkeiten verglichen.

Morais, Cary, Alegria und Bertelson (1985, S. 323ff.) überprüften Erwachsene portugiesische Analphabeten in Bezug auf ihre Fähigkeiten in der phonologischen Bewusstheit und verglichen sie mit einer Kontrollgruppe. Es zeigte sich, dass die Gruppe der Analphabeten durchweg in allen Bereichen (z. B. Anlautergänzung, Anlautentfernung, Phonemvertauschung, …) schlechter abschnitt als die Kontrollgruppe. Es ließ sich anhand der Ergebnisse jedoch vermuten, dass zumindest Teilfertigkeiten der phonologischen Bewusstheit ausgeprägt waren, da in der Gruppe der Analphabeten z. B. bei Aufgaben zur Silbenvertauschung fast die Hälfte der Teilnehmer die Aufgabe richtig bearbeitet hatte.

Derlei Gruppenvergleiche gerieten immer mehr in die Kritik der Forscher, da vorgefundene Defizite im Bereich der phonologischen Bewusstheit nicht als Ursache oder Konsequenz der Lese-Rechtschreibschwäche zuzuordnen waren (vgl. Weber 2003, S. 52). Wagner und Torgesen (1987, S. 196) bezweifeln auch, ob Ergebnisse, die mit Analphabeten gewonnen wurden, verallgemeinert werden können: “The results may apply only to the atypical individuals studies, in this case, Portugueses illiterates.”

Morais (1991, S. 5ff.) führt ebenfalls zwei Einzelfallstudien mit Dichtern an, die der Schriftsprache nicht mächtig waren, jedoch Reime bilden, aber keine Anlaute erkennen konnten. Dies wiederum spricht für einen wechselseitigen Einfluss von phonologischer Bewusstheit und dem Erwerb von schriftsprachlichen Fähigkeiten, wobei den zwei Einzelfällen aufgrund der geringen Verallgemeinerungsmöglichkeit keine allzu große Relevanz zugesprochen werden darf. Bradley und Bryant (1978, S. 746ff.) untersuchten Schüler mit Schwierigkeiten im Lesen und Rechtschreiben im Alter von zehn Jahren, die einen Rückstand von ca. zwei bis drei Jahren in ihren schriftsprachlichen Fähigkeiten aufwiesen. Die Autoren untersuchten, inwieweit die Schüler Alliterationen und Reime erkennen konnten. Es zeigte sich, dass die Schüler im Vergleich zu jüngeren, „normalen Schülern“ signifikant mehr Fehler bei diesen Aufgaben machten.

Vellutino, Scanlon und Spearing (1995, S. 85ff.) konnten in mehreren Gruppenvergleichen zwischen normalen Lesern und leseschwachen Schülern ebenfalls bestätigen, dass die meisten Schüler mit Leseproblemen auch Probleme mit der phonologischen Bewusstheit aufwiesen: “It was also concluded that phonological coding deficits are a probable cause of reading difficulties.” (a. a. O., S. 76).

Anhand einer Vergleichsstudie konnte Wimmer (1993, S. 1ff.) ermitteln, dass die Leistungen in der phonologischen Bewusstheit von leseschwachen Schülern zwar unter dem Niveau von Kindern in ihrem Alter lagen, sie jedoch im Vergleich zu Kindern, die ein vergleichbares Niveau der Schriftsprache aufwiesen, auf einem Niveau lagen. Landerl (1996, S. 56ff.) zeigte dagegen in einer Untersuchung, dass zwölfjährige Kinder den ca. drei Jahre jüngeren Kindern, die sich auf einem schriftsprachlichen Leistungsniveau befanden, deutlich unterlegen waren und somit keine aufgrund des Alters kognitive Kompensationsleistungen bei der Schüttelreimaufgabe unterstützen konnten. Sie folgert daraus, dass bei diesen Kindern wahrscheinlich ein grundsätzliches Problem im phonologischen Arbeitsspeicher vorliegt. Aufgrund der nicht eindeutigen Forschungslage scheinen in diesem Bereich weitere Untersuchungen von Nöten zu sein. Es scheint jedoch ein Defizit in der phonologischen Bewusstheit vorzuliegen, das allein durch Erfahrungen beim Schriftspracherwerb nicht ausgeglichen werden kann.

Marx, Weber und Schneider (2001, S. 91ff.) unterschieden in ihrer Studie drei Gruppen (N = 116 ausgewählt aus 823 Drittklässlern): Legastheniker (mit Diskrepanz zwischen IQ und Lese-Rechtschreibleistung), lese-rechtschreibschwache Kinder (Diskrepanz kleiner als 1,5 Standardabweichungen) und eine Kontrollgruppe. Sie konnten zum einen zeigen, dass sich beide Gruppen von der Kontrollgruppe unabhängig von ihrem Intelligenzniveau bzgl. ihrer Leistungen in der phonologischen Informationsverarbeitung signifikant unterschieden und zum anderen, dass die Gruppen untereinander jedoch nur im Pseudowortlesen eine Diskrepanz aufwiesen. Die Autoren betonen, dass durch dieses Versuchsmodell nicht eindeutig zu erfassen ist, inwieweit die Defizite eine Folge oder eine Ursache der Schwierigkeiten im Lesen und Schreiben sind.

Bei den Korrelations- und Trainingsstudien sowie bei den Studien zu den Gruppenunterschieden kam immer wieder die Frage auf, ob sich die phonologischen Fertigkeiten vor dem Schriftspracherwerb ausbilden, dies mit diesem einher geht oder erst durch diesen erworben wird. Im nächsten Abschnitt wird erörtert, wann und wie sich die phonologische Bewusstheit entwickelt.

1.2.1.4 Entwicklung der phonologischen Bewusstheit: Ursache oder Wirkung?

In der vorliegenden Arbeit geht es insbesondere um die Überprüfung der Fähigkeiten im Bereich der phonologischen Bewusstheit bei Vorschulkindern. Die oben dargestellten Korrelations- und Trainingsstudien ließen erkennen, dass Kinder bereits vor Schuleintritt

Vorläuferfertigkeiten bezüglich des Schriftspracherwerbs haben, die in den einzelnen Kategorien der phonologischen Bewusstheit unterschiedlich ausgeprägt sind. Es bedarf damit einer differenzierten Betrachtung der Fähigkeiten der phonologischen Bewusstheit.

Silbenebene

Vorschulkinder erzielen im Durchschnitt auf der Silbenebene auf der impliziten Bewusstseinsebene sehr gute Leistungen (vgl. Schnitzler 2008, S. 39). Jansen, Mannhaupt, Marx und Skowronek (1999, S. 37) verzeichneten, dass 82 % bzw. 78 % der Vorschulkinder beim Erkennen von Silben (Klatschen von Silben und Anlaut-Identifikation von Anlauten auf Silbenebene) richtige Antworten lieferten. Auch beim Zusammenziehen von Silben ließen sich gute Ergebnisse der Vorschulkinder nachweisen (vgl. Fricke 2005, zit. in Schnitzler 2008, S. 39). Es ist jedoch zu vermuten, dass die Kinder auch aufgrund von Rateprozessen so gute Leistungen erzielten und das Wort mit Hilfe von Klangassoziationen gefunden haben. Die Bewusstheit der Handlung kann nicht definitiv als explizit betrachtet werden. Manipulationsaufgaben sind für Vorschulkinder zu schwierig zu bewältigen.

Onset-Reimebene

Auf der Reimebene (implizite Bewusstheit) erzielen Vorschulkinder aufgrund von Erfahrungswerten und Umgang mit Reimen im Alltag im Durchschnitt sehr gute bis gute Leistungen, während die Leistungen auf der Onsetebene (implizite Bewusstheit) mäßig sind (vgl. Bryant & Goswami 1987, S. 213ff.; Schneider, Näslund & Weinert 1989, S. 157ff.; zusammenfassend Schnitzler 2008, S. 40). Treiman und Zukowski (1991, S. 67ff.) konnten herausfinden, dass sich die Bewusstheit für subsyllabische Einheiten Onset und Rime nach der Bewusstheit für Silben, jedoch vor der Bewusstheit für Phoneme entwickelt. Bezüglich einer Förderung konnten Nation und Hulme (1997, S. 154ff., vgl. auch Kap. II.1.2.1.2) zeigen, dass das Onset-Reim-Bewusstsein aktiv instruiert werden muss, damit es Leseanfängern einen Vorteil im Schriftspracherwerb verschafft: “Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills.” (a. a. O., S. 154).

Phonemebene

Auf der impliziten Ebene der Bewusstheit erzielten Vorschulkinder im Durchschnitt mäßige Leistungen. Van Bon und van Leeuvwe (2003, zit. in Schnitzler 2008, S. 40) zeigten in ihrer Studie in den Niederlanden, dass Kinder auch unabhängig von Schriftspracherfahrungen eine Bewusstheit für Phoneme entwickeln. Die Vorschulkinder erreichten jedoch bei Aufgaben auf der expliziten Bewusstheitsebene (Synthetisieren, Segmentieren, Manipulieren) schlechte Leistungen. Jansen und Thomé (1998, S. 23) konnten in einer Studie mit 59 Erstklässlern

zeigen, dass die Kenntnis der Wortschreibung („Äste“) nicht zwingend Auswirkungen auf die Qualität der Phonemanalyse hat. Die Kinder scheinen zusätzlich zur buchstabenorientierten Analyse größtenteils die orthografische Besonderheit im Gedächtnis markiert zu haben, um dies dann bei der Schreibung zu berücksichtigen.

Die dargestellten Befunde demonstrieren, dass die Fähigkeiten der phonologischen Bewusstheit zum einen differenziert werden müssen, zum anderen ist die Frage der Entwicklung dieser noch nicht geklärt. Eine zentrale Fragestellung in der Forschung ist: Entwickeln sich phonologische Fähigkeiten während des Schriftspracherwerbs oder existieren erst die phonologischen Fähigkeiten, die dann einen Beitrag als Voraussetzung zum Schriftspracherwerb leisten?

Bezüglich des Zusammenhangs von Fähigkeiten der phonologischen Bewusstheit und dem Erfolg beim Schriftspracherwerb gibt es in der Forschung verschiedene Ansätze (vgl. Ehri 1979, S. 63ff.; Schnitzler 2008, S. 55ff.):

- Die phonologische Bewusstheit wird als Voraussetzung zum Erlernen des Lesens und Schreibens betrachtet.
- Die phonologische Bewusstheit wird als Produkt des Schriftspracherwerbs betrachtet und wird erst dann erworben, wenn schon schriftsprachliche Fähigkeiten erworben wurden.
- Die phonologische Bewusstheit steht in einer Wechselwirkung mit dem Erwerb der Schriftsprache.

Ziegler und Goswami (2005, S. 3ff.) gehen davon aus, dass die Elemente im mentalen Lexikon zunächst nach ihrer größtmöglichen phonologischen Ähnlichkeit sortiert werden. Das mentale Lexikon bedarf dieser Struktur, damit das Suchsystem bei Abruf der Wörter auf redundante Einheiten zurückgreifen kann, um die Geschwindigkeit des Abrufs zu erhöhen. Vorschulkinder greifen zunächst auf größere Einheiten zurück. Je mehr sich der Wortschatz vergrößert, umso mehr Kriterien müssen gebildet werden, nach denen die Einheiten in Unterkategorien sortiert werden können (vgl. Schnitzler 2008, S. 37f.).

Marx (1992b, S. 225) konnte zeigen, dass nicht nur die phonologische Bewusstheit für die Prognose der Leistung im Schriftspracherwerb relevant ist. Auch für die Bereiche Aufmerksamkeit und Gedächtnis zeigen die Korrelationen hochsignifikante Verbindungen auf (vgl. Tab. 22).

Es zeigt sich, dass die Aufmerksamkeits- und Gedächtnisleistungen nur geringfügig weniger mit der Rechtschreibleistung korrelieren. In der Tabelle wird auch sichtbar, dass Aufgaben zur phonologischen Bewusstheit nach Schuleintritt (Testzeitpunkt 3 und 4) sehr früh von Schriftspracherfahrungen beeinflusst werden (vgl. auch Jansen 2008, S. 177ff.).

Tab. 22: Korrelationen zwischen Phonologischer Bewusstheit und Aufmerksamkeit und Gedächtnis zur Schreibfähigkeit (aus Marx 1992b, S. 225, Tabelle 20)

		Phonologische Bewusstheit				Aufmerksamkeit und Gedächtnis			
		T1	T2	T3	T4	T1	T2	T3	T4
Schreibfähigkeit	T3	.551	.514	.692	.465	.493	.503	.645	.602
	T4	.565	.552	.730	.708	.497	.439	.583	.601

Anmerkung: T1: Testzeitpunkt 1, 9 Monate vor der Einschulung, T2: Testzeitpunkt 2, 4 Monate vor der Einschulung, T3: Testzeitpunkt 3, Mitte der ersten Klasse, T4: Testzeitpunkt 4, Ende der ersten Klasse

Aufgrund dieser Ergebnisse beschreibt Jansen (2008, S. 185f.), dass die Früherfassung von phonologischer Bewusstheit unter der Validitätsproblematik leidet, da nicht abschätzbar ist, inwieweit schriftspezifische Teilprozesse bereits auf das Testergebnis einwirken:

> *„Messen die Aufgabentypen phonologische Bewusstheit oder bereits schriftsprachlich verankertes Wissen über die Graphemstruktur eines Wortes oder vielleicht beides oder noch etwas ganz anderes?“ (Jansen 2008, S. 185)*

Jansen fordert die Messung der Konfundierung der Tests, also inwieweit die Kinder ihre Strategien aufgrund von Kenntnissen der Schriftsprache schon verändert haben.

Verschiedene empirische Studien konnten zeigen, dass Kinder die Bewusstheit für Silben zu einem früheren Zeitpunkt als die Bewusstheit für Phoneme entwickeln (vgl. Gathercole & Baddeley 1993a, S. 131ff.; Liberman, Shankweiler, Fischer & Carter 1974, S. 201ff.; Fox & Routh 1975, S. 332ff.). Auch Ziegler und Goswami (2005, S. 4) konnten feststellen, dass bei Kindern in der Entwicklung der phonologischen Bewusstheit eine Entwicklung von großen zu kleinen Einheiten stattfindet:

> *“The emergence of phonological awareness can best be described along a continuum from shallow sensitivity of large phonological units to a deep awareness of small phonological units.”*

Die phonologische Bewusstheit im engeren Sinne ist dem eigentlichen Schriftspracherwerb näher als die phonologische Bewusstheit im weiteren Sinne, da sie erst im Schriftspracherwerb als relevante Leistung auftritt, während Reime und Silbensegmentierung schon in der täglichen Kommunikation der Kinder im Kindergarten Relevanz besitzt. In welchem genauen Zusammenhang die Fähigkeiten zur phonologischen Bewusstheit zum Erfolg im Schriftspracherwerb stehen, wird kontrovers diskutiert.

Die phonologische Bewusstheit im weiteren Sinn entwickelt sich scheinbar vor der phonologischen Bewusstheit im engeren Sinn. Die Übergänge sind dabei jedoch als fließend zu betrachten. Die Entwicklung der Vorschulkinder differiert stark. Stenzel (1999, zit. in Schnitzler 2008, S. 38) konnte die Ergebnisse englischer Untersuchungen (vgl. bspw. Carroll, Snowling, Hulme & Stevenson 2003, S. 923) bestätigen. Er konnte beobachten, dass sich die phonologische Bewusstheit von größeren phonologischen Einheiten zu kleineren phonologischen Einheiten entwickelt. Die Stichprobe, die für diese Untersuchung verwendet wurde, ist jedoch nicht repräsentativ (N = 17). Castles und Coltheart (2004, S. 77ff.) gehen davon aus, dass die Entwicklung des Lesens die Leistung in den Aufgaben zur phonologischen Bewusstheit steigert und erst der Erwerb der Graphem-Phonem-Korrespondenzen eine Grundlage für die Manipulation von Phonemen darstellt.

Schnitzler (2008, S. 35ff.) fasst in einer Überblicksstudie zusammen, dass die Strukturierung phonologischer Wortformen in abstrakte phonologische Einheiten ein Ergebnis der lexikalischen Restrukturierung bei der Wortschatzerweiterung und unabhängig vom Schriftspracherwerb ist, während die Identifikation der Einheiten in Phoneme aus dem Lernen der alphabetischen Strategie resultiert und eine weitere Ausdifferenzierung der phonologischen Wortformen dann durch den Erwerb der orthografischen Strategie vollzogen wird (vgl. a. a. O., S. 37). Die Entwicklung der phonologischen Bewusstheit beginnt meist im Vorschulalter und wird normalerweise erst Ende des vierten Schuljahres abgeschlossen (vgl. ebd.).

Wimmer, Landerl, Linortner und Hummer (1991, S. 243) schlagen auf der Basis ihrer Ergebnisse aus drei Studien (Studie 1: N = 50, Studie 2: N = 42 und Studie 3: N= 56) für die Klärung der Frage, ob die phonologische Bewusstheit als Konsequenz oder als Voraussetzung zu bewerten ist, die differenzielle Induktionshypothese vor. Es wird davon ausgegangen, dass Aufgaben zur phonologischen Bewusstheit im engeren Sinne, also auf der Phonemebene, auch von der Schriftsprache abgeleitet werden können (vgl. auch Schnitzler 2008, S. 60) und sie vermuten, dass die positiven Ergebnisse der schriftsprachlichen Kenntnisse nicht auf die Fähigkeiten im Vorschulalter, sondern auf den Erwerb der

Schriftsprache in den ersten beiden Schuljahren zurückzuführen ist. Sie bemängeln, dass in den Studien die Fähigkeiten der phonologischen Bewusstheit auf Phonemebene zu diesem Zeitpunkt als Variable nicht genauer erfasst werden. Die Autoren untersuchten Kinder des ersten Schuljahres, die bisher noch keinen Lese- und Schreibunterricht erhalten haben. Die Schüler konnten von sechs Items zur phonologischen Bewusstheit durchschnittlich nur 2,3 lösen. Frühe Leser dagegen konnten eine bessere Leistung vorweisen. Insbesondere der Misserfolg der Nicht-Leser in der Vokalersetzungsaufgabe lässt darauf schließen, dass dafür ein genauerer Einblick in die Graphem-Phonem-Korrespondenzen notwendig ist.
Aufgrund der Studien ist die phonologische Bewusstheit nicht eindeutig als Voraussetzung oder als Konsequenz des Schriftspracherwerbs einzuordnen. Die phonologische Bewusstheit scheint vom Erwerb der Schriftsprache beeinflusst zu sein und zum Teil mit dem Schriftspracherwerb einherzugehen. Die Trainingsstudien konnten jedoch deutlich zeigen, dass ein Training der phonologischen Bewusstheit im Vorschulalter von den Kindern zu besseren Leistungen im Lesen und Schreiben verhilft. Schnitzler (a. a. O., S. 65) rät davon ab, sich eindeutig für die Voraussetzungs- oder Konsequenzhypothese zu entscheiden. Es wäre ihrer Meinung nach

> *„[...] ein Fehlschluss, sich kategorisch entweder für die Voraussetzungs- oder für die Konsequenzhypothese zu entscheiden. Fähigkeiten der phonologischen Bewusstheit [...] sind nicht allein dazu in der Lage, einen erfolgreichen Schriftspracherwerb zu garantieren. Ebenso wenig kann man davon ausgehen, dass alle Kinder die explizite phonologische Bewusstheit auf der Phonemebene allein durch die aktive Auseinandersetzung mit der Schriftsprache erwerben. Beide Fähigkeiten entwickeln sich offensichtlich interaktiv.“*

Die grundlegenden Fähigkeiten der phonologischen Bewusstheit sind offensichtlich eine Voraussetzung, um im Erwerb der Schriftsprache erfolgreich zu sein. So glauben auch Tunmer und Hoover (1992, S. 175ff.), dass aufgrund dieses Fundaments und durch aktive Beschäftigung mit der Schriftsprache die Fähigkeiten erweitert und gestärkt werden und dies eine gute Voraussetzung für das Erlernen des Lesens darstellt. Implizite Fähigkeiten werden durch den aktiven Prozess explizit ins Bewusstsein gerufen.

Zusammenfassend konnten Korrelationsstudien nachweisen und Trainingsstudien bestätigen, dass sich durch gute Fähigkeiten in der phonologischen Bewusstheit Erfolge im Schriftspracherwerb vorhersagen lassen. Anhand von Gruppenvergleichen konnte gezeigt werden, dass Schwierigkeiten in der Schriftsprache bei Kindern und Erwachsenen auch auf unzureichende Kenntnisse der phonologischen Bewusstheit zurückgeführt werden können.

Phonologische Bewusstheit ist eine notwendige, aber nicht zwingend hinreichende Bedingung für einen erfolgreichen Schriftspracherwerb (vgl. Küspert 1998, S. 87). Es ist zu vermuten, dass ein Teil der phonologischen Bewusstheit als Voraussetzung für einen erfolgreichen Schriftspracherwerb gilt, ein Teil davon aber auch während des Schriftspracherwerbs erworben wird und somit eine Interaktion zwischen schriftsprachlichen Leistungen und phonologischer Bewusstheit besteht. Dies konnten auch Befunde von Tunmer und Nesdale (1985, S. 417ff.) bestätigen, in der Kinder schwache Leistungen im Pseudowortlesen erbrachten, obwohl sie über gute Fähigkeiten in der Phonemsegmentierung verfügten, aber wo andererseits kein Kind ohne die Fähigkeit der Phonemsegmentierung gute Leistungen im Pseudowortlesen vorweisen konnte. Auch Küspert (1998, S. 87) beschreibt, dass eine Unterscheidung zwischen der phonologischen Bewusstheit im weiteren und im engeren Sinne sinnvoll ist und die Gründe anhand der Untersuchungen auch gezeigt werden konnten. Die phonologische Bewusstheit auf Phonembasis ist eine wichtige Voraussetzung für gute schriftsprachliche Fähigkeiten in der Grundschule, während die phonologische Bewusstheit auf Silben- oder Wortebene eher für die orthografische Ebene als Voraussetzung gilt und damit ungefähr ab Ende des zweiten Schuljahrs Relevanz besitzt. Die Graphem-Phonem-Korrespondenzen werden mehr und mehr durch orthografische Regelmäßigkeiten ergänzt bzw. ausgetauscht, sodass der Zugriff auf größere phonologische Einheiten von Nöten ist (vgl. Abb. 12).

Schneider und Näslund (1999, S. 129) vermuten, dass sich ein Training phonologischer Bewusstheit im Vorschulalter oder in der ersten Klasse sowohl für gefährdete Kinder als auch für andere Kinder als positiv erweist. Das Training der phonologischen Bewusstheit ist ihrer Meinung nach eine wirksame Präventionsmaßnahme zur Vermeidung von Problemen im Schriftspracherwerb. Die Zusammenhänge der phonologischen Bewusstheit sind spezifisch in Bezug auf den Schriftspracherwerb und erweisen sich als irrelevante Einflüsse auf andere Leistungsbereiche der Kinder (vgl. Bradley & Bryant 1983, S. 419).

Lenel (2005, S. 169) konnte jedoch bei einem Vergleich von Trainingsprogrammen zeigen, dass die nachhaltigsten Effekte für das Lesen und Schreiben durch ein vorschulisches Förderprogramm erzielt werden konnten, das die direkte Schrifterfahrung integrierte. Sie schließt daraus, dass die phonologische Bewusstheit nur dann ein Prädiktor ist, wenn sie auf dem Schriftspracherwerb beruht. Auch Franzkowiak (2008, S. 163ff.) konnte keine Effekte für eine phonologische Förderung feststellen. Dies könnte jedoch auch von einem alphabetischen Anfangsunterricht abhängig sein.

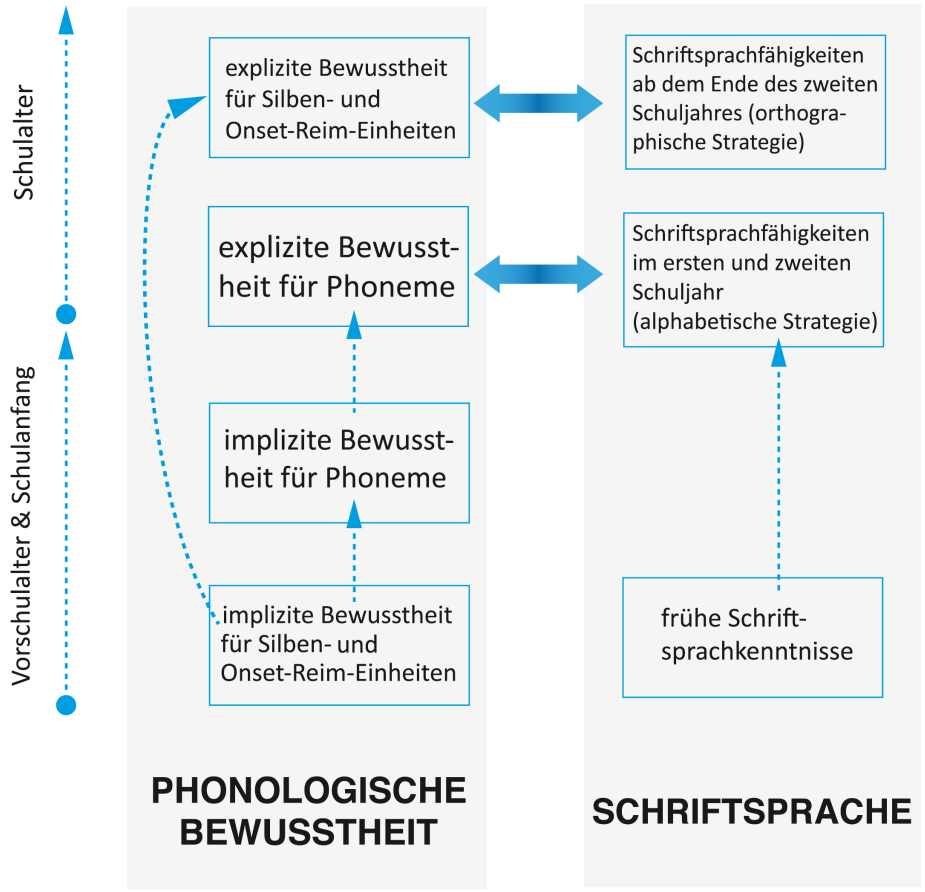


Abb. 12: Zusammenhänge zwischen der phonologischen Bewusstheit und schriftsprachlichen Fähigkeiten (nach Schnitzler 2008, S. 75)

Valtin (2010, S. 3) fasst aufgrund dieser gegenläufigen Ergebnisse zusammen:

> *„Der Nutzen einer expliziten und isolierten Förderung der phonologischen Bewusstheit im Vorschulalter konnte nicht nachgewiesen werden. Eine Förderung muss vielmehr direkt beim Schriftspracherwerb, das heißt in der Schule, ansetzen.“*

Kritisch angemerkt hat dies schon Walter (2002a, S. 48) in seinem Überblick über die Meta-Analyse von Ehri, Nunes, Schuster, Yaghoub-Zadeh und Shananhan (2001), in dem er feststellte, dass das phonologische Wissen nur 6,5 % der Varianz der Leseleistung aufklären kann. 10 % der Lesefähigkeit können erklärt werden, wenn das Trainingsprogramm mit Buchstabenwissen ergänzt wird (vgl. ebd.). Die Varianzaufklärung bei Vorschülern liegt bei 28 % und bei Risikokindern bei 31 %. Walter (ebd.) konstatiert: „Es bleibt also noch sehr viel Raum für andere Komponenten, die am Lese- und Schreibprozess beteiligt sind!“ Grundsätzlich sieht er die Wirksamkeit des phonologischen Trainings bestätigt, die

Instruktion zur phonologischen Bewusstheit sollte jedoch nicht primäres Ziel sein, sondern Mittel und Zweck für den Schriftspracherwerb (vgl. ebd.).

Abschließend beschreibt Schneider (1993, S. 200) die reziproke Beziehung zwischen der phonlogischen Bewusstheit und der Lese-Rechtschreibleistung wie folgt:

> *"[...] a majority of researchers seem to accept a view that emphasizes the reciprocal nature of the process: phonological awareness helps the beginning reader to take advantage of the alphabetic system at the same time the being taught to read may help to develop phonological awareness."*

Bezüglich des Einflusses der phonologischen Bewusstheit auf Rechtschreib- oder Leseleistungen weisen aktuelle Studien darauf hin, dass die phonologische Bewusstheit nicht der einzige relevante Faktor sein kann. Sie kann aufgrund der deutschen Orthografie scheinbar besser die Schwierigkeiten im Rechtschreiben als im Lesen vorhersagen (vgl. Wimmer & Mayringer 2002, S. 274). Kinder, die im Vorschulalter Schwierigkeiten in der phonologischen Bewusstheit aufwiesen, hatten im späteren Verlauf Defizite in ihrer Rechtschreibleistung, das Lesen war jedoch nicht beeinträchtigt. Andere Schüler, die später eine defizitäre Leseleistung aufwiesen, hatten dagegen im Vorschulalter keine Schwierigkeiten in der phonologischen Bewusstheit. Die Schüler mit einer defizitären Leseleistung hingegen hatten eine stark verlangsamte Benennungsgeschwindigkeit, die scheinbar im kausalen Zusammenhang mit einer verlangsamten Lesegeschwindigkeit steht. Die Schüler mit kombinierter Lese- und Rechtschreibschwäche wiesen in beiden Bereichen, der phonologischen Bewusstheit und in der Benennungsgeschwindigkeit, Defizite auf. Es scheint also, dass die beiden Bereiche auf die Lese-Rechtschreibschwäche einwirken.

Aufgrund dieser Befunde wird im nächsten Abschnitt auf die Benennungsgeschwindigkeit bzw. das phonologische Rekodieren beim Zugriff auf das semantische Lexikon als zweite Teilfertigkeit der phonologischen Informationsverarbeitung eingegangen.

1.2.2 Phonologisches Rekodieren beim Zugriff auf das semantische Lexikon

Das phonologische Rekodieren beim Zugriff auf das semantische Lexikon (auch in der einschlägigen Literatur unter „Schnellem Benennen" bekannt) ist die Fähigkeit, schriftliche Symbole in Laute zu rekodieren und sich damit einen Zugang zum semantischen Lexikon zu verschaffen (vgl. Wagner & Torgesen 1987, S. 202ff.).

Eines der bekanntesten Theorien zur Worterkennung ist die „Zwei-Wege-Theorie“ oder auch „Dual-Route-Theory“, die damalig erstmals von Coltheart (1978) präsentiert wurde. Das Dual-Route-Modell von Coltheart (1978, zusammenfassend in Coltheart, Curtis, Atkins & Haller 1993, S. 598) wird von Scheerer-Neumann (1997b, S. 283) als „bottom-up“-Modell eingestuft, da die Kontextnutzung außer Acht gelassen wurde. Nach dem Dual-Route-Modell von Coltheart (1987, zusammenfassend in Coltheart, Curtis, Atkins & Haller 1993, S. 598ff. und zusammenfassend Roth 1999, S. 82) werden zwei unterschiedliche Zugangswege zum semantischen Lexikon unterschieden. Es beschreibt zwei Wege des Worterkennens, einen direkten (lexikalischen) und einen indirekten (nicht-lexikalischen) Weg (vgl. Abb. 13).

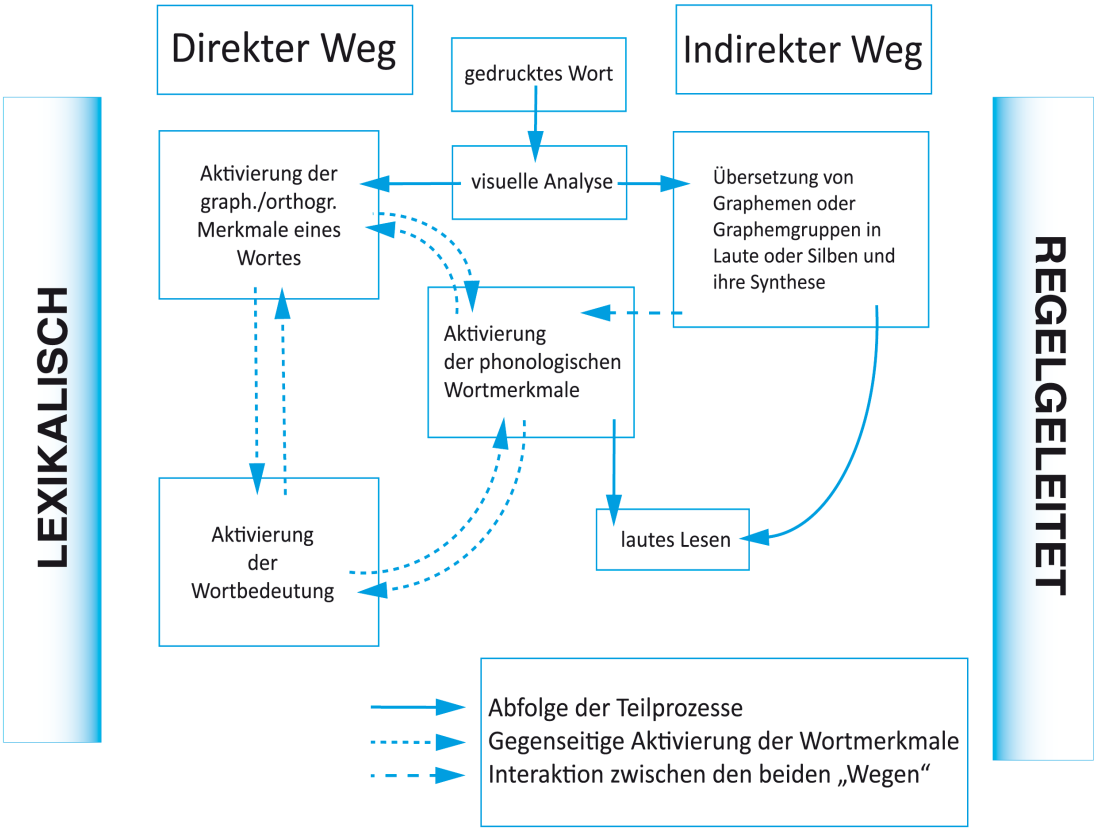


Abb. 13: Zwei-Wege-Modell (nach Scheerer-Neumann 1989b, S. 18)

Coltheart (zit. nach Coltheart & Rastle 1994, S. 1197ff.) geht von einem „inneren Lexikon“ aus, das alle semantischen, phonologischen und orthografischen Elemente des Wortschatzes beinhaltet und somit als eine Art Langzeitgedächtnis fungiert.

Beim *direkten Weg* erkennt der Leser das zu lesende Wort direkt anhand der grafischen Merkmale und kann es mit Hilfe des inneren Lexikons entschlüsseln. Somit wird die Wortbedeutung aktiviert und parallel dazu werden im inneren Lexikon die restlichen Bestandteile der Kodierung zugeordnet. Damit können die phonologischen Merkmale des

Wortes kodiert und das Wort gelesen werden. Auch beim orthografisch korrekten Schreiben setzt dieser Weg der Route voraus, dass das zu schreibende Wort im inneren Lexikon abgespeichert ist und abgerufen werden kann.

Beim *indirekten Weg* hingegen erkennt der Leser das Wort nicht direkt anhand des visuellen Reizes, da kein Eintrag im internen Lexikon vorliegt, und muss das Wort in einzelne Grapheme oder Graphemgruppen zerlegen. Das Wort wird erst am Ende des Lautierprozesses identifiziert. Anhand der im Arbeitsspeicher hinterlegten Graphem-Phonem-Korrespondenzen ordnet der Leser den identifizierten Graphemen die entsprechenden Phoneme zu. Das Wort kann bereits jetzt laut gelesen werden, ohne dass eine Sinnentnahme stattgefunden hat, da bisher kein Zugang zum inneren Lexikon stattgefunden hat.

In den 70er-Jahren konnte gezeigt werden, dass auch die Geschwindigkeit des Abrufs aus dem Langzeitgedächtnis für viele kognitive Komponenten hohe Korrelationen zu den Lese- und Rechtschreibleistungen aufwies (vgl. Denckla & Rudel 1974, S. 168ff.). Ein verlangsamter Abruf aus dem Langzeitgedächtnis kann die Arbeitsleistung des Kurzzeitgedächtnisses verringern, da die dort gespeicherten Inhalte bis zum Abruf der notwendigen Informationen aus dem Langzeitgedächtnis aufgrund der zu großen Zeitspanne nicht mehr vorhanden sind. Dies wiederum kann die Leseleistung beeinflussen (vgl. Mannhaupt 2008, S. 141).

Ist diese Fähigkeit des phonologischen Rekodierens beim Zugriff auf das semantische Lexikon nicht ausreichend ausgeprägt, hat das Kind Schwierigkeiten Sätze oder Texte zu verstehen, da es die Phonem-Graphem-Korrespondenzen nicht abrufen kann. Die phonologische Rekodierungsfähigkeit kann schon im Vorschulalter erfasst werden. Im Bielefelder Screening wurde dies zur Prognose einer Lese-Rechtschreibschwäche anhand der Geschwindigkeit beim Benennen von Farben zu Objekten bzw. auch an Farben zu farbig inkongruenten Objekten erfasst (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002, S. 11). Es wird die Zeit beim Benennen von Farben gemessen, in der Annahme, dass die Zeit die phonologische Repräsentation und die hohe Frequenz der Abrufbarkeit symbolisiert. Inwieweit das „schnelle Benennen“ z. B. von Objekten im direkten Zusammenhang mit dem Lesen steht, ist umstritten.

Das schnelle Benennen („speech rate“) konnte neben der phonologischen Bewusstheit auch von Muter und Snowling (1998, S. 332) als wichtiger spezifischer Prädiktor (prädiktive Werte für die Mathematik konnten ausgeschlossen werden) für die Lesegenauigkeit im Alter

von neun Jahren identifiziert werden. Aufgrund der geringen Stichprobengröße (N = 34) kann dieses Ergebnis jedoch nur als Tendenz gewertet werden.

Auch Schatschneider, Fletcher, Francis, Carlson und Foorman (2004, S. 279f.) bestätigten diesen Befund in ihrer Untersuchung mit 384 Kindern. Sie konnten feststellen, dass das schnelle Benennen von Buchstaben der beste Prädiktor der Leseflüssigkeit war (r = .51) und die Korrelation besser war als die des Benennens von Objekten (r = .40) am Ende der ersten Klasse. Das schnelle Benennen von Objekten und Buchstaben in Kombination mit der phonologischen Bewusstheit und dem Buchstabenwissen ergaben die höchste Korrelation (r = .87).

Moll, Fussenegger, Willburger und Landerl (2009, S. 17ff.) untersuchten in drei unabhängigen Studien an die 1 200 Kinder bezüglich des prädiktiven Wertes der phonologischen Bewusstheit und des „rapid automized naming“ (RAN), des schnellen Benennens. Sie konnten feststellen, dass RAN mehr die Varianz des Wort- (14 – 40 %) und Pseudowortlesens (15 – 32 %) vorhersagt als die phonologische Bewusstheit, während die phonologische Bewusstheit ein besserer Prädiktor für die Rechtschreibleistungen ist und RAN auf diese Fähigkeit fast keinen Einfluss hat.

Wolf und Bowers (1999, S. 423ff.) fanden in einer englischsprachigen Studie heraus, dass das schnelle Benennen und die phonologische Bewusstheit voneinander unabhängige Komponenten sind und dass bei einer Lese-Rechtschreibschwäche ein phonologisches Defizit oder ein Defizit im schnellen Benennen oder beides gleichzeitig vorliegen kann (vgl. Tab. 23, Doppel-Defizit-Hypothese).

Tab. 23: Die Subgruppen der „Double-Deficit“-Hypothese (nach A. Berglez 2002, S. 41)

		Phonologische Bewusstheit	
		Kein Defizit	Defizit
Schnelles Benennen	kein Defizit	keine Leseschwierigkeiten	Leseschwierigkeiten
	Defizit	Leseschwierigkeiten	Erhebliche Leseschwierigkeiten

Nach diesem Modell könnten schriftsprachliche Kompetenzen durch die phonologische Bewusstheit und durch das schnelle Benennen unabhängig voneinander vorhergesagt werden (vgl. Klicpera, Schabmann & Gasteiger-Klicpera 2010, S. 162f.). Wolf und Obregon (1992, S. 224ff.) konnten in ihrer Längsschnittstudie (N = 89 Kinder von der Vorschule bis zur vierten Klasse der Grundschule) nachweisen, dass leseschwache Kinder auch in Bezug auf ihre Genauigkeit schlechter abschnitten als unauffällige Kinder.

Auch im deutschen Sprachraum wurden dazu Ergebnisse gefunden (vgl. Wimmer und Mayringer 1999, S. 141ff.), die nach Aussagen der Studie von Landerl (1996, S. 169f.) nicht handfest sind. Die Theorie lässt einen Zusammenhang vermuten, da in beiden Prozessen aufgrund visueller Informationen eine phonologische und eine semantische Repräsentation aktiviert wird (vgl. Schnitzler 2008, S. 16f.). Beim Benennen von Objekten ist es beispielsweise die Farbe, beim Benennen von Wörtern sind es beispielsweise die Silben. Das schnelle Benennen wird somit nicht als Teil der phonologischen Arbeitsprozesse verstanden, sondern ist ein allgemeiner Geschwindigkeitsfaktor, der nicht spezifisch wirkt (vgl. ebd.). Vukovic und Siegel (vgl. 2006, S. 25ff.) konnten in ihrer Metaanalyse feststellen, dass zusätzlich zu den phonologischen Defiziten auch die visuelle Verarbeitung eine Rolle spielt. Sie nehmen an, dass das „schnelle Benennen“ nie isoliert für Probleme im Schriftspracherwerb verantwortlich ist, sondern bei Kindern mit einem phonologischen Defizit je nach Ausprägung und Schwere des Defizits nicht nur die phonologische Bewusstheit, sondern auch die Benennungsgeschwindigkeit berührt sein kann.

Landerl und Wimmer (2008, S. 154ff.) konnten in einer achtjährigen Follow-Up-Studie mit 115 Kindern feststellen, dass der beste Prädiktor bis hin zur achten Klasse für die Leseflüssigkeit das „Schnelle Benennen“ (rapid automatized naming) und für das Recht-schreiben die phonologische Bewusstheit ist. Die langfristige Entwicklung der schriftsprachlichen Kompetenzen der Kinder wurde mehr vom „Schnellen Benennen“ als von der phonologischen Bewusstheit beeinflusst. Auch Ennemoser, Marx, Weber und Schneider (2012, S. 63f.) konnten in ihren zwei jüngsten längsschnittlichen Untersuchungen zwischen der Benenngeschwindigkeit und der Lesegeschwindigkeit eine Korrelation von $r = .43$ ($p < .01$, Studie 1) und $r = .28$ ($p < .01$, Studie 2) feststellen, sodass die Benennungsgeschwindigkeit damit in der Studie mit der phonologischen Bewusstheit die besten prädiktiven Aussagen erzielen konnte. Die Auswirkungen auf die Rechtschreibleistungen der ersten und zweiten Klasse waren mit $r = .31$ ($p < .01$, Studie 1) und $r = .11$ ($p < .01$, Studie 2) geringer.

Die Studien im Bereich der Benennungsgeschwindigkeit sind jedoch bisher nicht ausreichend, um sich ein genaues Bild über den Zusammenhang von schriftsprachlichen, phonologischen Fähigkeiten oder der Benennungsgeschwindigkeit zu machen (vgl. Roth 1999, S. 85, Schnitzler 2008, S. 17). Die Untersuchungen bieten keinen Aufschluss darüber, ob die Benennensleistung von Objekten eine Ursache oder eine Konsequenz der Lese-Rechtschreibschwäche ist (beispielsweise die Resultate der beiden Studien von Snowling, van Wagtendonk & Stafford 1988, S. 67ff.). Es bleibt festzuhalten, dass die Benenngeschwindigkeit insbesondere durch die Erkenntnisse von Ennemoser, Marx, Weber und Schneider (2012, S. 63f.) eine wichtige Komponente speziell für die Vorhersage der Lesegeschwindigkeit darstellt.

1.2.3 Phonetisches Rekodieren im Arbeitsgedächtnis

Die Kognitionspsychologie konnte schon vor vielen Jahren herausarbeiten, dass das Kurzzeitgedächtnis bei jeder kognitiven Informationsverarbeitung eine bedeutende Rolle spielt (vgl. Baddeley 1986, S. 169). Das Kurzzeitgedächtnis ist in der Menge der Einheiten sowie in der Zeit, in der eine Information behalten wird, beschränkt. Außerdem kann immer nur eine Aktivität zur gleichen Zeit durchgeführt werden.

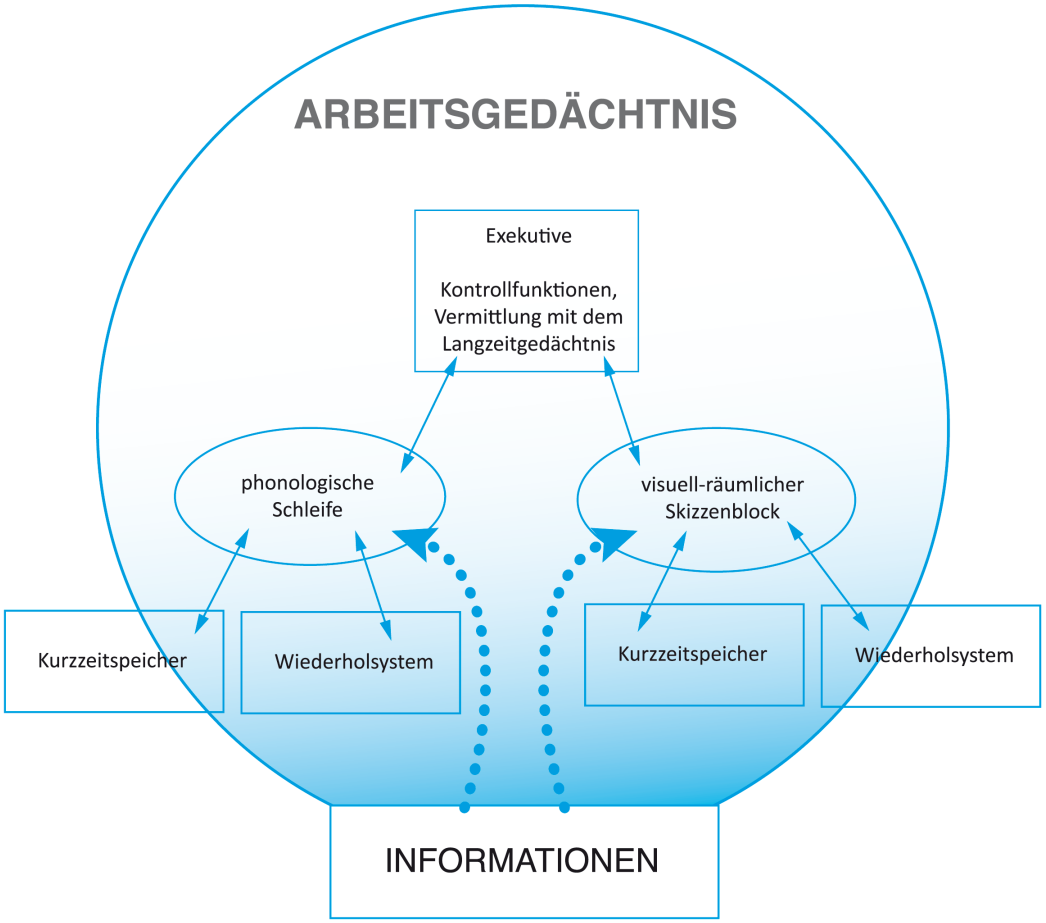
Es wurden in einigen Studien bei lese-rechtschreibschwachen Kindern Defizite im Kurzzeitgedächtnis festgestellt, jedoch lag keine grundsätzliche allgemeine Speicherschwäche bei diesen Schülern vor (vgl. Angermaier 1974, S. 27ff.). Auch Katz, Shankweiler und Libermann (1981, S. 474) konnten zeigen, dass sich lese-rechtschreibschwache Kinder in ihrer Leistung des Arbeitsgedächtnisses unterschieden. Sie legten den lese-rechtschreibschwachen und den normal-lesenden und -schreibenden Schülern abstrakte Bilder und Bilder mit konkreten Objekten vor. Bei den abstrakten Zeichnungen zeigte sich bei den beiden Gruppen kein Unterschied, während die benennbaren Abbildungen von beiden Gruppen unterschiedlich gut bewältigt worden sind. Somit kann davon ausgegangen werden, dass es sich nicht um ein generelles Problem bei der Speicherung handelt, sondern dass lese-rechtschreibschwache Schüler insbesondere bei phonologischen Einheiten ein Problem mit der Speicherung haben.

Ein relevanter spezifischer Prädiktor für die Lese-Rechtschreibleistungen ist das phonologische Rekodieren im Arbeitsgedächtnis. Kinder benötigen das Kurzzeitgedächtnis zum Behalten und Konstruieren der Wörter anhand kleinerer Einheiten (z. B. Silben). Für flüssiges Lesen und Schreiben ist es wichtig, dass die Schüler die bereits dekodierten Informationen so lange im Kurzzeitgedächtnis präsent halten, bis der Lese- oder Schreib-

prozess beendet ist. Bei einem nicht ausreichend ausgeprägtem Kurzzeitgedächtnis werden die Einzellaute vom Wortbeginn am Wortende nicht mehr erinnert (vgl. Barth & Gomm 2006, S. 13). Die Fähigkeit des kurzfristigen Präsenthaltens von Informationen ist besonders im Schriftspracherwerb eine bedeutende Kompetenz, da die Schüler zu Beginn des Lernprozesses viele Wörter noch durch einzelnes Lautieren der Buchstaben erlesen und diese dann zu Lautfolgen zusammenfügen. Wenn ein Kind beispielsweise eine kürzere Behaltensspanne als vier Slots (Einheiten) aufweist, dann bekommt es Schwierigkeiten Wörter zu lesen, bei denen mehr Einheiten behalten werden müssen.

Dem Kurzzeitgedächtnis wird in der Forschung schon länger die zentrale Rolle des „Arbeitsgedächtnisses“ (working memory) zugeschrieben (vgl. Atkinson & Shiffrin 1968, S. 89ff.), das Informationen aufnimmt und zum Langzeitgedächtnis transportiert. Die Informationen werden dann dort mit verfügbaren Wissenseinheiten verknüpft. Zur Erklärung der Rolle des Arbeitsgedächtnisses wird in der Schriftspracherwerbsforschung häufig das Modell von Baddeley (1992, S. 557 und Baddeley & Hitch 1974, S. 74) genutzt. Demnach sind drei Bereiche des Arbeitsgedächtnisses für die Fähigkeiten des Lesens und Rechtschreibens besonders wichtig. Die zentrale Exekutive (central executive) steuert die Weiterverarbeitung der Informationen im Arbeitsgedächtnis und besitzt nur eine begrenzte Kapazität. Der zentralen Exekutive gehören die Subsysteme „visuell-räumlicher Notizblock“ (visuo-spatial sketch pad) und die „phonologische Schleife“ (phonological loop oder auch auditory rehearsal loop) an (vgl. Abb. 14).

Der visuell-räumliche Notizblock ist für die Verarbeitung und Speicherung von optischen Informationen zuständig, die phonologische Schleife dagegen verarbeitet und speichert auditiv-verbale Informationen. Die phonologische Schleife ist nochmals in zwei Teilbereiche aufgeteilt. Im phonologischen Speicher sind phonologisch kodierte Informationen gespeichert. Zusätzlich dazu ist dort ein artikulatorischer Wiederholungsprozess (eine Art Tonbandendlosschleife) verankert, der das Präsenthalten von Informationen gewährleisten soll, bevor sie im Kurzzeitspeicher vergessen werden. Über die Wiederholung gelangen die Informationen in den phonologischen Speicher (vgl. Abb. 14).



Anmerkung: Die Aufnahme ist auf max. 7 Einheiten begrenzt. Bei gleichzeitiger Verarbeitung reduziert sich die Verarbeitung auf 2-3 Einheiten.

Abb. 14: Struktur des Arbeitsgedächtnisses nach Baddeley (nach Wellenreuther 2005, S. 81)

Baddeley (2003, S. 190ff.) ergänzte das Modell später noch um eine weitere Komponente (vgl. Abb. 15), den episodischen Puffer (episodic LTM), der für die Informationsspeicherung zuständig ist. Er hängt stark von der zentralen Exekutiven ab und nutzt multidimensionale Codes, um Repräsentationen des Arbeits- oder Langzeitgedächtnisses zu integrieren. Diese korrespondieren direkt mit der individuellen Erfahrung.

Individuelle Unterschiede in Teilen dieses Komplexes haben direkte unspezifische Effekte auf den Erwerb von Fähigkeiten oder auf die kognitive Entwicklung. Die phonologische Schleife dagegen ist eine spezifische Teilkomponente für den Sprach- und Schriftspracherwerb, da die Auswirkung auf mathematische Leistungen widerlegt werden konnte (vgl. Krajewski 2003, S. 19ff.).

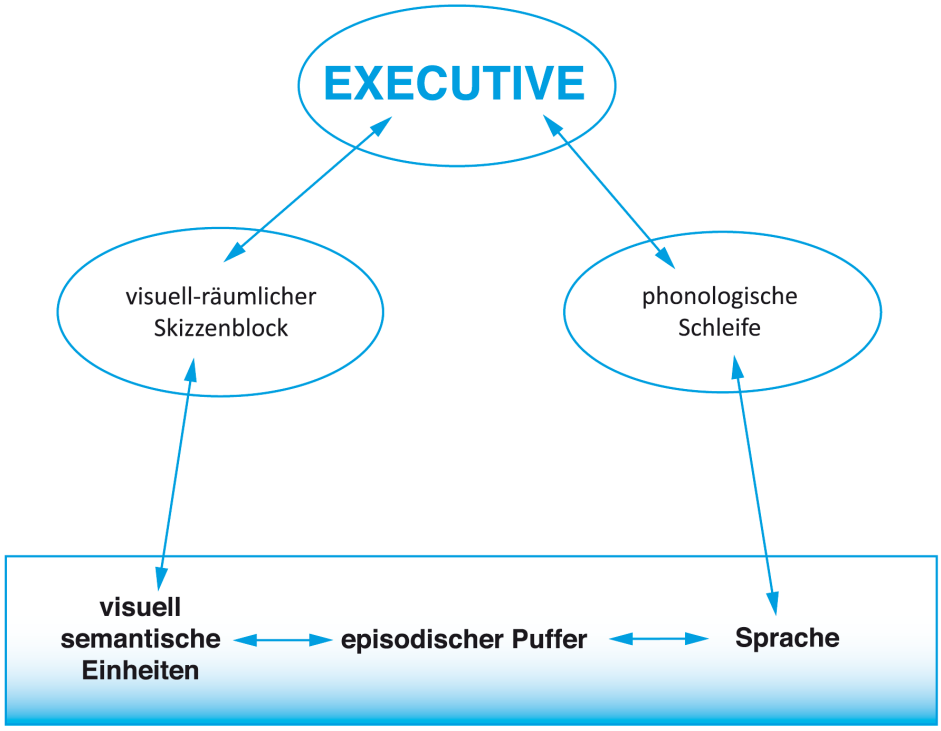


Abb. 15: Das Multikomponentenmodell von Baddeley 2003 (übersetzt aus dem Englischen, verändert nach Hippmann 2008, S. 45)

Die Autoren Hasselhorn, Grube und Mähler (2000, S. 174) schlagen aufgrund ihrer Studien vor, das Modell von Baddeley um zwei Komponenten für den Rehearsalprozess zu erweitern. Sie wollen anhand des Automatisierungsgrades bzw. der Aktivierungsleichtigkeit des Rehearsal-Prozesses und der Geschwindigkeit der subvokalen Teilprozesse aufklären, warum einige Kinder mit phonologischen Defiziten Lernbehinderungen und andere Kinder mit den gleichen Defiziten Sprachentwicklungsstörungen ausbilden.

Den phonetischen Speicher wollen sie zudem in zwei Komponenten aufteilen. Zum einen kann man über das Nachsprechen von Pseudowörtern die Größe dieses Speichers messen, zum anderen lässt sich darüber aber auch die Verarbeitungspräzision feststellen (vgl. Abb. 16).

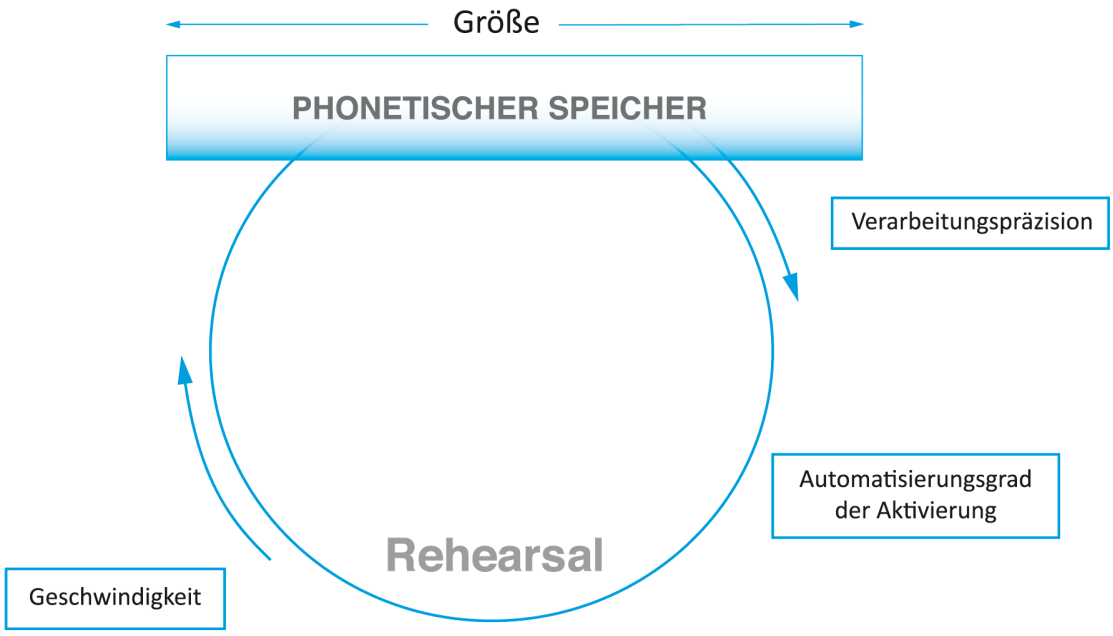


Abb. 16: Erweitertes Zwei-Komponenten-Modell des phonologischen Arbeitsgedächtnisses von Baddeley (nach Hasselhorn, Grube & Mähler 2000, S. 174)

In Tabelle 24 sind die Ergänzungskomponenten sowie deren empirische Indikatoren nochmals aufgezählt. Diese Überlegungen können nach Aussage der Autoren bisher ausschließlich als ein Modell verstanden werden und müssen noch durch einschlägige Untersuchungen bestätigt werden.

Tab. 24: Komponenten der phonologischen Schleife, hypothetisch zu den funktionalen Merkmalen und empirischen Indikatoren zugeordnet (nach Hasselhorn, Grube & Mähler 2000, S. 176)

Komponente	Funktionales Merkmal	Empirische Indikatoren
Phonetischer Speicher	Größe	Kunstwörter Nachsprechen
	Verarbeitungspräzision	Akustischer Ähnlichkeitseffekt Effekt irrelevanter Sprache (Kunstwörter Nachsprechen)
Subvokales Rehearsal	Geschwindigkeit	Sprechrate
	Automatisierungsgrad der Aktivierung	Wortlängeneffekt (Effekt artikulatorischer Unterdrückung)

Bee-Göttsche (1993, S. 185ff.) und Küspert (1998, S. 188f.) konnten in ihren wissenschaftlichen Studien zeigen, dass das Arbeitsgedächtnis eine beträchtliche Bedeutung für den

Schriftspracherwerb trägt. Bee-Göttsche (1993, S. 183) konzipierte eine Trainingsstudie mit 75 Kindergartenkindern und führte eine Förderung des Kurzzeitgedächtnisses durch. Die Effekte zeigten Auswirkungen auf die phonemische Bewusstheit der Kinder. Schneider und Näslund (1992, S. 267ff.) konnten herausfinden, dass eine signifikante Beziehung zwischen der Arbeitsgedächtnisleistung im letzten Jahr vor Schulbeginn und den Leistungen im Schriftspracherwerb besteht.

Shankweiler, Liberman, Mark, Fowler und Fischer (1979, S. 531) konnten durch einen Gruppenvergleich mit Zweitklässlern nachweisen, dass Kinder mit Lese-Rechtschreibschwäche, bei denen ein Gedächtnisdefizit ausgeschlossen wurde, deutlich schlechtere Leistungen bei der phonetischen Rekodierung von Informationseinheiten hatten als Kinder mit normalen Lese-Rechtschreibleistungen. Sie vermuten, dass gute und schlechte Leser das Arbeitsgedächtnis auf unterschiedliche Art und Weise nutzen. Die drei Untersuchungsgruppen bestanden jedoch nur aus N = 17, N = 16 und N = 13 Kindern.

Die genaue Rolle der phonologischen Rekodierung in Verbindung mit schriftsprachlichen Leistungen konnte bisher nicht eindeutig geklärt werden. Gruppenvergleiche konnten zwar mehrfach nachweisen (vgl. Liberman, Mann, Shankweiler & Werfelmann 1982, S. 367ff.), dass lese-rechtschreibschwache Kinder nachweislich Schwierigkeiten in der phonologischen Rekodierung aufweisen, jedoch konnten die kausalen Zusammenhänge im Detail noch nicht ermittelt werden. Johnston (1982, S. 455) untersuchte 9-, 12- und 14-jährige leseschwache Kinder und verglich diese mit einer Kontrollgruppe chronologischen Alters und einer weiteren hinsichtlich des Lesealters. Die leseschwachen Kinder schnitten durchweg schlechter ab als die Kinder mit durchschnittlichen Leseleistungen mit gleichem chronologischem Alter. Es wurde jedoch kein Unterschied zu den Kindern mit gleichem Lesealter vermerkt. Die Ergebnisse der Untersuchung sind nicht eindeutig zu interpretieren. Die Gedächtnisleistungen könnten das Erlernen des Lesens fördern, umgekehrt wäre es allerdings auch möglich. Gathercole und Baddeley (1993b, S. 259ff.) vermuten eine beidseitige Einwirkung der Variablen. Schuchardt, Kunze, Grube und Hasselhorn (2006, S. 265ff.) konnten bei einer Stichprobe von N = 232 Drittklässlern herausfinden, dass bei Kindern mit Minderleistungen in der Schriftsprache sowohl phonologische als auch zentralexekutive Defizite zu beobachten waren. Es zeigten sich zu Kindern ohne Defizite keine Unterschiede im Bereich des visuell-räumlichen Arbeitsgedächtnisses. Die Autoren vermuten daraufhin, dass Kinder mit Schwierigkeiten im Schriftsprachbereich nicht ausschließlich Defizite im phonologischen Arbeitsgedächtnis, sondern umfassendere Defizite im Arbeitsgedächtnis aufweisen (vgl. a. a. O., S. 267).

Die Überprüfung der Fähigkeit des phonetischen Rekodierens im Arbeitsgedächtnis kann unter anderem durch das Wiederholen von Pseudowörtern (z.B. ba-re-mi-fuk) gemessen werden. Gathercole und Baddeley (1993a, S. 70ff.) gehen davon aus, dass effiziente Arbeitsgedächtnisprozesse den Schriftspracherwerb begünstigen, dass der Schriftspracherwerb zudem aber auch die Steigerung und Effizienz der Arbeitsgedächtnisprozesse voranbringt. Insbesondere im Alter von sechs bis zehn Jahren beobachteten sie ein enormes Wachstum in Bezug auf die Dauer des Präsenthaltens und der Menge an Informationseinheiten, die gespeichert werden können (vgl. ebd.). Gathercole und Baddeley (a. a. O., S. 71ff.) legten dar, dass das Nachsprechen von Pseudowörtern signifikant einen Teil der Varianz der Leseleistungen im fünften Schuljahr aufklärt. Hasselhorn, Seidler-Brandler und Körner (2000, S. 119ff.; vgl. auch Hasselhorn, Grube & Mähler 2000 S. 175f.) konnten in zwei Studien ($N = 26$ und $N = 40$) diesen Befund bestätigen. Längere Pseudowörter konnten schlechter nachgesprochen werden als kürzere („akustischer Ähnlichkeitseffekt“, vgl. auch Hasselhorn, Grube & Mähler 2000, S. 174) und es ergab sich eine positive Korrelation zwischen der unabhängig erhobenen Wortspanne und der Leistung im Nachsprechen. Beides spricht für einen Einfluss des phonologischen Arbeitsgedächtnisses auf die Prozesse. Außerdem wurde festgestellt, dass wortähnliche Pseudowörter einfacher reproduziert wurden als wortunähnliche Pseudowörter. Somit scheint ein lexikalischer Einfluss auf den Behaltensprozess plausibel zu sein. Aus einer weiteren Untersuchung konnten Hasselhorn, Tiffin-Richards, Woerner, Banaschewski und Rothenberger (2000, S. 163) zusätzlich folgern, dass die Funktionstüchtigkeit des Speichers im phonologischen Arbeitsgedächtnis auch für die Diagnose von LRS relevant scheint. Dies konnte auch Goswami (2001, S. 70) bestätigen.

Insgesamt existieren in diesem Bereich wesentlich weniger empirische Arbeiten, die das phonologische Rekodieren beim Zugriff auf das Arbeitsgedächtnis als prognostische Variable untersuchen als bei der phonologischen Bewusstheit. Ebenso wie bei der phonologischen Bewusstheit ist im Bereich des phonetischen Rekodierens im Arbeitsgedächtnis die Frage nach Ursache und Wirkung noch nicht geklärt. Gathercole und Baddeley (1993a, S. 194ff.) schlagen eine integrative Sichtweise vor, bei der sich die beiden Prozesse gegenseitig bedingen.

Zur Vorhersage von Lese-Rechtschreibfähigkeiten konnten Muter und Snowling (1998, S. 332) in ihrer längsschnittlichen Untersuchung als beste langfristige Prädiktoren für neunjährige Kinder die Aufgaben zum Auslassen von Phonemen und zur Wiederholung von Pseudowörtern im Alter von fünf bis sechs Jahren identifizieren. Sie vermuten jedoch aufgrund ihrer Studie, dass die Wiederholung von Pseudowörtern erst für die langfristigen

Prozesse als Vorhersage gültig ist, während es sich auf die kurzfristigen Prozesse nicht sonderlich auswirkt. Die Stichprobengröße (N = 34) ist hingegen nicht repräsentativ. Swanson und Berninger (1996, S. 358) maßen in einer Studie an ca. 300 Schülern der vierten bis sechsten Klasse die Arbeitsgedächtniskapazität durch Aufgaben zum Schreiben und Nachsprechen von Wörtern, Pseudowörtern, Buchstaben oder Folgen von Buchstaben. Sie konnten mit der Gedächtniskapazität 4 bis 12 % der Varianz im Rechtschreiben aufklären. Die Gedächtniskapazität bewerteten sie als signifikanten Prädiktor der Rechtschreibleistung.

In der Studie von Cutting und Denckla (2001, S. 673ff.) dagegen konnte das Arbeitsgedächtnis als prädiktiver Wert nicht bestätigt werden. Sie fanden einen signifikanten Zusammenhang zwischen der Gedächtnisspanne und der phonologischen Bewusstheit heraus. Diese konnte auch keinen signifikanten Beitrag zum Wortlesen leisten. Die beiden Komponenten scheinen sich nach Meinung der Autoren zu überlappen. Windfuhr und Snowling (2001, S. 160ff.) kommen zu einem ähnlichen Resultat.

Alloway, Gathercole, Willis und Adams (2004, S. 85) interpretieren die Resultate ihrer Untersuchung anders und glauben, dass phonologische Bewusstheit und die Kapazität des Arbeitsgedächtnisses für unterschiedliche Prozesse zuständig sind. Das phonologische Kurzzeitgedächtnis ist diesen Autoren zufolge eine wichtige Komponente zum Erlernen von Graphem-Phonem-Korrespondenzen und für die Speicherung von phonologischen Sequenzen vor dem Output während des Rekodierungsvorgangs zuständig. Die phonologische Bewusstheit dagegen unterstützt die Segmentation der phonologischen Repräsentation von zu schreibenden Wörtern (vgl. auch Alloway, Gathercole, Adams, Willis, Eaglen & Lamont 2005, S. 422). Die Trennung dieser Komponenten konnten Roebers und Zoelch (2005, S. 119) schon bei vierjährigen Kindern ebenfalls nachweisen.

Die Studien sprechen insgesamt für einen Einfluss des phonetischen Rekodierens im Arbeitsgedächtnis auf die Leistungen im Lesen und Schreiben. Bei dieser Komponente verhält es sich jedoch anders als bei der phonologischen Bewusstheit. Die Autoren gehen nicht davon aus, dass ein Training zur effizienteren Speicherung Einfluss auf die Komplexität des Lese- und Schreibprozesses hat (vgl. Torgesen, Kistner & Morgan 1987 aus Roth 1999, S. 88). Trotz der positiven Ergebnisse darf außerdem nicht davon ausgegangen werden, dass die Arbeitsgedächtnisleistung allein die Entstehung einer Lese-Rechtschreibschwäche erklärt, da es auch Untersuchungen gab, die das Arbeitsgedächtnis nicht als Prädiktor identifizierten und der Anteil der Aufklärung der Varianz noch weitere Faktoren vermuten lässt. Hippmann (2008, S. 48) geht davon aus, dass die unterschiedlichen Befunde auch aufgrund der Problematik der unterschiedlichen Erfassung der Kapazität des

Arbeitsgedächtnisses entstanden sind. Die zentrale Exekutive arbeitet als modalitätsunabhängiges Aufmerksamkeitskontrollsystem. Die artikulatorische Schleife dagegen ist für die Speicherung von sprachbasierter Information zuständig (vgl. ebd., S. 47). Bei einigen Verfahren kommt es zu einer Vermischung dieser beiden Eigenschaften. Wenn die Kinder Wörter, Buchstaben, Zahlen oder Pseudowörter wiederholen, dann wird ausschließlich die artikulatorische Schleife angesprochen, sollen aber bspw. Zahlen rückwärts aufgesagt werden, wird automatisch auch die zentrale Exekutive verwendet. Es ist also zu empfehlen, einer Vermischung der beiden Systeme vorzubeugen. Außerdem ist die Arbeitsgedächtnisleistung zusätzlich abhängig von der jeweiligen Stufe im Schriftspracherwerbsprozess, in der sich das jeweilige Kind in seiner Entwicklung befindet.

1.2.4 Wechselseitige Abhängigkeiten der Komponenten der phonologischen Informationsverarbeitung

Die phonologische Informationsverarbeitung bietet ein spezifisches Konstrukt, auf dessen Basis die Prävention und Intervention von Schwierigkeiten im Schriftspracherwerb stattfinden.

Fähigkeiten zur *phonologischen Bewusstheit* scheinen eine notwendige Bedingung für Erfolg im Schriftspracherwerb zu sein. Jedoch sind sie keine hinreichende Bedingung, da andere Faktoren, wie Intelligenz, sozialer Hintergrund, Beschulung usw. ebenfalls von Gewicht sind, auch wenn diesen Faktoren ein weitreichend geringeres Maß an Bedeutung zukommt (vgl. Kap. II.2: Endogene und exogene Einflüsse auf die Schülerleistungen).

Bei der Fähigkeit des *phonologischen Rekodierens beim Zugriff auf das semantische Lexikon* gibt es einen Großteil an Studien, die der Vorhersagekraft des Prädiktors zusprechen. Es gibt jedoch auch Gegenstimmen (vgl. Kap. II.1.2.2) und insgesamt nur relativ wenige Studien im Vergleich der Vielzahl an Untersuchungen zur phonologischen Bewusstheit.

Das *phonetische Rekodieren im Arbeitsgedächtnis* dagegen wurde anhand eines sinnvollen theoretischen Modells erstellt und durch zahlreiche Studien als Prädiktor identifiziert. Der Nachteil ist, dass sich die Fähigkeit wahrscheinlich nicht isoliert trainieren lässt und somit nur der Identifizierung von Risikokindern zuträglich sein kann.

Schneider und Näslund (1999, S. 126ff.) zeigten, dass die Fähigkeiten der phonologischen Informationsverarbeitung 70 % der Varianz in den Rechtschreibleistungen in der Grundschule erklären können. Die phonologische Bewusstheit ist der stärkste dieser drei

Faktoren, da ihm prognostisch die größte Bedeutung zugemessen werden kann. Wagner (1988, S. 261ff.) erstellte aus neun korrelativen Längsschnittstudien und aus sieben Trainingsstudien eine Meta-Analyse. Er überprüfte die kausalen Beziehungen zwischen den Komponenten der phonologischen Informationsverarbeitung. Anhand einer Pfadanalyse konnte er feststellen, dass 75 % der Varianz der Wortleseaufgabe durch die Komponenten der phonologischen Informationsverarbeitung aufgeklärt werden konnten. Für jede einzelne Variable konnte er auch nach Kontrolle der Intelligenz einen direkten signifikanten Zusammenhang zum Kriterium herstellen. Eine großangelegte Studie fand auch im Bielefelder Raum zur Konstruktion des Bielefelder Screenings statt (vgl. dazu Kap. I.2.2.1). Die Autoren (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002, S. 48f.) konnten für die Komponenten der phonologischen Informationsverarbeitung einen RATZ-Index für alle Lese-Rechtschreibleistungen zwischen 34 % und 66 % ermitteln, was laut Marx (zit. a. a. O., S. 47) im guten Vorhersagebereich liegt. Diese Ergebnisse konnten jedoch nicht repliziert werden (vgl. Marx & Weber 2006, S. 256ff.).

Die Forschungsgruppe der LOGIK-Studie des Max-Planck-Instituts in München analysierte Daten von 121 Kindern, um die Interdependenzen der Komponenten der phonologischen Informationsverarbeitung zu erfassen. Kognitive Leistungen, phonologische Bewusstheit, Rekodiergeschwindigkeit, Gedächtniskapazität und frühe Schriftkenntnis wurden als Prädiktoren im Vorschulalter herangezogen. Die Kriteriumsleistung zum Leseverständnis und zum Rechtschreiben wurde am Ende des zweiten Schuljahres erhoben. Die Autoren berechneten nach der LISREL-Methode (Analysis of Linear Structural Relationships, Jöreskog & Sörbom, 1984) Strukturgleichungsmodelle, die aus einer Kombination von Faktoren- und Regressionsanalysen bestehen. Das Modell zum Leseverständnis (vgl. Abb. 17) konnte 47 % der Varianz des Kriteriums aufklären. Das Modell zum Rechtschreiben (vgl. Abb. 18) konnte sogar 62 % der Varianz des Kriteriums erklären. Im Unterschied zum Modell zum Leseverständnis hat der verbale IQ einen direkten Einfluss auf das Arbeitsgedächtnis und die phonologische Rekodierung im lexikalen Zugriff. Somit konnte ein starker indirekter Einfluss der Intelligenz vermerkt werden. In beiden Modellen wurde zudem kein direkter Einfluss des Arbeitsgedächtnisses auf die Kriteriumsleistungen identifiziert.

Somit vermerken Schneider und Näslund (1992, S. 271f.) folgende Erkenntnisse (vgl. dazu Roth 1999, S. 93):

- Die drei Komponenten der phonologischen Informationsverarbeitung wirken signifikant auf die Lese- und Rechtschreibleistungen ein.

- Des Weiteren sind verbale Intelligenz und frühe Schriftkenntnis ebenfalls wichtige Komponenten, um die Lese- und Rechtschreibleistungen zu erklären (vgl. hierzu Kap. II.1.3 und Abb. 17 und Abb. 18).
- Die Kausalmodelle von Leseverständnis und Rechtschreibung sind verschieden.

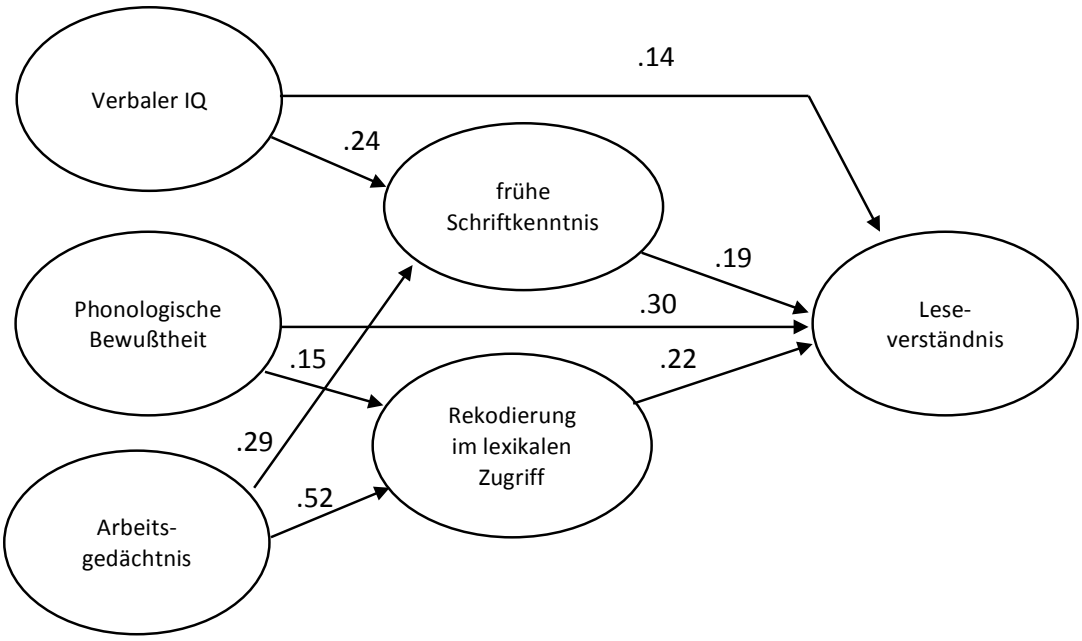


Abb. 17: Strukturgleichungsmodell, Kriterium: Leseverständnis (nach Schneider und Näslund 1992, S. 268)

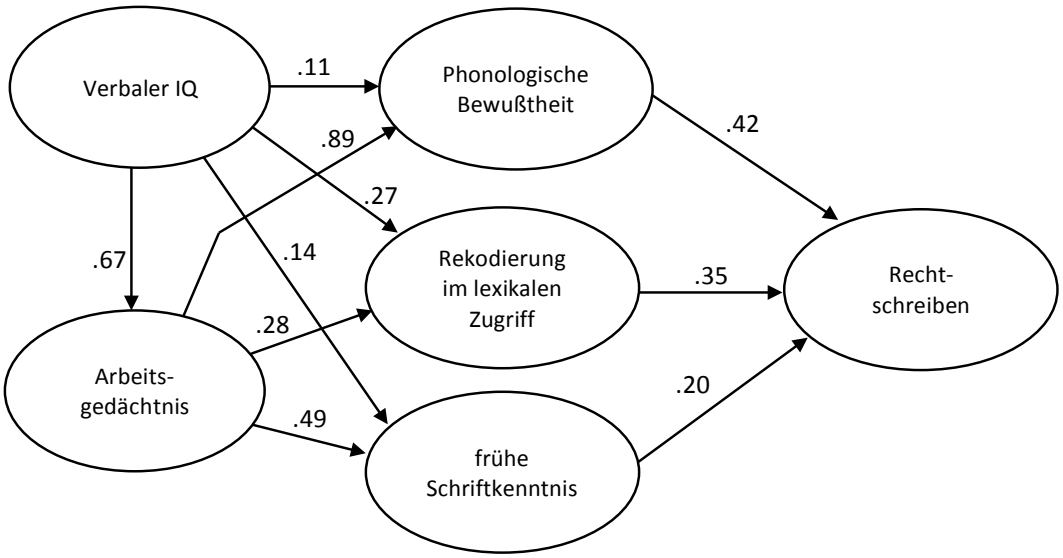


Abb. 18: Strukturgleichungsmodell, Kriterium: Rechtschreibung (nach Schneider und Näslund 1992, S. 270)

Auch Ennemoser, Marx, Weber und Schneider (2012, S. 63f.) konnten in ihren jüngsten längsschnittlich angelegten Studien einen korrelativen Zusammenhang zwischen dem Arbeitsgedächtnis und der phonologischen Bewusstheit (Studie 1 mit N = 165: r = .39; Studie 2 mit N = 175: r = .23) sowie der linguistischen Kompetenz (Studie 1 mit N = 165: r = .40; Studie 2 mit N = 175: r = .44), die wiederum korrelativ mit der Lesegeschwindigkeit und dem Textverständnis zusammenhängen. Das Strukturgleichungsmodell für die Vorhersage der Rechtschreibleistungen in der ersten Klasse kann ebenfalls einen korrelativen Zusammenhang zur phonologischen Bewusstheit (Studie 1 mit N = 165: r = .39; Studie 2 mit N = 175: r = .28) und zur linguistischen Kompetenz (Studie 1 mit N = 165: r = .45; Studie 2 mit N = 175: r = .42) herstellen, die sich wiederum auf die Rechtschreibleistungen auswirken. Ebenfalls die Benennungsleistung spielt eine zentrale Rolle und steht mit der Lesegeschwindigkeit in direktem korrelativen Zusammenhang (Studie 1 mit N = 165: r = .43; Studie 2 mit N = 175: r = .28), ebenso mit den Rechtschreibleistungen der ersten Klasse (Studie 1 mit N = 165: r = .31; Studie 2 mit N = 175: r = .11). So finden sich in den Strukturgleichungsmodellen zur Vorhersage von Lesegeschwindigkeit, Textverständnis und Rechtschreibkompetenz alle drei Komponenten der phonologischen Informationsverarbeitung als relevante Prädiktoren wieder.

Die Ergebnisse aller Studien sind jedoch im Netzwerk der vielen den Schriftspracherwerb beeinflussenden Faktoren zu betrachten, die unmöglich alle Berücksichtigung finden können. Zudem wurden viele der Studien im Bereich der phonologischen Informationsverarbeitung an englischsprachigen Kindern durchgeführt und es ist nicht eindeutig geklärt, ob sich dies auf den deutschen Sprachraum auf direktem Weg übertragen lässt. Prinzipiell ist zu beachten, dass bei der Diagnostik der phonologischen Bewusstheit immer auch falsche Prognosen geliefert werden können (vgl. Schnitzler 2008, S. 59). Insgesamt liefert sie ein großes Potenzial, Risikokinder zu identifizieren und eine entsprechende Förderung der Fähigkeiten möglichst vor dem Schuleintritt zu etablieren. Die phonologische Informationsverarbeitung scheint damit in Bezug auf schriftsprachliche Leistungen ein spezifischer Indikator zu sein. Es existieren jedoch weitere Indikatoren, die einen Einfluss auf die schriftsprachlichen Leistungen haben. In wissenschaftlichen Studien (z.B. Wagner & Torgesen 1987, S. 202ff. u. a.) konnte nachgewiesen werden, dass die phonologische Bewusstheit die schriftsprachlichen Fähigkeiten vorhersagen konnte, wenn andere Faktoren wie Intelligenz, Sprachvermögen, Hörverständnis oder sozialer Hintergrund beeinflussend auf die Untersuchung einwirkten. Dies und weitere mögliche Einflussfaktoren, die Gründe für die schlechten Prognoseergebnisse der beschriebenen Diagnostikverfahren sein können, werden in Kap. II.2 erklärt. Im Folgenden wird nun eine Auswahl weiterer möglicher Prädiktoren analysiert.

1.3 Buchstabenkenntnis

In US-amerikanischen Untersuchungen wird häufig die Buchstabenkenntnis als weiterer wichtiger Prädiktor in Früherkennungsprogramme und Fördermaßnahmen miteinbezogen. Die hier vorgestellten Programme berücksichtigen diese nur am Rande (Namen schreiben, Buchstaben visuell wiedererkennen, ...) bzw. die meisten gar nicht.
Um zu überprüfen, ob dies einen wichtigen Prädiktor darstellen könnte, wird die Buchstabenkenntnis im Vorschulalter in Bezug auf die Lese- und Rechtschreibleistungen im Schulalter beleuchtet. Die frühe Buchstabenkenntnis scheint eine Rolle bei der Prognose von Lese-Rechtschreibfähigkeiten im Grundschulalter zu spielen (vgl. Schneider & Küspert 2003, S. 114f.). Schneider und Näslund (1993, S. 273ff.) konnten zeigen, dass das Wissen in der phonologischen Bewusstheit unter anderem auch von der Buchstabenkenntnis der Kinder abhängig ist. Das Buchstabenwissen scheint einen Einfluss auf die phonologischen Fähigkeiten der Kinder zu haben. Muter und Snowling (1998, S. 332) dagegen stellten zwar ebenfalls die Buchstabenkenntnis als einen Prädiktor für den Erfolg im Lesen und Schreiben in der Grundschule fest, jedoch konnten sie diesen nicht als spezifisch identifizieren. Die Korrelationen zu den mathematischen Fähigkeiten der ersten Klasse waren ebenfalls signifikant. Sie führen dies auf die Nähe zum Zahlen Benennen zurück. Die Stichprobengröße betrug nur N = 34 Schüler und kann nicht als repräsentativ angesehen werden.

May und Okwumo (1999, S. 8ff.) verglichen in ihrer Studie 72 Vorschulkinder, die in zwei Experimental- und eine Kontrollgruppe aufgeteilt wurden. Die erste Trainingsgruppe wurde einem phonologischen Training unterzogen, der zweiten wiederum wurde ein Training zur Buchstabenvermittlung erteilt. Sie konnten feststellen, dass die zweite Trainingsgruppe deutliche Vorteile gegenüber der ersten Trainingsgruppe und der Kontrollgruppe hatte. Ein Transfereffekt auf die Rechtschreibung erwies sich in beiden Gruppen als nicht signifikant. Die Autoren vermuten, dass ein vorschulisches Training umso effektiver ist, je expliziter dieses mit dem Erwerb der Buchstaben verbunden ist. Aufgrund des hohen Drop-Outs (zum Ende der Untersuchung im ersten Schuljahr waren nur noch 36 Kinder übrig) kann die Untersuchung nicht als repräsentativ angesehen werden, sie zeigt jedoch eine Tendenz auf. Da die Langzeiteffekte des Trainings nicht überprüft wurden, kann auch nicht ausgeschlossen werden, dass die Effekte des phonologischen Trainings erst in höheren Klassenstufen zum Tragen kommen, da aufgrund der alphabetischen Strategie die Kinder mit dem Vorteil im Buchstabenwissen zunächst bessere Chancen im Schriftspracherwerb aufweisen, während das phonologische Wissen vielleicht explizit für den Erfolg im späteren Schriftsprachverlauf noch hilfreich sein könnte. Schatschneider, Fletcher, Francis, Carlson

und Foorman (2004, S. 279) konnten in ihrer Untersuchung mit 384 Kindern für das Buchstabenwissen eine Korrelation zur Leseflüssigkeit von r = .30 feststellen.

Die längsschnittlichen Untersuchung mit 69 Grundschulklassen zum „Modell Schriftsprach-Moderatoren“ (MSM) von Deimel, Ziegler und Schulte-Körne (2005, S. 7) zeigte, dass die Buchstabenkenntnis in Kombination mit den Vorkenntnissen der phonologischen Informationsverarbeitung der in der Untersuchung stärkste Prädiktor der Rechtschreibleistungen am Ende der zweiten Klasse ist. Bei der Einschulung übertrifft die Buchstabenkenntnis noch den Faktor der phonologischen Informationsverarbeitung bzgl. der Rechtschreibleistungen und ist damit der stärkste Einzelprädiktor in der angegebenen Untersuchung.

In einer dritten Studie zum Würzburger Training (vgl. Roth 1999, S. 171ff.; Schneider, Roth & Ennemoser 2000, S. 290ff.; Roth & Schneider 2002, S. 101ff.) wurden nur Risikokinder (N = 126 aus N = 726), die aufgrund des Bielefelder Screenings (vgl. Jansen, Mannhaupt, Marx und Skowronek 1999, S. 86f.) identifiziert worden sind, mit dem Würzburger Programm gefördert. In der Kontrollgruppe (N = 108) musste aus moralischen Gründen auf die Unterscheidung zwischen Risiko- und Nichtrisikokindern verzichtet werden, was wiederum die Aussagekraft der Ergebnisse mindert. Die Vorschulkinder wurden in drei Gruppen eingeteilt. Bei der ersten Trainingsgruppe wurde das 20-wöchige Würzburger Trainingsprogramm eingesetzt, mit dem Fähigkeiten der phonologischen Bewusstheit trainiert werden. Die zweite Trainingsgruppe wurde über zehn Wochen lang mit spielerischen Übungen zur Laut-Buchstabenverbindung bedacht und die dritte Gruppe erfuhr über 20 Wochen eine Kombination aus den beiden Trainingseinheiten. Die Kinder, die ein reines Training der phonologischen Bewusstheit erhalten haben, erzielten in einem Test zu phonologischen Fertigkeiten im Vergleich zu den anderen beiden Gruppen auch die besten Ergebnisse, bzgl. der schriftsprachlichen Fähigkeiten erwies sich das kombinierte Training am erfolgreichsten. Von den trainierten Risikokindern hatten nur 5 % der Kinder in einem Rechtschreibtest am Ende des zweiten Schuljahres einen Prozentrang kleiner als 15. Jedoch konnten nur die Kinder, die ein kombiniertes Training erfahren haben, die durchschnittlichen schriftsprachlichen Leistungen der Kinder der Kontrollgruppe erreichen. Diese Ergebnisse konnten auch noch am Ende der dritten Klasse aufrechterhalten werden. Zudem waren zum Ende der zweiten Klasse nur 6 % der Risikokinder der kombinierten Trainingsgruppe dem unteren Leistungsbereich zuzuordnen. Damit konnte diese Gruppe mit dem Anteil der Kontrollgruppe von 8 % im unteren Leistungsbereich mithalten. In der Gruppe mit dem Training der Buchstabenkenntnisse lag der Anteil der unteren Leistungsgruppe bei 22 %, bei der Gruppe mit dem isolierten Training der phonologischen Bewusstheit bei 20 %. Aufgrund der

zu kleinen Datenbasis konnten Schneider und Küspert (2003, S. 108ff.) nur vermuten, dass ein kombiniertes Training den größten Vorteil bringt. Die Befunde von Roth (1999, S. 250) weisen damit auch auf die Bestätigung der phonologischen Verknüpfungshypothese (Phonological-Linkage-Hypothese) hin, die bisher noch nicht im deutschen Sprachraum untersucht worden war. Die Vertreter der Phonological-Linkage-Hypothese (vgl. Hatcher, Hulme & Ellis 1994, S. 41ff.) gehen davon aus, dass die besten Ergebnisse in Trainingsstudien erzielt werden können, die ein phonologisches Training mit dem Training von Buchstabenkenntnis, bzw. Graphem-Phonem-Korrespondenzen vereinen. Das reine Buchstabenwissen allein scheint somit keinen allzu großen Einfluss auf die späteren Lese- und Rechtschreibleistungen zu haben. Die phonologische Verknüpfungshypothese vereint die Vorstellung, dass die phonologischen Fähigkeiten und der Schriftspracherwerb wechselseitig einhergehen (vgl. Roth 1999, S. 249).

> *„Zusammenfassend lässt sich sagen, […] dass die gezielte vorschulische Förderung der phonologischen Bewusstheit und der Buchstabenkenntnis einen wirksamen Präventionsansatz zur Vorbeugung von Schwierigkeiten beim Erwerb der Schriftsprache darstellt.“ (Roth & Schneider 2002, S. 106).*

Jenkins (2002, S. 99ff.) dagegen konstatiert, dass ein hohes Maß an Schrifterfahrung auch von Risikofaktoren ablenken kann, die den Schriftspracherwerb negativ beeinflussen. Es handele sich nicht um eine kausale Beziehung zwischen der Buchstabenkenntnis und dem Erfolg im Lesen und Rechtschreiben, sondern die Korrelationen könnten auf eine literarische Bildung und Förderung durch die Eltern oder andere Bezugspersonen hinweisen. Das Buchstabenwissen könnte jedoch bzgl. des Verständnisses und der Ablenkung der Aufmerksamkeit des Kindes auf das abstrakt-symbolische System hinweisen und ist damit auch dem Schriftspracherwerb zuträglich. Lotz, Schoreit und Kemper (2013, S. 87) konnten dagegen feststellen, dass die Buchstabenkenntnis für die Vorhersage des Leseverständnisses in der zweiten Klasse keinen signifikanten Einfluss hat, während phonologische Vorläuferfertigkeiten eine Vorhersage zulassen.

1.4 Sprache

In diesem Kapitel gilt es zunächst den Zusammenhang zwischen geschriebener und gesprochener Sprache zu klären, um dann im Anschluss anhand ausgewählter Studien zu analysieren, inwieweit linguistische Kenntnisse und Fähigkeiten die schriftsprachlichen Fähigkeiten in der Schule vorhersagen können. Hierbei wird auch explizit auf den gut

evaluierten „Simple-View-of-Reading“-Ansatz hingewiesen, der bedeutsam die sprachlichen Komponenten der Leseleistung in den Mittelpunkt rückt.

Die deutsche Schrift ist eine alphabetische Schrift, die eine besondere Beziehung zur Lautsprache aufzeigt. Die mündliche Sprache wird auch als Primär-, die Schrift hingegen als Sekundärsystem bezeichnet (vgl. Coulmas 1982, S. 17). Diese beiden Systeme sind unterschiedlich strukturiert. Es handelt sich aber um Systeme, die miteinander korrespondieren und die in einem gemeinsamen Spannungsfeld stehen. Günther (1995, S. 30) betont, dass die Elemente des Sprechhandelns in die Schrifthandlung mit einfließen, jedoch umgekehrt uns die sprachlichen Grundkonzepte (Wort, Satz, Silbe …) erst durch den Erwerb der Schriftsprache ins Bewusstsein gerufen werden.

Der Erfolg beim Schriftspracherwerb ist unter anderem auch von der Abstraktionsleistung des Schülers abhängig, von den Charakteristika der mündlichen Sprache abzusehen und die Unterschiede zwischen Laut- und Schriftsprache zu entdecken, die verschiedenen Funktionen der Schriftsprache zu erkennen und zu nutzen (vgl. Wygotski 1977, S. 91f.). Die Besonderheit der Schriftsprache besteht darin, dass es sich mit der Schriftsprache um eine vollkommen neue Sprachsituation handelt. Es ist kein Gesprächspartner vorhanden, das Sprechbedürfnis bzw. das Motiv, das sich normalerweise aus einer kommunikativen Situation ergibt, ist vom Kind selbst zu schaffen und es ist ein bewussterer und willkürlicher Gebrauch der grammatischen, semantischen und phonetischen Ebene nötig als beim Gebrauch der Sprechsprache (vgl. Crämer & Schumann 2002, S. 258f.). Der Zusammenhang zwischen mündlicher und schriftlicher Sprache setzt für den Lernenden voraus, dass er lernt die Rechts-Links-Orientierung der Schriftsprache auch der gesprochenen Sprache zu entnehmen (das, was zuerst gehört wird, wird auch als erstes aufgeschrieben) und dass er in der Lage ist, den Lautstrom in einzelne kleinere Einheiten (Wort, Satz, Silbe, …) zu unterteilen bzw. beim Lesen diesen Lautstrom der distinktiven Elemente wiederherzustellen (vgl. Meiers 1998, S. 151ff.). Die Schriftsprache unterliegt einem hohen Abstraktionsgrad, der Bezug zur Lautsprache stellt jedoch eine entscheidende Grundlage dar. Die Abstraktion vom handelnden Kontext fällt Kindern oft noch schwer. So konnte Valtin (1988, S. 13f.) zeigen, dass einige Vorschulkinder bei der Frage nach der Wortlänge beim Vergleich der Wörter „Kuh“ und „Piepvögelchen“ das Wort „Kuh“ als länger identifizierten, weil die Kuh größer ist als ein Vogel. Die Kinder müssen lernen, dass sich die Laut- und Schriftstruktur nicht an Wortbedeutungen orientiert (vgl. Crämer & Schumann 2002, S. 267f.). Im Rahmen der Forschung zum Spracherfahrungsansatz wurde zudem diskutiert, inwiefern der Spracherwerb mit dem Schriftspracherwerb zusammenhängt. Die Verfechter des Spracherfahrungsansatzes (als didaktisches Mittel im Anfangsunterricht) gehen davon aus, dass die

Sprachproduktionen der Kinder selbst das Verbindungsglied zwischen gesprochener und geschriebener Sprache herstellen und der Schriftspracherwerb in ähnlicher Form analog zum Spracherwerb verläuft. Verschiedene Aussagen sprechen jedoch dagegen (vgl. zusammenfassend Walter 1996, S. 166ff.):

- In allen menschlichen Populationen hat sich zwar eine Sprache entwickelt, jedoch nutzen nicht alle Individuen eine Schriftsprache.
- Die Schriftsprache ist kulturell entstanden, während sich die gesprochene Sprache aus der biologischen Evolution herausgebildet hat.
- Gesprochene Sprache entwickelt sich anhand eines universellen Musters als Kombination aus verschiedenen Phonemen, während sich die geschriebene Sprache unterschiedlich realisieren lässt (alphabetische Schriften, chinesische Schriftbilder, …).
- Für die Sprachentwicklung benötigt das Kind nur eine entsprechende Umgebung, das Lesen und Schreiben bedarf jedoch einer spezifischen Instruktion.

In mehreren Untersuchungen konnte gezeigt werden (vgl. zusammenfassend Walter 1996, S. 166ff.), dass sich dieser Ansatz als nicht richtig erwies: „Lesen und Schreiben entwickeln sich nicht so natürlich wie gesprochene Sprache.“ (a. a. O., S. 171).

Nun sei zu überlegen, ob Schwierigkeiten in der Sprachkompetenz sich als ursächliche Faktoren für den Misserfolg im Schriftspracherwerb herausstellen. Verspätete Sprachentwicklung oder auch Artikulationsstörungen, Dysgrammatismus, Wortfindungsstörungen können ebenfalls vorschulische Risikoindikatoren für eine Lese-Rechtschreibschwäche sein (vgl. Warnke, Hemminger, Plume 2004, S. 7), müssen aber nicht zwangsläufig zu einer Lese-Rechtschreibschwäche führen (vgl. Schründer-Lenzen 2009, S. 228). Schneider, Marx und Weber (2002, S. 188) fanden Hinweise darauf, dass Kinder mit spezifischen Sprachentwicklungsstörungen besonders gefährdet sind, Probleme im Schriftspracherwerb zu entwickeln. Verschiedene Studien konnten Hinweise liefern (vgl. zusammenfassend Schneider, Marx & Weber 2002, S. 191f.), dass die Kinder mit Sprachentwicklungsstörungen im Schulalter unterdurchschnittliche Leistungen zeigen. Das theoretische Modell nach Grimm (1995, S. 53ff.), das anhand verschiedener Forschungsbefunde erstellt wurde (vgl. Abb. 19), zeigt, dass vorschulische Sprachstörungen in zwei Problembereiche unterteilt werden können. Vorschulische phonologische und metaphonologische Defizite können demnach kausal mit einer Störung des frühen Leseprozesses im Schulalter interpretiert werden (Graphem-Phonem-Korrespondenzen, ...).

Syntax- und Diskursprobleme werden dagegen als ursächlich für die Störung des späteren Leselernprozesses gesehen, insbesondere dem Leseverständnis.

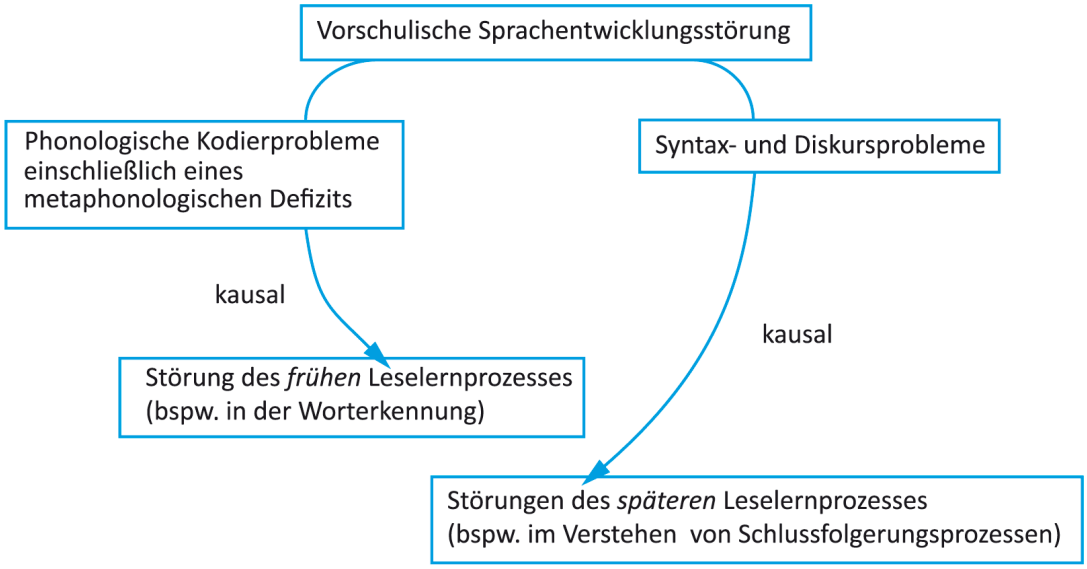


Abb. 19: Theoretisches Modell zum Zusammenhang von Sprachentwicklung und Schriftspracherwerb nach Grimm 1995 (verändert nach Schneider, Marx & Weber 2002, S. 192)

Der Ansatz von Grimm (1995) stellt eine Brücke zwischen den Forschungsbereichen der phonologischen Informationsverarbeitung und den Sprachentwicklungsstörungen dar. Es bleibt jedoch zu betonen, dass nicht jedes Kind mit Sprachentwicklungsstörungen automatisch Schwierigkeiten im Schriftspracherwerb hat und anders herum erst recht nicht, dass automatisch bei jedem Kind mit Problemen in den schriftsprachlichen Leistungen die Sprachentwicklungsverzögerung als ursächlicher Faktor heranzuziehen ist. Bishop und Edmundson (1987a, S. 156ff.; 1987b, S. 444ff.) dagegen konnten aufzeigen, dass Sprachschwierigkeiten die Entwicklung der phonologischen Fähigkeiten nicht zwingend beeinflussen müssen.

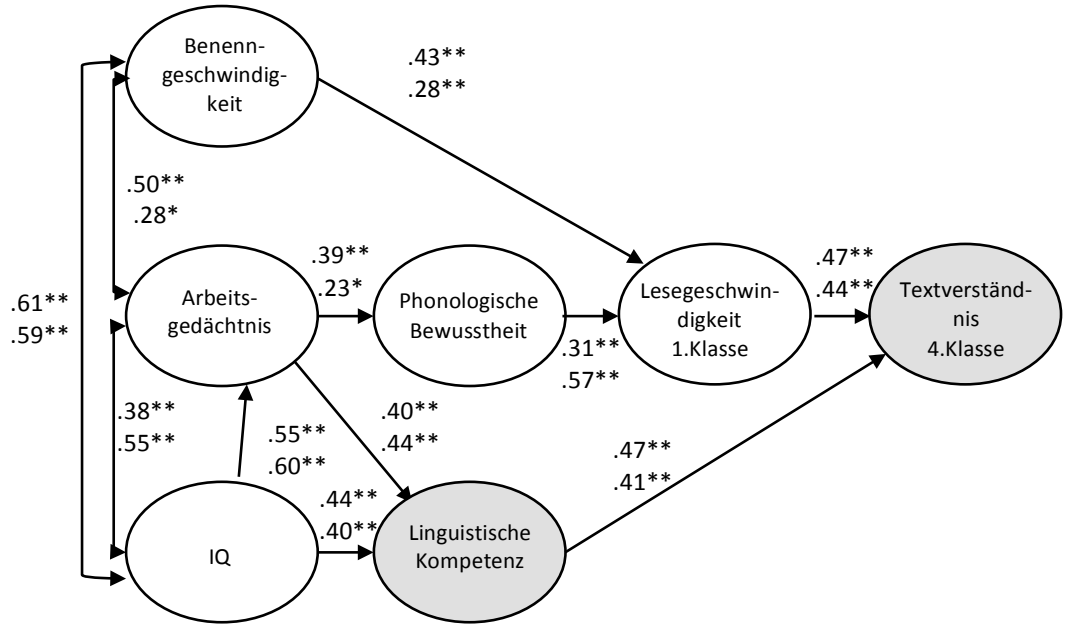
Hagtvet (1997, S. 174) konnte anhand einer Längsschnittstudie mit 74 Kindern, die im Alter von vier bis neun Jahren begleitet worden sind, herausfinden, dass auch Semantik und Syntax einen Einfluss auf die spätere Lese- und Rechtschreibentwicklung haben, dies sich aber im Vergleich zur phonologischen Bewusstheit relativiert: "[…] phonemic awareness stood out as theoretically the most important variable in reading development, syntax and semantics also played their part."

Der Reynell Score, der Syntax und Semantik der Kinder im Alter von vier Jahren bestimmt, konnte besser als die kognitiven Leistungen der Kinder signifikant die Varianz des Lesens

von Sätzen aufklären ($p < 0.05$). Auch Scarborough (1991, S. 207) verglich in seiner Langzeitstudie 22 Kinder, die später lese-rechtschreibschwach wurden mit 22 Kindern, die in der Schule keine Probleme entwickelten. Er stellte fest, dass sich das Satzverständnis, die syntaktische Komplexität sowie die Länge der Äußerungen aus dem Kindergartenalter auf die Lese- und Rechtschreibfähigkeit im Grundschulalter auswirken. Die Stichprobe war bzgl. Intelligenz, sozioökonomischem Status und Geschlecht vergleichbar. Kim und Pallante (2010, o. S.) untersuchten 468 spanischsprachige Kindergartenkinder und Erstklässler und konnten für den Wortschatz keinen prädiktiven Wert erfassen. Als positive Faktoren ergaben sich dagegen das Benennen von Buchstaben und die phonemische Segmentierung. Schulte-Körne, Deimel, Bartling und Remschmidt (1998, S. 28) fanden in ihren Untersuchungen unter anderem heraus, dass die Sprachwahrnehmung und -unterscheidung eine Voraussetzung für die phonologische Bewusstheit zu sein scheint.

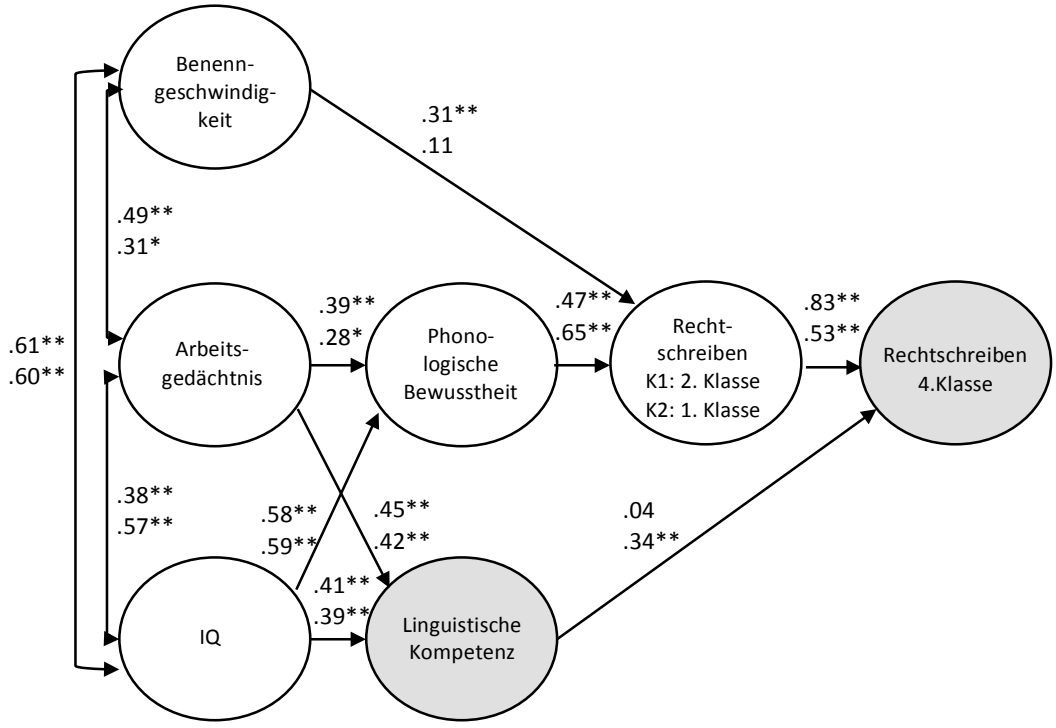
Auch Roth, Speece und Cooper (2002, S. 262ff.) konnten anhand von 88 Kindern, deren Lernprozess sie bis zum Ende der zweiten Klasse verfolgten, zum einen bestätigen, dass die phonologische Bewusstheit 61 % der Varianz der Fähigkeit des Pseudowortlesens in der ersten und zweiten Klassenstufe aufklärt, konnten aber auch feststellen, dass die Fähigkeiten der mündlichen Sprache im Bereich Satzverständnis und Sprachstruktur im Kindergarten diese Fähigkeiten ebenfalls zu 23 % Varianz aufklärten. Die mündliche Sprache im Kindergarten kann zudem die Lesekompetenz am Ende der zweiten Klasse besser vorhersagen als die phonologische Bewusstheit. Andere Autoren unterscheiden die Prädiktorleistung je nach Kriterium. Wortschatz und Grammatik seien demnach bessere Prädiktoren für das Leseverständnis, während die phonologische Informationsverarbeitung besser die Lesegenauigkeit und -geschwindigkeit erklären (vgl. Catts 1993, S. 951; Catts & Hogan 2003, S. 237f.).

Goldammer, Mähler, Bockmann und Hasselhorn (2010, S. 54) konnten anhand einer Längsschnittstudie mit 53 Kindern von fünf bis acht Jahren zeigen, dass die Komponente des Satzgedächtnisses der in ihrer Untersuchung stärkste Prädiktor zur Vorhersage von Lese-Rechtschreibfähigkeiten im zweiten Schuljahr ist. Die phonologische Informationsverarbeitung dagegen nahm nur eine Nebenrolle ein. Die Leistung des Satzgedächtnisses lässt sich aus den sprachlichen Bereichen der Größe des Wortschatzes und der Kapazität des phonologischen Arbeitsgedächtnisses erklären. In zwei unlängst veröffentlichten Studien von Ennemoser, Marx, Weber und Schneider (2012, S. 63) stellte die linguistische Kompetenz eine bedeutende Komponente zur Prognose von schriftsprachlichen Fähigkeiten dar (vgl. Abb. 20 und Abb. 21).



Anmerkung: Studie 1: oberer Koeffizient; Studie 2: unterer Koeffizient; *p < .05, **p < .01

Abb. 20: Strukturgleichungsmodell zur Vorhersage von Lesegeschwindigkeit und Textverständnis (nach Ennemoser, Marx, Weber & Schneider 2012, S. 63)



Anmerkung: Studie 1: oberer Koeffizient; Studie 2: unterer Koeffizient; *p < .05, **p < .01

Abb. 21: Strukturgleichungsmodell zur Vorhersage von der Rechtschreibkompetenz (nach Ennemoser, Marx, Weber & Schneider 2012, S. 64)

Die Effekte der linguistischen Kompetenz konnten durch die längsschnittlich über vier Jahre angelegte Studie eindrucksvoll aufgezeigt werden. Die Autoren konnten zeigen, dass sich gute linguistische Kompetenzen nicht direkt im Anfangsunterricht positiv auswirken, sondern sich besonders in Bezug auf das Textverständnis und die Rechtschreibkenntnisse in der 4. Klasse und damit langfristig gesehen als vorteilhaft erweisen. Die Autoren bewerten die linguistische Kompetenz als einen wichtigen weiteren spezifischen Ansatzpunkt, der auch in Präventionsmaßnahmen Berücksichtigung finden sollte (vgl. a. a. O., S. 66).

Auch der Simple-View-of-Reading Ansatz (vgl. Gough & Tunmer 1986 und Gough, Hoover & Peterson 1996), der in US-amerikanischen Studien mehrfach gut evaluiert werden konnte, beschreibt, dass für die Lesekompetenz das Hörverstehen (print-independent) mit der Dekodierfähigkeit (print-dependent) einhergehen. Bei diesem Ansatz wird der gesamte Leselernprozess in einem theoretischen Modell inkludiert, dessen Ziel der erworbenen Lesefähigkeit, das Leseverständnis auf der Textebene ist. Dies bedeutet, dass auch grundlegende sprachliche Verstehensprozesse zu einer höheren Lesekompetenz beitragen. Hierbei wird davon ausgegangen, dass die Verarbeitungsprozesse für gehörte und gelesene Informationen vergleichbar sind, dass also für gesprochene und geschriebene Sprache ähnliche Prozesse ablaufen. Hörverstehen beschreibt dabei die Fähigkeit, einen Satz syntaktisch analysieren zu können sowie inhaltlich zu verstehen, die Dekodierfähigkeit beschreibt die basalen Fähigkeiten des Leselernprozesses, die Worterkennung. Sind beide Fähigkeiten ausgeprägt, ist das Kind in der Lage sinnentnehmend zu lesen. Hoover und Gough (1990) führten eine fünfjährige Studie mit 254 Schülern durch und konnten die Befunde positiv evaluieren. Marx und Jungmann (2000, S. 90), die den Ansatz den monistischen Positionen zuordnen, konnten die besondere Bedeutung der Dekodierfähigkeit und des Hörverstehens für Leseverständnisleistungen in der Grundschule empirisch bestätigen. Sie konnten nachweisen, dass sich die Hörverstehensleistung bis zum Ende der vierten Klasse im Wesentlichen der Leseverständnisleistung angeglichen hat, mit der Vermutung, dass dies mit der fortgeschrittenen Dekodierfähigkeit im Zusammenhang steht (vgl. Marx & Jungmann 2006, S. 82ff.). Dies konnte bis in die 6. Klasse hinein nachgewiesen werden. Es wird angenommen, dass Worterkennen und Hörverstehen multiplikativ verknüpft sind. Hoover und Gough (2000) postulieren, dass durch diese multiplikative Verknüpfung auch deutlich wird, dass ein Fortschritt beim Lesen nur dann entstehen kann, wenn die Komponenten der Dekodierfähigkeit bzw. des Sprachverständnisses nicht Null sind. Beide Komponenten werden als gleichwertig angesehen.

Gough und Tunmer (1986, S. 9) beschreiben, dass eine gute Dekodierfähigkeit und ein gutes Sprachverständnis (oder Hörverstehen) nicht zwingend miteinander einhergehen. Aus dem Ansatz sind auch die Probleme im Leseverständnis leicht zu beschreiben (vgl. Tab. 25).

Tab. 25: Mögliche Probleme des Leseverständnisses (abgewandelt nach Harrison-Palmer 2012)

		Hörverstehen	
Dekodierfähigkeit		+	-
	+	gutes Leseverständnis	Hyperlexie
	-	Leseschwäche (Legasthenie)	Allgemeine Leseschwierigkeit (schlechte Dekodierfähigkeit und schlechtes Hörverstehen)

Anmerkung: + = gute/s Hörverstehen oder Dekodierfähigkeit, - = schlechte/s Hörverstehen oder Dekodierfähigkeit

Für den Vorschulbereich ist zu vermuten, dass zunächst Probleme im Bereich des Dekodierens für Schwierigkeiten im Lesen verantwortlich sind, sich dies im fortgeschrittenen Leselernprozess auf Probleme im Hörverstehen verschiebt. Dies konnte von Marx und Jungmann (2000, S. 90) durch die zunehmende Korrelation zwischen Hör- und Leseverständnis bei zunehmender Erfahrung im Lesen nachgewiesen werden. Außerdem konnten Rost und Hartmann (1992, S. 345ff.) nachweisen, dass das Hörverstehen eine bessere Varianzaufklärung (38 %) für das Leseverstehen hat, als die sprachliche Intelligenz (28 %). Stanovich (2000) geht davon aus, dass das Hörverstehen eine Art obere Grenze des Leseverstehens darstellt und Leseleistungen besser als die Intelligenz vorhersagen kann.

Die Sprache ist somit zusätzlich zu den phonologischen Komponenten ebenfalls ein weiterer wichtiger Teilbereich, der zur Erklärung der Lese- und Rechtschreibfähigkeiten einen Beitrag leistet. Als Ursache für die Lese-Rechtschreib-Schwäche stellen somit Schwierigkeiten in der phonologischen Verarbeitung bisher die am besten empirisch belegte Hypothese dar (vgl. Kap. II.1.2.1). Es gibt jedoch auch Anzeichen, dass die mangelhaften sprachlichen Fähigkeiten nicht nur eine Ursache der Lese-Rechtschreibschwäche sind, sondern auch eine Folge dieser sein können (vgl. Tacke 2008, S. 154). Neurophysiologische Befunde sprechen zudem für eine Auffälligkeit im Planum temporale, das für die Sprachverarbeitungsprozesse eine wichtige Bedeutung hat (vgl. Weber 2003, S. 67).

Aus den oben dargestellten Befunden lassen sich noch weitere Gründe insbesondere für eine Messung der phonologischen Informationsverarbeitung ableiten:

- Bestimmte Aspekte der phonologischen Informationsverarbeitung entwickeln sich bereits im Vorschulalter und lassen sich messen und zum Teil (insbesondere die phonologische Bewusstheit) auch fördern.
- Die phonologische Informationsverarbeitung zeigt in gewissem Maße eine prädiktive Validität bzgl. der schriftsprachlichen Fähigkeiten.
- Kinder mit Lese- Rechtschreibschwierigkeiten unterscheiden sich von Kindern ohne diese Schwierigkeiten häufig durch Probleme in der phonologischen Informationsverarbeitung.

Die Untersuchungen sind nicht zu allen drei Gesichtspunkten konsistent, sodass eine frühe Identifikation von Risikokindern immer einen Teil der Fragen offen lassen wird und es aufgrund der unterschiedlichen Befunde zur Erklärung der Validität des Lese-Rechtschreiberwerbs sicher ist, dass noch weitere Aspekte außerhalb der phonologischen Informationsverarbeitung als prädiktiv für den Erfolg im Lesen und Schreiben in der Schule wirken. Sprachkompetenzen sowie möglicherweise auch die Buchstabenkenntnis scheinen eine Rolle für die Prognose von Lese-Rechtschreibleistungen in der Schule zu spielen. Schneider, Marx und Weber (2002, S. 186ff.) glauben jedoch, dass die phonologische Bewusstheit zwar ein wichtiger Bereich für die Vorhersage, sie aber keine hinreichende Voraussetzung für die Entwicklung von Lese- und Schreibleistungen ist, da auch andere Bereiche wie Schutz- und Risikofaktoren von Relevanz sind.

In diesem Abschnitt wurde nun vorgestellt, ob die Inhalte der Früherkennungsverfahren einen Beitrag zu Fehlklassifikationen im Vorschulalter geleistet haben. Es scheint nicht ganz ausgeschlossen, dass die Früherkennungsverfahren nicht ausreichend mögliche Prädiktoren für den Schriftspracherwerb in ihren Aufgabenkompositionen abbilden. Die meisten Früherkennungsverfahren haben bereits eine der wichtigsten Komponenten, die phonologische Informationsverarbeitung, als Prädiktor berücksichtigt. Es könnte jedoch sein, dass dies die Lese- und Rechtschreibleistungen der Schüler nicht ausreichend vorhersagen kann und Komponenten wie die Sprachkompetenz oder auch mögliche weitere Schutz- bzw. Risikofaktoren die Kinder beeinflussen könnten. Die endogenen und exogenen Einflüsse, die die Lese- und Rechtschreibleistungen der Kinder zwischen der Prädiktormessung und der Messung des Kriteriums verändern und ihre Relevanz in Bezug auf die Prognoseleistung der Diagnostikverfahren werden in nächsten Abschnitt beleuchtet.

2 Endogene und exogene Einflüsse auf die Schülerleistungen

Neben den inhaltlichen Schwächen eines Diagnostikinstruments müssen auch mögliche Einflussfaktoren und Variablen überprüft werden, die zwischen der Messung des Prädiktors im Kindergarten und des Kriteriums in der Schule die Leistungen des Kindes beeinflussen können. Poskiparta, Niemi und Vauras (1999, S. 444) vermuten, dass verschiedene Einflüsse (Intelligenz, mangelnde Buchstabenkenntnis, schlechtes Arbeitsgedächtnis, ...) in Kombination die schlechten Leseleistungen im Schulalter provozieren. Sind nur einige Faktoren betroffen, seien Kinder dazu in der Lage, Schwierigkeiten in bestimmten Bereichen zu kompensieren. Brügelmann (2003a, o. S.) konstatiert sogar, dass aufgrund der Kompensationsleistung der Schüler die Leistung der phonologischen Bewusstheit nur marginal Einfluss auf die Lese- und Rechtschreibleistung der Schüler hat und die Defizite auf Dauer nicht durch mangelnde Fähigkeiten in Bereichen der phonologischen Bewusstheit erklärt werden können, sondern durchaus andere Komponenten in Betracht gezogen werden sollen. Endogene wie exogene Einflüsse auf die Lese- und Rechtschreibleistungen der Kinder, die zu Fehlklassifikationen führen können, werden nun im Folgenden erörtert.

Die Ursachen einer Lese-Rechtschreibschwäche sind bis heute nicht vollständig analysiert. Standen in der Vergangenheit eher basale Funktionen als Ursache im Mittelpunkt, wird heute der Gesamtkomplex einer multivariaten Ursachensuche als Bedingungsgefüge in den Fokus der Untersuchungen gestellt. Einfache Kausalitätsbeziehungen sind bei der Lese-Rechtschreibschwäche nicht hinreichend geklärt. Teilleistungsschwächen konnten nicht als Ursache nachgewiesen werden, da sowohl Kinder mit visuellen und auditiven Wahrnehmungsdefiziten dennoch dazu in der Lage waren, das Lesen und Rechtschreiben adäquat zu erlernen, wie auch Kinder mit Lese-Rechtschreibschwäche keine Wahrnehmungsschwächen in dem Bereich aufwiesen. Genauso wenig konnten legasthenie-spezifische Fehler als Ursache identifiziert werden, da diese auch im Entwicklungsprozess von normal lesenden und schreibenden Kindern nachgewiesen werden können. Ein Unterschied zwischen intelligenten und weniger intelligenten Kindern mit Lese-Rechtschreibschwäche wurde schon in der Diskrepanzforschung widerlegt. Kognitive Ursachen konnten nur in bestimmten Teilbereichen nachgewiesen werden, wobei auch hier der Zusammenhang zwischen Ursache und Wirkung noch nicht geklärt ist. Umweltbedingungen wie Familie und Unterricht haben auf die Lese-Rechtschreibleistung einen Einfluss, dieser steht aber auch in Zusammenhang mit anderen Schulleistungen und wirkt damit unspezifisch. Motivation, Angst und weitere Persönlichkeitsvariablen sind Teil des Ursachengefüges, treten aber auch durch die Schwierigkeiten komorbid auf. Walter (2005a, o. S.; vgl. Abb. 22) verweist auf drei Perspektiven des gestörten

Schriftspracherwerbs, die zu je unterschiedlichen didaktisch-methodischen Konsequenzen für Förderung und Unterricht führen. Er unterscheidet den medizinischen, den kognitiven und den ganzheitlichen Ansatz, die alle drei in der Vergangenheit vermehrt als Erklärungsansätze genutzt worden sind.

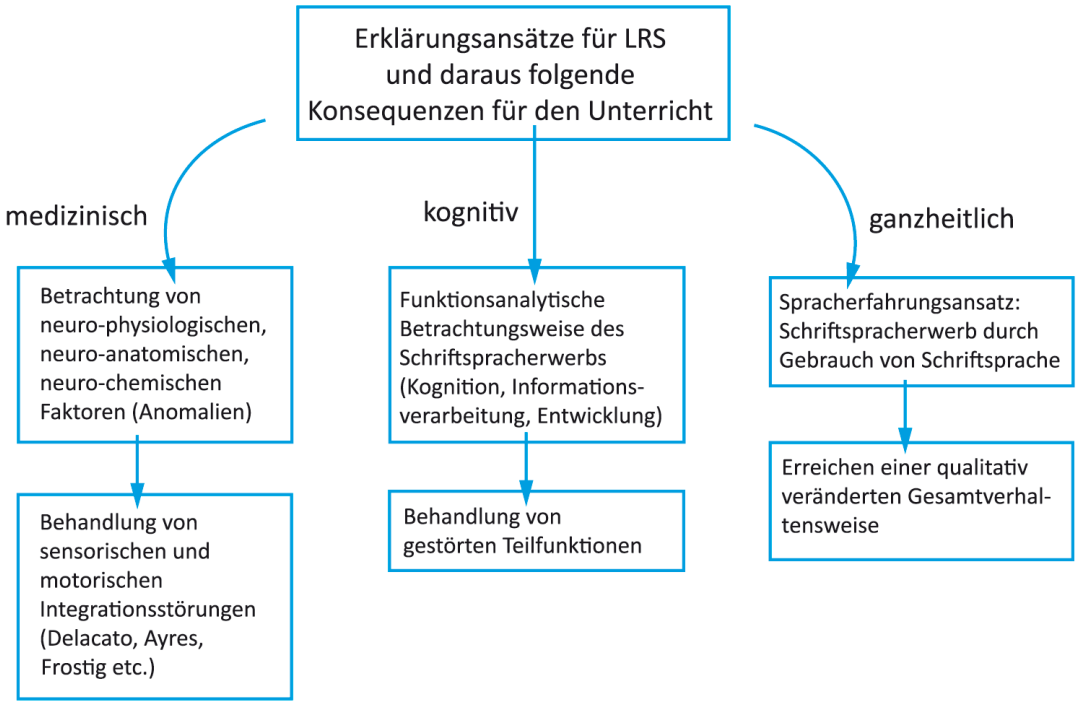


Abb. 22: Erklärungsansätze für Lese-Rechtschreibschwierigkeiten (nach Walter 2005a, o. S.)

An diesen drei Ansätzen wird die Gefahr einer einseitigen Ursachensuche deutlich. Wird nur eine Komponente betrachtet, wirkt sich dies automatisch auf die Förderung der Kinder aus und kann bspw. zu Interventionsformen wie edu-kinesiologische Trainings im medizinischen Bereich führen, die sich später nachweislich als unspezifisch und ineffektiv bzgl. des Lesens und Schreibens erwiesen (vgl. Walter 2002a, S. 40f.).

Der Schriftspracherwerb wird in der gegenwärtigen Diskussion häufig als Entwicklungsprozess gesehen, somit werden auch die Schwierigkeiten im Schriftspracherwerb zunehmend als verzögerte Entwicklung betrachtet. Die Ursachenforschung konnte aufzeigen, dass sich die Kinder mit Schwierigkeiten im Bereich Lesen und Schreiben von anderen Kindern unterscheiden, es jedoch nicht genau festzustellen ist, welche dieser Merkmalsunterschiede vor oder nach dem Eintreten der ersten Schwierigkeiten mit der Schriftsprache aufgetreten sind. Scheerer-Neumann (2004, S. 38) bestätigt ebenfalls die Befunde der Hirnforschung, bestreitet aber ihren praktischen Nutzen für den Bereich der Lese-Rechtschreibschwäche. Phonologische Defizite (vgl. Kap. II.1.2) sind heute bei lese-rechtschreibschwachen Kindern

genauso anzunehmen wie ihre Sichtbarkeit in verschiedenen kognitiven Prozessen. Bisher sind keine der Befunde praktisch relevant geworden, sodass es für die pädagogische Arbeit wichtig erscheint, sich auf die kognitiven Prozesse des Schriftspracherwerbs, sowie auf Motivation und Emotion der Kinder zu konzentrieren, da aus diesen Ebenen Fördermaßnahmen abgeleitet werden können (vgl. a. a. O., S. 38). Marx (2007, S. 99) stellt die Vermutung auf, dass zwar auch inadäquate Aneignungsversuche einen Teil der Problematik erklären, dass jedoch die Annahme bestehen bleibt, dass auch vorschriftsprachliche Voraussetzungen mit zu der Lese-Rechtschreibschwäche beigetragen haben, dies aber weiterer Prüfung und neuer Untersuchungsmethoden bedarf. Walter (2007a, S. 481) beschreibt ein Ursachenmodell von Lundberg und Hsien (1989, S. 589ff.; 1990, S. 231ff.) und Lundberg (1995, S. 89ff.), das umweltbedingte Faktoren und genetische Disposition zur phonologischen Informationsverarbeitung in einen Zusammenhang bringt und Probleme im Bereich der phonologischen Bewusstheit als Symptomatik versteht (vgl. Abb. 23).

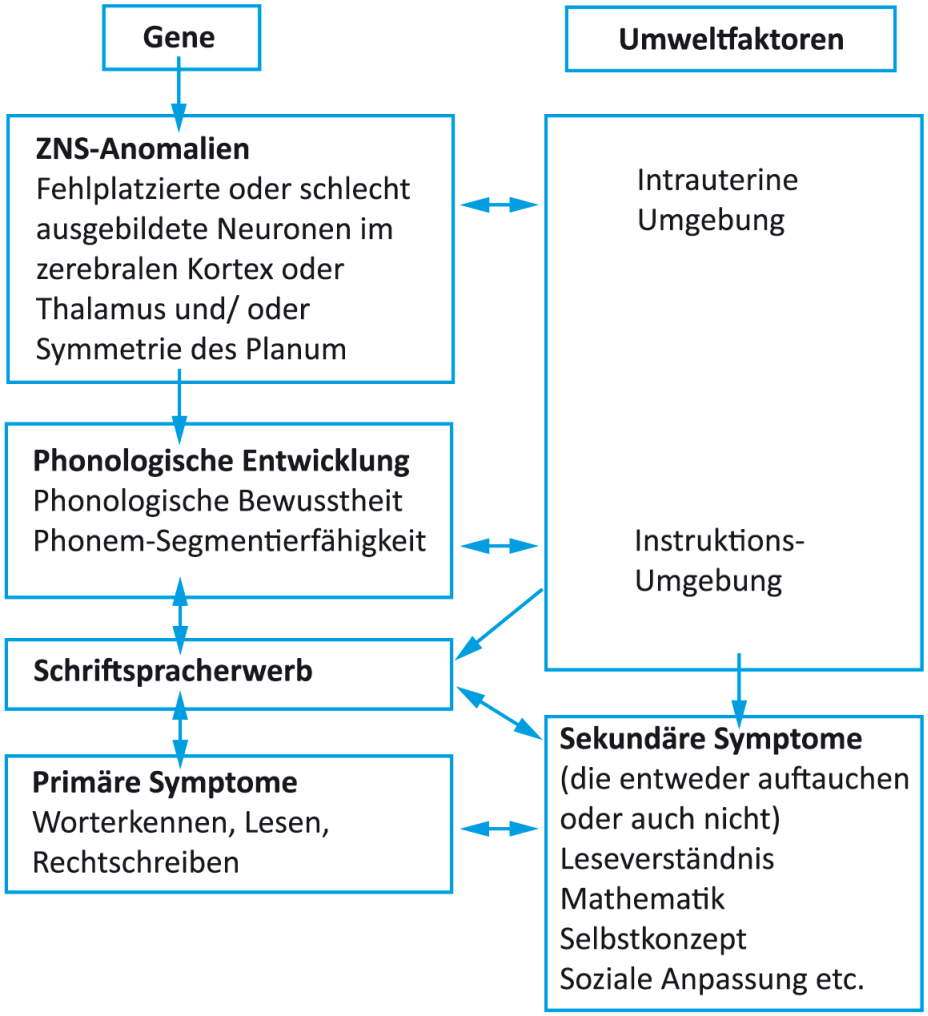


Abb. 23: Hauptkomponenten der Legasthenie aus der Sicht des phonologisch orientierten Ansatzes, verändert nach Lundberg 1995, S. 91 (nach Walter 2007a, S. 481)

Walter (2007a, S. 482) führt bzgl. der neurologischen Komponente eine Reihe von Untersuchungen auf, die die Befunde in Frage stellen, sodass es sich um eine noch nicht bestätigte Hypothese handelt. Insbesondere die Befunde von Castro-Caldas, Petersson, Reis, Stone, Elander und Ingvar (1998, S. 1053ff.) zeigten, dass die funktionalen Organisationen des Gehirns auch von der Beschäftigung mit Schriftsprache und dem Lernen von Lesen und Schreiben abhängig sind und sich von diesen beeinflussen lassen. Außerdem weisen medizinische Studien häufig auch Probleme bei der Parallelisierung auf oder es wurden keine Kontrollgruppen hinzugezogen bzw. die Unterschiede zwischen Kontroll- und Experimentalgruppe waren zu groß. Die Diskrepanzen in den unterschiedlichen Untersuchungsansätzen machen deutlich, dass es offensichtlich nicht die eine Ursache für Lese-Rechtschreibschwäche gibt, sondern die Schwierigkeiten ein komplexeres Ursachengefüge darstellen und die Forschung trotz der zahlreichen Studien diesbezüglich noch in den Kinderschuhen steckt.

Viele unterschiedliche Ursachen einer Lese-Rechtschreibschwäche wurden über die Jahre vermutet, zahlreiche Variablen diesbezüglich überprüft und wieder verworfen. Die Erscheinungsbilder der Lese-Rechtschreibschwäche sind vielfältig. Probleme beim Lesen können auf die Dekodierungsgenauigkeit oder auch die Geschwindigkeit beim Dekodieren zurückzuführen sein. Unzureichende Graphem-Phonem-Korrespondenzen, sowie mangelnde Kontextausnutzung können ebenfalls Teil der Problematik sein. Bei Schwierigkeiten im Bereich der Rechtschreibung sind nicht nur Falschschreibungen, sondern auch Schwierigkeiten im Ausdruck zu beobachten (vgl. Weber 2003, S. 46). Einige Autoren vermuten auf der anderen Seite eine genetische Disposition, andere unterschiedliche biologische Voraussetzungen, weitere wiederum glauben, dass es sich bei der Lese-Rechtschreibschwäche eher um eine Entwicklungsverzögerung handelt.

Die Lese-Rechtschreibschwäche kann auf diverse Ursachen zurückgeführt werden (vgl. Tacke 2008, S. 153). Sie kann durch ein Zusammenspiel von neurobiologischen Lernvoraussetzungen und umweltbedingter schriftsprachlicher Einflüsse beschrieben werden. Hartke und Koch (2008, S. 40f.) unterteilen (ohne Anspruch auf Vollständigkeit) in drei Variablengruppen, die in der Forschung zu Schulproblemen relevant sind:

- Schülerspezifische Variablen
- Variablen der familiären Umwelt
- Schulische und unterrichtliche Variablen

Zu den schülerspezifischen Variablen zählen beispielsweise die kognitiven Lernvoraussetzungen, das Vorwissen, aber auch Konzentration, Arbeitsverhalten oder Metakognition. Die familiären Variablen sind insbesondere auf das Bildungsniveau im Elternhaus, das Einkommen oder den Erziehungsstil bezogen. Schulische und unterrichtliche Variablen hängen von der Art, dem Ort, der Quantität und Qualität des Unterrichts sowie von der Lehrkraft ab (vgl. bspw. die aktuelle Diskussion der Studien von John Hattie 2003, S. 1ff.). Die einzelnen Variablen sind jedoch nicht getrennt voneinander zu betrachten, sondern beeinflussen sich untereinander (vgl. Hartke & Koch 2008, S. 40).

Lese-Rechtschreibstörungen, wie Lernstörungen im Allgemeinen, sind im Sinne neurologischer Annahmen eine Manifestation ineffizienter Gehirnsysteme bzw. -regionen (vgl. Weber 2003, S. 62). In der Entwicklungspsychologie werden Fehlentwicklungen, zu denen auch die Lese-Rechtschreibschwäche gezählt werden kann, dagegen durch Risiko- und Schutzfaktoren determiniert. Risikofaktoren sind Personenmerkmale und beeinflussende Faktoren aus der Umwelt, die die Auftretenswahrscheinlichkeit der Lese-Rechtschreibschwäche vergrößern, während Schutzfaktoren diese verringern (vgl. Hartke & Koch 2008, S. 39f.). Laut Warnke, Hemminger und Plume (2004, S. 10ff.) ist die Lese-Rechtschreibschwäche als Kombination aus beidem eine Veranlagung, die durch Schutzfaktoren in der Umwelt beeinflusst und meistens kompensiert werden kann.

Somit wird ein multivariates Bedingungsgefüge vermutet. Im Folgenden werden weitere Variablen beleuchtet, die ein Teil des Wirkungsnetzwerkes sein können und/oder als Risiko- und Schutzfaktoren auf die Schülerleistungen einwirken.

Kretschmann (2002, S. 5) plädiert im Zuge dessen für eine Auflösung des traditionellen Verständnisses von Ursache und Wirkung:

> *„Ursache und Wirkung können sich – zirkulär – gegenseitig beeinflussen, so kann z. B. der Lernerfolg eines Kindes (oder dessen Ausbleiben) mit dazu beitragen, wie wohlwollend, besorgt oder gar ablehnend sich die Eltern, Lehrer oder Mitschüler dem Kind gegenüber verhalten. Und dieses Verhalten wiederum kann sich auf den künftigen Lernerfolg auswirken.“*

Im Folgenden wird eine Auswahl der drei Variablen, die eine Früherkennung als Risiko- bzw. Schutzfaktoren beeinflussen können, indem sie sich auf die Schülerleistungen auswirken, vorgestellt. Eine umfassende Auflistung möglicher Risiko- und Schutzfaktoren ist in Kretschmann (2002, S. 3f.) beschrieben.

2.1 Schülerspezifische Variablen

Im folgenden Abschnitt wird eine Auswahl möglicher schülerspezifischer Einflussfaktoren analysiert, die die Resultate der Früherkennungsverfahren beeinflussen könnten. Zunächst wird die genetische Prädisposition für Schwierigkeiten im Lesen und Rechtschreiben als eine allgemeine Einflussvariable betrachtet, die als schülerspezifische Variable möglicherweise eine Rolle spielen könnte. Im Fokus dieses Abschnitts stehen jedoch die kognitiven Voraussetzungen der Kinder, die sich möglicherweise als Risiko- oder Schutzfaktoren auf die Schülerleistung Einfluss nehmen können.

2.1.1 Genetische Prädisposition

Die Annahme der genetischen Veranlagung bzw. der Vererbbarkeit von Lese-Rechtschreibschwäche beruht darauf, dass die Schwierigkeiten sich oft manifestieren und in bestimmten Familien gehäuft auftreten. Aufgrund von Familien-, Zwillings- und molekulargenetischen Studien kommen Warnke, Hemminger & Plume (2004, S. 20 ff.) zu dem Schluss, dass genetische Umstände sehr wahrscheinlich für manche Form der Lese-Rechtschreibstörung begründend sind. In Studien konnte eine familiäre Häufung nachgewiesen werden, es wird ein autosomal dominanter Erbgang vermutet. 40 % der Eltern und ungefähr 40 % der Geschwister betroffener Kinder haben ebenfalls eine Lese-Rechtschreibschwäche (vgl. Warnke, Hemminger, Roth & Schneck 2002, S. 36). Bei eineiigen Zwillingen trat die Lese-Rechtschreibstörung zu fast 100 % auf (vgl. Warnke, Hemminger & Plume 2004, S. 21; Warnke, Hemminger, Roth & Schneck 2002, S. 36, Abb. 24). Warnke, Hemminger und Plume (2004, S. 21) weisen jedoch darauf hin, dass sich die genetische Suggestion durch die Alterung der Personen aufgrund von individuellen Umwelteinflüssen verringert.

Auch Pennington, Gilger, Pauls, Smith, Smith und DeFries (1991, S. 1 527) sowie für den deutschsprachigen Raum Schulte-Körne, Remschmidt und Hebebrand (1993, S. 242ff.) konnten eine genetische Prädisposition der betroffenen lese-rechtschreibschwachen Kinder bestätigen. Schlussfolgernd fassen sie zusammen, dass sich durch die Studien kein „Legasthenie-Gen“ nachweisen ließ, sondern die Lese-Rechtschreibstörung ein Zusammenwirken von mehreren Genen ist und diese zu Besonderheiten in der Hirnentwicklung führen. Es gibt Hinweise dafür, dass das Chromosom 6 für lautsprachliches Lernen und das Chromosom 15 für schriftsprachlich relevante Prozesse verantwortlich ist, die Ergebnisse der zugehörigen Studien sind jedoch nicht eindeutig (vgl. Warnke, Hemminger, Roth & Schneck 2002, S. 36f., ein Überblick zu den Untersuchungen der einzelnen Gene: Caylak 2007, S. 929ff.). Schulte-Körne (2004, S. 72) zählt zusätzlich auch

noch die Chromosomen 1, 2, 3 und 18 auf, in deren Region Gene vermutet werden, die eine wichtige Funktion für zentralnervöse Prozesse haben. Dieses Gefüge ist auch abhängig von der jeweiligen Sprache. Außerdem wurde die Mehrzahl der Untersuchungen mit diskrepanten Kindern durchgeführt und es ist nicht klar, ob eine genetische Prädisposition nur für diese Kinder vorhanden ist oder auch für nicht-diskrepante Kinder (vgl. Weber 2003, S. 101). Es ist jedoch zu betonen, dass nicht-genetische Faktoren, wie der schulische Schriftspracherwerb, einen bedeutenderen Einfluss haben (vgl. Warnke, Hemminger & Plume 2004, S. 21).

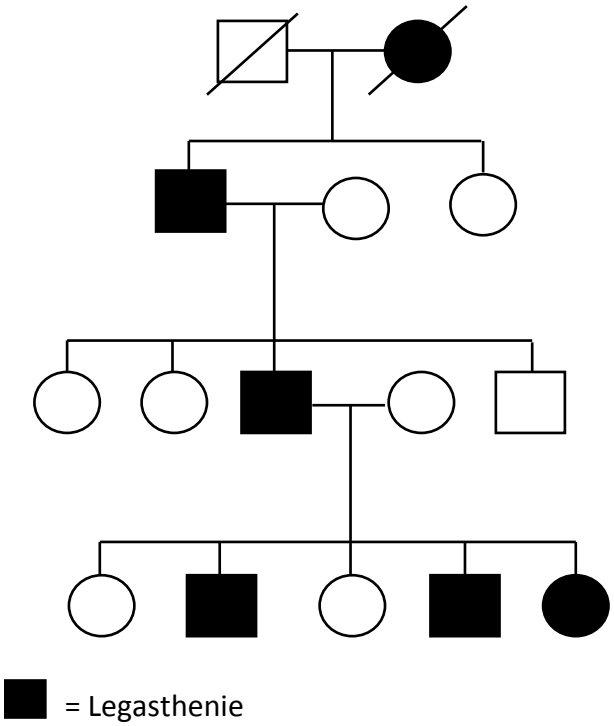


Abb. 24: Das Auftreten von Familienmitgliedern mit Legasthenie über vier Generationen nach Grimm, Noethen & Schulte-Körne 1998 (nach Warnke, Hemminger, Roth & Schneck 2002, S. 37)

Puolakanaho, Poikkeus, Ahonen, Tolvanen und Lyytinen (2004, S. 238ff.) führten einen Vergleich der Leistungen in der phonologischen Bewusstheit von 98 Kindern mit familiärem Risiko zu 91 Kindern ohne familiäres Risiko durch. Die Kinder konnten im Alter von 3,5 Jahren einen signifikanten Unterschied in den phonologischen Fertigkeiten aufweisen, auch wenn andere Sprachfertigkeiten kontrolliert wurden. Im ersten Schuljahr unterschieden sich die Dekodierleistungen der Kinder mit familiärem Risiko ebenfalls signifikant von denen ohne familiäres Risiko. Auch wenn die jeweils besseren Kinder der Gruppe mit der Kontrollgruppe verglichen wurden, waren deren Dekodierfähigkeiten signifikant schwächer ausgeprägt (vgl. Lyytinen, Aro, Eklund, Erskine, Guttorm, Laakso, Leppänen, Lyytinen, Poikkeus, Richardson & Torppa 2004, S. 211; Lyytinen, Ahonen, Eklund, Guttorm, Kulju, Laakso, Leiwo, Leppänen, Lyytinen, Poikkeus, Richardson, Torppa & Viholainen 2004,

S. 168). Snowling, Gallagher und Frith (2003, S. 363) untersuchten Kinder im Alter von drei Jahren bis zu ihrem achten Lebensjahr. 66 % der Kinder, bei denen bereits in der Familie ein Fall von Lese-Rechtschreibschwäche bekannt war, hatten mit acht Jahren häufiger Schwierigkeiten im Lesen als die Kinder der Vergleichsgruppe (13 %). Das Fazit der Autoren ist jedoch, dass die Resultate die Frage unbeantwortet lassen, inwieweit es möglich ist, dass einige „family-risk children" trotz des Risikos ihre Schwächen kompensieren können und normale Leser werden.

Bei den Studien zur familiären Disposition sei auch anzumerken, dass die Kinder nicht nur die Gene der Eltern erben, sondern sich auch ihre Eltern nicht aussuchen können. Mit ihnen einhergehen Umwelteinflüsse aus dem direkten familiären Umfeld, die sozio-ökonomischen Bedingungen, die soziale Umgebung, die Erziehung und vieles mehr (vgl. Hofmann 2006, S. 104). So ist aus der Fülle der Studien zur Prädisposition nicht ganz ersichtlich, ob es sich tatsächlich um eine biologische Vererbung handelt oder ob nicht vielmehr die Umwelt-bedingungen von Generation zu Generation weitergereicht werden.

Hofmann (2006, S. 103) ist der Auffassung, dass Gehirnstrukturen im Wesentlichen genetisch bedingt sind, die Feinstrukturen dann durch neuronale Aktivitäten ausgebaut und miteinander vernetzt werden:

> *„Diese soweit entstandene neurobiologische Basis kann durchaus für eine optimale Schriftsprachverarbeitung unzureichend sein. Schwächen im Lesen und in der Rechtschreibung können somit wohl eine neurobiologische Basis haben, sie können jedoch nicht direkt auf eine genetische Disposition mit einem oder mehreren genau umschriebenen Genen („Legasthenie-Gen") zurückgeführt werden."*

Schulte-Körne, Warnke und Remschmidt (2006, S. 441) bestätigen, dass es zwar erste „Kandidatengene" gebe, es jedoch zurzeit auf individueller Ebene nicht möglich sei, genetische Risikofaktoren zu bestimmen.

Dies konnte Sirch (1975, S. 14) bereits feststellen. Er bemängelt, dass nicht erkannt wurde, dass „[...] Lesen und Rechtschreiben nicht biologisch bedingte Fähigkeiten sind, sondern sekundäre kulturelle Fähigkeiten". Nach Sirch (ebd.) sollten die Lernprozesse der Kinder stärker beleuchtet werden als die Ursache in der Genetik zu suchen. Auch Bjorklund und Schneider (2006, S. 773) fassen zusammen, dass ein Verhalten nicht ausschließlich genetisch oder ausschließlich erfahrungsbedingt ist, sondern die Beziehung zwischen Organismus und

Kontext betrachtet werden muss, um Entwicklung richtig zu verstehen. Dies wird durch die Aussage von Johnson (1998, S. 4, zit. aus Bjorklund & Schneider 2006, S. 773) gestützt:

> *„Da es inzwischen klar geworden ist, dass die Gene mit ihrer Umgebung auf jeder Ebene – einschließlich der molekularen – interagieren, lässt sich kein Aspekt der Entwicklung als absolut 'genetisch' bezeichnen."*

Es ist somit nicht anzunehmen, dass eine reine genetische Prädisposition die Schwierigkeiten im Lesen und Rechtschreiben in der Schule verursachen und damit die Ergebnisse der Früherkennungsverfahren stark beeinflussen könnte. Insbesondere kann nicht geklärt werden, an welcher Stelle sich diese benachteiligenden Voraussetzungen bei einer Prädisposition auf die Schülerleistungen auswirken. Bezüglich der Früherkennungsverfahren wäre es interessant zu erfahren, in welchem Zeitraum eine schwierige genetische Grundlage sich aktiv auf die Schülerleistungen auswirken würde. Nur wenn dies explizit im Vorschulalter oder in der ersten Klasse eine hohe Relevanz hat, würde es die Ergebnisse der Studien zu den Screening-Verfahren möglicherweise beeinflussen. Wie deutlich gemacht wurde, scheinen gerade genetische Prädispositionen als Risikofaktor dennoch durch andere Schutzfaktoren kompensiert werden zu können.

Einige andere Schutz- und Risikofaktoren werden in den weiteren Abschnitten vorgestellt. Zunächst werden die kognitiven Voraussetzungen der Kinder beleuchtet.

2.1.2 Kognitive Voraussetzungen

In der Vergangenheit wurden Schwierigkeiten im Lesen und Schreiben häufig als eine Hirnstörung beschrieben. Die damalige Hypothese, die Lese-Rechtschreibstörung als minimale cerebrale Dysfunktion (MCD) zu verstehen, konnte in wissenschaftlichen Studien jedoch nicht gehalten werden (vgl. zusammenfassend Warnke, Hemminger & Plume 2004, S. 10ff.). Es konnten bei Kindern mit einer Lese-Rechtschreib-Schwäche zwar häufiger Symptome gefunden werden (vgl. bspw. Punt, de Jong, de Groot & Hadders-Algra 2010, S. 1ff.), die auf eine Dysfunktion im Hirn hinweisen, sie kommen jedoch nicht gehäuft in dieser Gruppe vor, sodass andere Faktoren einen bedeutend größeren Einfluss haben müssen (vgl. Warnke, Hemminger & Plume 2004, S. 11f.). Die rein ätiologische Ursachensuche im Rahmen einer medizinischen Dysfunktion wurde aufgegeben (vgl. Weber 2003, S. 47f.). Die Theorie der legasthenietypischen Fehler (Reversionsfehler als Unterscheidungskriterium) bzw. einer Raum-Lagelabilität bei lese-rechtschreibschwachen Schülern wurde damals von Schenk-Danziger (1975, S. 120ff.) als wesentliche Ursache konstatiert, konnte aber in Folge-

untersuchungen als nicht haltbare Annahme nachgewiesen werden. Die Untersuchungen konnten weder legasthenietypische Fehler noch Raumorientierungsprobleme oder Linkshändigkeit bei legasthenen Kindern gehäuft feststellen. Die Fehlerarten treten sowohl bei Legasthenikern als auch bei den Kontrollkindern der Untersuchungen auf (vgl. hierzu zusammenfassend Walter 1996, S. 26ff. sowie Müller 1980a, S. 34). Die Gefahr dieser Ansätze ist, dass sich die Förderung der betroffenen Kinder weniger mit dem Hauptaspekt - der Sprache selbst - als mit Funktionsübungen beschäftigt. Walter (2002b, S. 444ff.) stufte Wahrnehmungstrainings aufgrund US-amerikanischer Meta- und Megaanalyse als nicht effektiv ein.

Die hirnanatomischen Korrelate für die schriftsprachliche Informationsverarbeitung scheinen in dem Grad ihrer Wirkung auf die Lese- und Rechtschreibleistung noch nicht hinreichend erforscht. Die Abweichungen betreffen scheinbar vorwiegend linkshemisphärische Nervenbahnen, die für die visuelle und akustisch-sprachliche Informationsverarbeitung zuständig sind (vgl. Warnke, Hemminger & Plume 2004, S. 12f.). Grigorenko (2001, S. 91ff.) dagegen vermutet, dass die gefundenen strukturellen Unterschiede von Personen mit einer Lese-Rechtschreibschwäche im Vergleich zu Personen ohne Schwierigkeiten auch auf eine biologische Ursache der Lese-Rechtschreibschwäche hinweisen, da diese in einer Vielzahl von Studien auf je unterschiedlicher Basis festgestellt werden konnten. Neurobiologische Untersuchungsverfahren eignen sich jedoch aufgrund der Diversität der Ergebnisse nicht für eine Frühdiagnostik.

Neurophysiologische Befunde konnten zeigen, dass nur bei einem bestimmten Personenkreis der Betroffenen die visuelle Informationsbearbeitung beeinträchtigt und auch von Bedeutung ist. Bildgebende Verfahren konnten dagegen einen höheren Anteil an Betroffenen feststellen, die Schwierigkeiten in der sprachlichen Informationsverarbeitung vorwiesen. Das Gehirn wies bei Personen mit einer Lese-Rechtschreibstörung hirnfunktionelle Besonderheiten im linken okzipitalen-temporalen Hirnrindenbereich auf. Der Großteil der Ergebnisse wissenschaftlicher Untersuchungen zeigt, dass die defizitäre Verarbeitung von phonologischen Informationen größtenteils auch auf physiologischer Ebene sichtbar ist (vgl. Warnke, Hemminger & Plume 2004, S. 14ff.). Zugleich wird jedoch darauf hingewiesen, dass aufgrund der Streuung der Ergebnisse und der geringen Zahl der Replikationen die kausale Interpretation nicht zulässig sei.

Um neuropsychologische Befunde zur Lese-Rechtschreibschwäche zu erhalten, müssen die Befunde bei betroffenen Schülern mit den modellartigen Abläufen im normalen Schrift-spracherwerbsprozess verglichen werden. Hierzu dienen die Modelle des Schrift-

spracherwerbs. Die neuropsychologischen Ansätze zur Ursachenforschung sind sehr stark abhängig vom zugrunde liegenden Modell. In Stufenmodellen (vgl. bspw. Frith 1985, S. 300ff.; Scheerer-Neumann 1989a, S. 32f.; Goswami 1993, S. 443ff.) kann eine Störung auf jeder einzelnen Ebene auftreten, während sich in Netzwerkmodellen (vgl. bspw. Seidenberg & McClelland 1989, S. 523; Ziegler, Perry & Coltheart 2000, S. 416) die Störung in der zentralnervösen schriftsprachlichen Informationsverarbeitung vollzieht (vgl. Warnke, Hemminger & Plume 2004, S. 15ff.).

Warnke, Hemminger und Plume (2004, S. 16) stellen ein vereinfachtes Ursachenmodell (vgl. Abb. 25) zur Lese-Rechtschreibstörung aus neuropsychologischer Sicht vor. Die genetische Disposition wirkt sowohl auf die visuellen als auch auf die auditiven Informationsverarbeitungsprozesse ein. Aus diesen gestörten Prozessen und einer Entwicklungsstörung des Lesens und Schreibens folgen eine Verlangsamung und defizitäre Sinnentnahme, sowie etwaige Fehler beim Lesen und Schreiben.

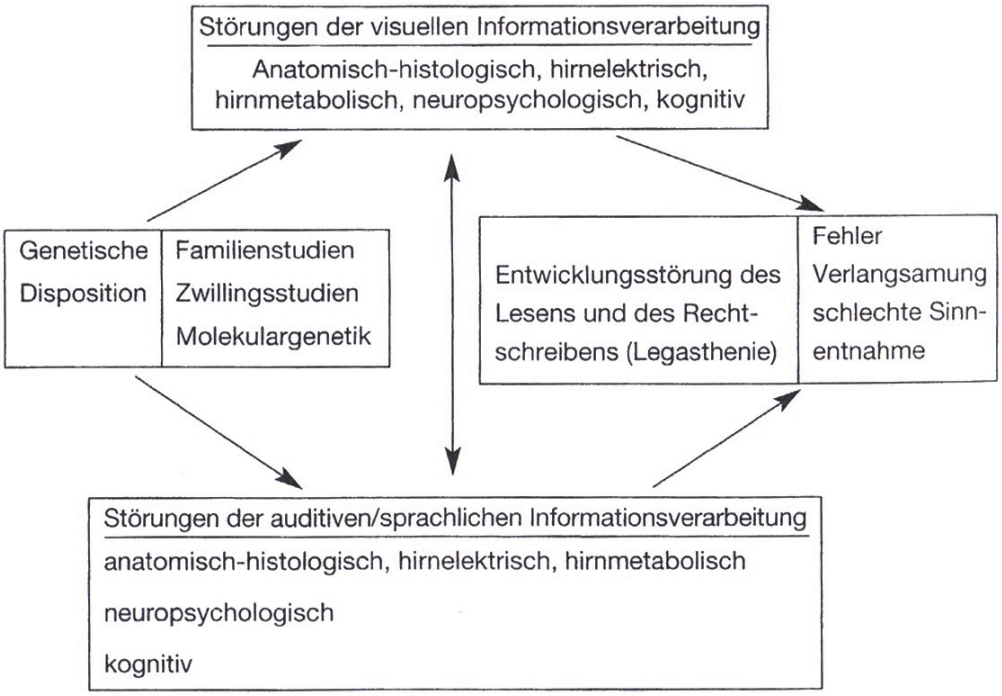


Abb. 25: Vereinfachtes Ursachenmodell der Lese-Rechtschreibstörung aus neurologischer Sichtweise (nach Warnke, Hemminger & Plume 2004, S. 16)

Castles, McLean und McArthur (2010, S. 426ff.) unterscheiden in Bezug auf die neuropsychologische Lese-Rechtschreibschwäche in „high-level deficits“ und „low-level deficits“. Bei den „high-level deficits“ (vgl. a. a. O., S. 427) stellen sie die phonologischen Defizite sowie Schwierigkeiten beim schnellen Benennen, die ja teilweise auch schon in den Früherkennungsverfahren Verwendung finden (vgl. Kap. I.2.2) in den Vordergrund der Ursachenforschung, während sie unter den „low-level deficits“ (vgl. a. a. O., S. 428)

Schwierigkeiten in der auditiven sowie visuellen Verarbeitung und abnorme Gehirnstrukturen verstehen. Sie vermuten, dass sich die auditive und visuelle Verarbeitung nur indirekt auf die Lese- und Rechtschreibfähigkeiten auswirken und diese hauptsächlich auf das schnelle Benennen und die phonologischen Defizite übertragen, die wiederum zu Lese- und Schreibschwierigkeiten führen können (vgl. Abb. 26). In der Abbildung 26 wird diese Sichtweise der Autoren deutlich. Sie schließen jedoch weitere Faktoren (Umwelteinflüsse usw.) nicht aus und propagieren durch die unterschiedlichen Pfade ein multikausales Ursachenmodell.

Im neuropsychologischen Bereich spielt auch die phonologische Informationsverarbeitung eine große Rolle. Warnke, Hemminger und Plume (2004, S. 18) beschreiben, dass sich für die phonologische Bewusstheit hirnanatomische und hirnfunktionelle Korrelate finden ließen, die jedoch ebenfalls noch nicht ausreichend erforscht worden sind. Sie stellen zudem die Vermutung auf, dass die Lese-Rechtschreibstörung auch eine Störung des Übersetzungsprozesses von visuell schriftsprachlichen Informationen hin zu akustisch-sprachlichen Informationen ist.

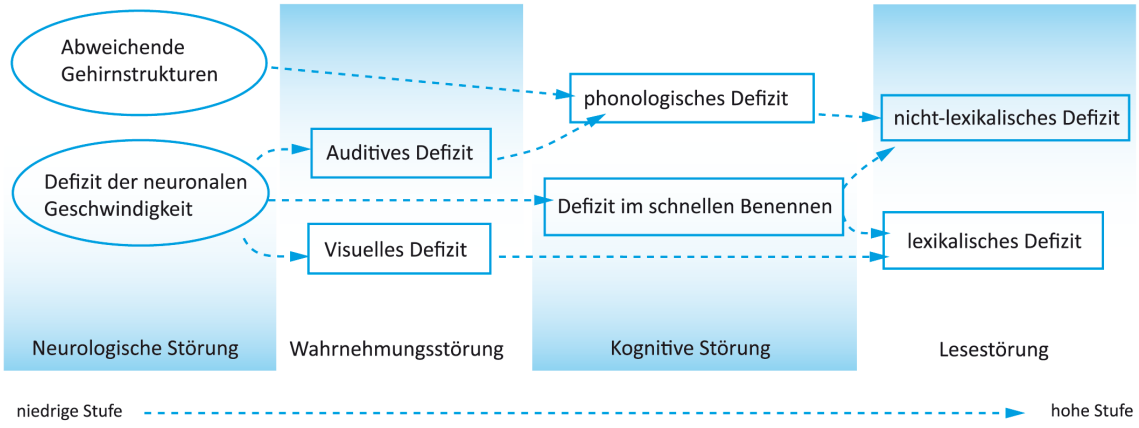


Abb. 26: Ursachenmodell bei einer Lese- und Rechtschreibschwäche (aus dem Englischen übersetzt, nach Castles, McLean & McArthur 2010, S. 429)

Ligges und Blanz (2007, S. 109ff.) vermuteten anhand ihrer Übersicht zu bildgebenden Verfahren in der Lese-Rechtschreibforschung, dass bei der Verarbeitung von phonologischen Aufgaben ein linkshemisphärisches neurobiologisches Defizit bei der Lese-Rechtschreibschwäche vorliegt, während die rechtshemisphärischen Aktivierungsunterschiede nur im geringen Maße bei Kindern vorkommen und wahrscheinlich entwicklungsbedingte Kompensationsprozesse darstellen. Bei der Untersuchung wurde jedoch weiterhin von der Definition der Lese-Rechtschreibschwäche über das Diskrepanzkriterium ausgegangen. Auch Larsen, Høien, Lundberg und Ødegaard (1990, S. 289) konnten in ihrer Studie nachweisen, dass atypische Symmetrien mit Schwierigkeiten

auf phonologischer Ebene in Beziehung gesetzt werden können. Breitenbach und Lenhard (2001, S. 167ff.) definieren dagegen den Ansatz der ungewöhnlichen Asymmetrien nach der Sichtung von 120 Artikeln zu neurobiologischen Korrelaten als zu unspezifisch. Schuchardt, Kunze, Grube und Hasselhorn (2006, S. 265ff.) konnten zeigen, dass Kinder mit schwachen schriftsprachlichen Leistungen sowohl phonologische als auch zentralexekutive Defizite aufwiesen. Die bisher dargestellten Untersuchungen zu kognitiven Defiziten konnten aufzeigen, dass kognitive Komponenten scheinbar einen Risikofaktor darstellen können. Sie sollten nicht als Ursache für Lese-Rechtschreibschwierigkeiten bezeichnet werden, auch in den referierten Untersuchungen wird größtenteils darauf hingewiesen, dass noch weitere Bedingungen die Kinder beeinflussen. Es kann jedoch möglich sein, das bestimmte kognitive Komponenten auf die Kinder einwirken und damit die Prognosegenauigkeit der Verfahren verändert wird.

Im Folgenden werden die in der Forschung kontrovers diskutierten kognitiven Komponenten der auditiven und visuellen Informationsverarbeitung sowie die Intelligenz, die alle kompensatorische Wirkung auf mögliche Defizite in den Vorläuferfertigkeiten haben könnten und damit die Ergebnisse der Schüler in der Grundschule trotz guter/schlechter Vorhersagewerte verschlechtern/verbessern könnten, diesbezüglich nochmals genauer betrachtet.

2.1.2.1 Auditive Informationsverarbeitung

Bei Menschen mit Lese-Rechtschreibschwierigkeiten treten anatomische und histologische Veränderungen häufiger auf als bei Menschen ohne Schwierigkeiten mit der Schriftsprache. In erster Linie sind in der linken Hirnhälfte Veränderungen zu sehen. Die linke Hemisphäre ist insbesondere für die akustische, sprachliche und sprachlich-visuelle Informationsverarbeitung verantwortlich. Zudem scheinen Personen mit Lese-Rechtschreibschwäche sprachliche Laute langsamer akustisch zu unterscheiden, als Personen ohne Schwierigkeiten mit der Schriftsprache. Durch bildgebende Verfahren ergaben sich Anzeichen, dass bei einem Teil der betroffenen Personen beispielsweise bei Reimaufgaben die Anregung der Hirnbereiche unterschiedlich ist. Die Befunde sind jedoch nicht kongruent und zum Teil widersprüchlich. Es bedarf weiterer Untersuchungen, um hier eine wissenschaftlich fundierte Aussage treffen zu können (vgl. zusammenfassend Warnke, Hemminger, Roth & Schneck 2002, S. 34ff.). Bezüglich des Zusammenhangs von Lese-Rechtschreibschwäche und auditiver Verarbeitungs- und Wahrnehmungsstörungen konnten Suchodoletz, Berwanger und Mayer (2004, S. 24ff.) anhand von 58 Kindern aufzeigen, dass

es keinen Grund für die Annahme eines kausalen Zusammenhangs besteht. Es konnte keine Korrelation zwischen auditiver Wahrnehmung und Rechtschreibleistung nachgewiesen werden.

In den letzten Jahren stand auch immer häufiger die „Phonologie-Defizit-Hypothese“ im Raum, bei der angenommen wird, dass lese-rechtschreibschwache Kinder insbesondere Schwierigkeiten bei Unterscheidung von lautlichen Segmenten in der Sprache sowie bei der Graphem-Phonem-Zuordnung haben. Es wird angenommen, dass die phonologische Informationsverarbeitung ein Prädiktor für den Erfolg vom Schriftspracherwerb in der Schule ist (vgl. Kap. II.1.2). Marx, Weber und Schneider (2001, S. 91f.) konnten in einer Studie mit 82 lese-rechtschreibschwachen Kindern aufzeigen, dass alle Kinder (mit und ohne IQ-Diskrepanz) im Bereich der phonologischen Informationsverarbeitung der Kontrollgruppe unterlagen. Es scheinen sich neurologische Auffälligkeiten in den für die Sprachverarbeitung zuständigen Arealen zu bestätigen (vgl. Landerl 2001, S. 65). Breitenbach und Lenhard (2001, S. 174f.) konstatieren, dass eine herabgesetzte Auflösungsfähigkeit auditiver Stimuli in Verbindung mit der Lese-Rechtschreibschwäche zurzeit noch ungeklärt bleibt, da das auditive System noch nicht ausreichend erforscht ist. Dagegen steht eine Untersuchung von Tallal (1993; 1997; zit. aus Breitenbach und Lenhard 2001, S. 175), die besagt, dass die herabgesetzte Auflösungsfähigkeit die Diskriminations- und Segmentierungsfähigkeit mindert und dies die Entwicklung der phonologischen Bewusstheit beeinträchtigt.

Schwierigkeiten bei der zeitlichen Auflösung akustischer Signale können zu einer Beeinträchtigung des Erkennens von Lauten führen. Schneider und Näslund (2003, S. 136ff.) zeigten, dass ein Training der Zeitverarbeitung in verschiedenen Studien zwar zu einem signifikanten Unterschied in der Zeitverarbeitung selbst führte, sich diese Effekte jedoch nicht auf die Lese- und Schreibkompetenzen bzw. deren Vorläuferfertigkeiten bzgl. der Lauterkennung transferieren ließen. Auch unter Berücksichtigung von Einflussfaktoren wie Intelligenz oder Alter der Kinder konnten die Mittelwertdifferenzen zwischen den unauffälligen und den in der Schriftsprache entwicklungsgefährdeten Kindern nicht mehr nachgewiesen werden (vgl. a. a. O., S. 164). Die Untersuchungen zeigen, dass Zeitverarbeitungsdefizite offensichtlich nicht die Ursache einer Lese- und Rechtschreibschwäche sein können.

Castles, McLean und McArthur (2010, S. 428) fassen die Erkenntnisse neuropsychologischer Studien der letzten 30 Jahre zur auditiven Verarbeitung zusammen und kommen zu der Erkenntnis, dass die Rate der Kinder mit Lese-Rechtschreibschwäche und Schwierigkeiten in

der auditiven Verarbeitung je nach Aufbau der Untersuchung stark schwankt. Es sei zu vermuten, dass 30 bis 50 % der Kinder Schwierigkeiten in der auditiven Verarbeitung haben, dies jedoch stark abhängig von der jeweiligen Stichprobe und den spezifischen Untersuchungsmerkmalen ist. Außerdem scheinen die Kinder mit Lese-Rechtschreibschwierigkeiten nicht alle das gleiche Defizit in der auditiven Verarbeitung aufzuzeigen. Es sei zum Lese- und Schreiberwerb jedoch anzumerken, dass auch bei Defiziten in der auditiven Wahrnehmung die Graphem-Phonem-Korrespondenzen aktiv erlernt werden müssen und ein ausschließlich „gutes Hinhören" nicht ausreichend ist (vgl. Jung 1981, S. 36ff. und zusammenfassend Walter 1996, S. 54ff.).

Walter (a. a. O., S. 55) konstatiert:

> *„Die Pathologisierung bestimmter schriftsprachlicher Realisierungen von Schülern als Wahrnehmungsfehler beruht häufig auf einer falschen Beurteilung des Verhältnisses von Laut und Schrift und ist daher ungerechtfertigt. Die möglichen Formen von Graphem-Phonem-Korrespondenzen müssen explizit gelehrt werden und können nicht ohne weiteres vorausgesetzt werden."*

Die Forschungslage scheint nicht eindeutig zu sein. Es ist anhand einiger Untersuchungen jedoch gezeigt worden, dass einige Kinder mit Lese- und Rechtschreibproblemen ein auditives Defizit verzeichnen. Es könnte sein, dass auch die auditive Informationsverarbeitung ein Risiko- oder Schutzfaktor darstellen könnte.

Im nächsten Teil wird mit der visuellen Informationsverarbeitung und ihrem Einfluss auf die Lese- und Schreibfähigkeiten von Kindern eine weitere kognitive Variable betrachtet.

2.1.2.2 Visuelle Informationsverarbeitung

Schon zu Beginn des 20. Jahrhunderts wurde den legasthenen Kindern ein visuelles Defizit, damals unter dem Begriff der „Wortblindheit" bekannt, unterstellt (vgl. bspw. Hinshelwood 1904, S. 1 303f.). Später wurde dann vermutet, dass lese-rechtschreibschwache Kinder ein Defizit in der visuellen Wahrnehmungsleistung aufweisen, das jedoch unspezifisch wirkt. Dies bedeutet, dass sich das Defizit bei der optischen Gliederung nicht nur auf Buchstaben beschränken solle, sondern ein umfassendes Bild darstelle (vgl. bspw. Müller 1967, S. 37ff.). Die Untersuchungen wiesen jedoch Mängel auf und konnten teilweise durch andere Untersuchungen (vgl. bspw. Vellutino, Steger & Kaman 1975, S. 22ff.) widerlegt werden.

Olsen (1966, S. 179ff.) führte eine Studie durch, die mit 71 Legasthenikern der zweiten Klasse den Test von Frostig (1964) zur Augen-Hand-Koordination, Wahrnehmung der Formkonstanz, Figur-Grund-Unterscheidung und der Raum-Lage-Labilität anwendete. Die Autoren konnten nur einen geringen Zusammenhang aufzeigen. Auch in Trainingsstudien zur visuellen Wahrnehmung konnten keine relevanten Effekte erzielt werden (vgl. zusammenfassend Walter 2000, S. 383). In Untersuchungen zum Transfer der Förderung visueller Wahrnehmung konnten zwar Effekte bzgl. der Leseleistung erzielt werden, allerdings wurde zugleich keine Verbesserung der visuellen Wahrnehmung festgestellt, so dass Walter (ebd.) anmerkt, dass die Effekte somit nicht unbedingt auf die verbesserte visuelle Wahrnehmung zurückzuführen sind. Trotzdem finden sich auch heute noch Hinweise in der Literatur, dass das Konzept der visuellen Wahrnehmungsschwäche noch nicht abgelegt worden ist, so wie beispielsweise Squires und McKeown (2006, S. 18) in ihrer Veröffentlichung „Supporting Children with dyslexia - Practical approaches for Teachers and Parents“ weiterhin Hinweise auf ursächliche Bedingungen der visuellen Wahrnehmung geben.

Die Hypothese der visuellen Wahrnehmungsschwäche konnte über die Jahrzehnte nicht gänzlich widerlegt werden. Bezüglich des Sehvermögens konnte in Untersuchungen festgestellt werden, dass die Augenbewegung von Personen mit Lese-Rechtschreibschwäche verändert ist. Die Fixationen sind länger und ungezielter als bei normalen Lesern und Schreibern. Es ist zu vermuten, dass diese Befunde die Folgen einer Lese-Rechtschreibschwäche, jedoch nicht ihre Ursache darstellen. Störungen der Sehfunktion des Auges als Ursache einer Lese-Rechtschreibschwäche scheinen damit eher die Ausnahme zu bilden (vgl. Warnke, Hemminger, Roth & Schneck 2002, S. 31; Singleton 1988, S. 112).

Auch Evans (2004, S. 20) kann zusammenfassend aus englischen Studien berichten:

> *“Ocular pathology is rare in childhood and is not a major cause of reading difficulties. Poor visual acuities and refractive errors are only rarely severe enough to interfere with reading, although longsightedness is often missed by school vision screening and can cause visual discomfort and blurred vision with reading.”*

Besonderheiten der visuellen Wahrnehmung, also der Verarbeitung der visuellen Informationen, können möglich sein. Es wurden hirnanatomische Ausnahmen in Bezug auf die Verarbeitung in der linken Hirnhälfte gefunden, in der hirnelektrische Reize beim Lesen von Wörtern und Buchstaben verlangsamt waren. Histologische Befunde konnten zeigen, dass

der Sehkern eine Besonderheit der Zellstruktur bei Kindern mit Lese-Rechtschreibschwäche aufweist und dass das Bewegungssehen, das zum Lesen benötigt wird, erschwert ist. Die Befunde bedürfen einer weiteren wissenschaftlichen Überprüfung (vgl. Warnke, Hemminger, Roth & Schneck 2002, S. 31f.). Warnke (1999, S. 2) vermutet bei ca. 3 bis 10 % der lese-rechtschreibschwachen Kinder ein Defizit in der visuellen Informationsverarbeitung (vgl. Warnke & Roth 2000, S. 465 mit 5 – 10 %). Marx, Weber und Schneider (2001, S. 91) konnten dagegen in einer Studie mit 82 lese-rechtschreibschwachen Kindern (mit und ohne IQ-Diskrepanz) feststellen, dass sich im Vergleich zur Kontrollgruppe keine Unterschiede in der visuellen Wahrnehmungsleistung zeigten. Auch Metz, Marx, Weber und Schneider (2003, S. 130ff.) bestätigten, dass lese-rechtschreibschwache Kinder in der Studie unabhängig von ihrer Intelligenz keine Defizite im Bereich der visuellen Informationsverarbeitung hatten.
Mähler und Schuchardt (2014, S. 84) stellten anhand der Studie von Hasselhorn et al. (2012) zu dem Verfahren AGTB 5-12[15] fest, dass Kinder mit einer Lese-Rechtschreibschwäche in den visuellen Gedächtnisaufgaben zum visuell-räumlichen Arbeitsgedächtnis keine Schwierigkeiten aufwiesen. Dies passe zu der Annahme, das die Lese-Rechtschreibschwäche eine Störung der phonologischen Verarbeitung sei (vgl. a. a. O.).

Die visuelle Wahrnehmung wird als unspezifisch bewertet und der kausale Zusammenhang von einigen Autoren angezweifelt. Gasteiger-Klicpera und Klicpera (2004, S. 49) vermuten, dass nur eine kleine Untergruppe der Kinder betroffen ist. Ihre Bedeutung wird im Vergleich zur auditiven Wahrnehmungsleistung als gering bewertet (vgl. Schulte-Körne 2004, S. 73). Die Klärung, ob die visuelle Informationsverarbeitung als Folge oder Ursache der Lese-Rechtschreibstörung interpretiert werden kann, steht zudem noch aus. Überdies wurde in den Studien häufig nicht konsequent zwischen visueller und artikulatorischer Informationsverarbeitung getrennt. Beim Merken eines Bildpunkts bzgl. der Raum-Lage greifen viele Kinder auf die artikulatorische Merkstrategie zurück (der Punkt liegt „oben rechts“), womit auch phonologische bzw. verbale Gesichtspunkte eine Rolle spielen könnten. Ebenfalls könnte bei der Benennung eines Buchstabens auch ein Benennungsproblem statt einer Unfähigkeit visuelle Muster zu erkennen vorliegen (vgl. bspw. Vellutino 1980, S. 251ff.). Visuelle bzw. visuo-motorische Schwierigkeiten von Kindern mit einer Lese-Rechtschreibschwäche können damit derzeit nicht im Sinne eines kausalen Zusammenhangs interpretiert werden.

Anhand der hier dargestellten Untersuchungen zum Zusammenhang der visuellen und auditiven Informationsverarbeitung zu schriftsprachlichen Leistungen ist eher unwahr-

15 AGTB 5-12: Arbeitsgedächtnistestbatterie für Kinder von 5 bis 12 Jahren

scheinlich, dass sich durch die beiden kognitiven Verarbeitungsmechanismen diese Vielzahl an fehlklassifizierter Kinder in den Früherkennungsuntersuchungen erklären lassen. Es ist durchaus möglich, dass bei vereinzelten Kindern die visuelle oder vor allem die auditive Verarbeitung kompensatorisch Schwächen ausgleicht oder sie noch verschärft, doch der Anteil fehlklassifizierter Kinder lässt noch weitere oder andere einwirkende Faktoren vermuten.

Im Folgenden wird die Intelligenz betrachtet. Sie wird aufgrund ihres bereits nachgewiesenen Zusammenhangs mit schulischen Leistungen in der vorliegenden Arbeit besonders in den Fokus gerückt (vgl. bspw. Valerius und Sparfeldt 2015, S. 101).

2.1.2.3 Intelligenz

Es konnte nachgewiesen werden, dass die kognitiven Leistungen in der industrialisierten Welt den akademischen Grad und die späteren beruflichen Leistungen relativ zuverlässig abbilden (vgl. Sternberg, Grigorenko & Bundy 2001, S. 1ff.). Somit können die Intelligenz bzw. die allgemeinen kognitiven Fähigkeiten eines Kindes einen weiteren möglichen unspezifischen Prädiktor und damit einen möglichen Risiko- oder Schutzfaktor für Lese-Rechtschreibleistungen darstellen. Valerius und Sparfeldt (2015, S. 101) konnten feststellen, dass der allgemeine Intelligenzfaktor mit den Schulleistungen in sechs Fächern substantiell zusammenhing ($r = .44$).

Dem Begriff der Intelligenz liegen eine Reihe von Definitionen zugrunde, die sich zum Teil widersprechen. Die Intelligenz eines Menschen ist nicht als unumstößlicher Wert messbar, sondern bedingt sich immer aus den Bestandteilen des Intelligenztests und dessen zugrunde gelegter Theorie bzw. Intelligenzdefinition. „Intelligenz ist vielmehr ein Phänomen mit vielen Facetten, das über kein einzelnes Messverfahren adäquat erfasst werden kann.“ (Bjorklund & Schneider 2006, S. 771-772)

Zudem handelt es sich bei der Intelligenz um eine veränderbare Größe. Sie kann in Abhängigkeit von den Umweltbedingungen variieren (vgl. Bjorklund & Schneider 2006, S. 807). Dies sollte insbesondere auch bei der Interpretation der Untersuchungen zur prädiktiven Kraft der Intelligenz berücksichtigt werden. Es ist also theoretisch möglich, dass die Lese- Rechtschreibleistungen die Intelligenz und die Intelligenz die Lese-Rechtschreib-leistungen positiv wie negativ beeinflussen und damit für einige Kinder resilient oder vulnerabel wirken können.

Schlechte Lese- oder Rechtschreibleistungen wurden in der Vergangenheit häufig auf mangelnde kognitive Leistungen zurückgeführt (vgl. Pfeiffer & Zielinski 1975, S. 1ff.). Diese Ansicht wurde später revidiert, da nachgewiesen werden konnte, dass Lese-Rechtschreibprobleme bei Schülern mit unterschiedlichsten kognitiven Leistungen auftreten können. Nur eine deutlich unterdurchschnittliche Intelligenzleistung konnte mit Lese- und Rechtschreibschwierigkeiten in Verbindung gebracht werden (vgl. Röhr 1978, S. 43, zit. nach Brügelmann 1984, S. 73).

Werner (o. J., S. 5) konstatiert: „[…], dass erfolgreiches Lernen nur indirekt von der Intelligenz abhängt: entscheidender ist […] das (fachspezifische) Vorwissen." Intelligenz kann den Erwerb und den Abruf der Wissensbasis steuern, es kann aber fehlendes Wissen nicht ersetzen (vgl. ebd.).

Die Korrelation zwischen Intelligenzquotienten und Lese- Rechtschreibfähigkeit schwankt in den unterschiedlichen Untersuchungen stark. Bei Warnke, Hemminger, Roth und Schneck (2002, S. 25) liegt die Korrelation bei r = .04. Malmquist (1958, zit. nach Zimmermann 1980, S. 33) erreicht dagegen eine Korrelation zwischen r= .40 und r= .60, wobei die höheren Korrelationen bei älteren Schülern oder Erwachsenen aufzufinden waren, da in den Intelligenztests die Lese- sowie sprachliche Leistung von Belang war. Pfeiffer und Zielinski (1975, S. 1ff.) konnten dagegen aufzeigen, dass die Rechtschreibleistung sich zu 34 % aus den Intelligenzvariablen, sich aber die Varianz der Intelligenz nur zu 10 % aus den Rechtschreibleistungen erklären lässt. Nach näherer Analyse der Zusammenhänge konnten sie nachweisen, dass die Varianz hauptsächlich aufgrund der Regelfehler zustande kommt, weniger aufgrund von Merk- und Wahrnehmungsfehlern. Klicpera und Gasteiger-Klicpera (vgl. 1993, S. 45ff.) erhielten in einer Studie mit Viertklässlern eine Korrelation von r = .52 von Intelligenz (nonverbaler IQ-Test) und Rechtschreibung und eine Korrelation von r = .41 bzgl. Intelligenz und Lesen. Landerl und Wimmer (1994, S. 160) konnten in ihrer Studie mit 60 Zweit- bis Viertklässlern keinen systematischen Zusammenhang zwischen IQ und Leseleistung feststellen. Nur der Koeffizient „Nichtwortlesen in der 3. Klasse" konnte von insgesamt 15 Koeffizienten einen signifikanten Zusammenhang (r = .28) aufzeigen. Bei der Rechtschreibleistung dagegen ergaben sich zwei von drei Koeffizienten als signifikant (Rechtschreibleistung der 2. Klasse: r = .30 und Rechtschreibleistung der 4. Klasse: r= .36). In einer weiteren Studie konnten Klicpera und Gasteiger-Klicpera (2001, S. 45f.) anhand einer Stichprobe von 1 800 Kindern der zweiten bis vierten Klasse der Grundschule zeigen, dass sich zwar zwischen lese-rechtschreibschwachen und durchschnittlichen Schülern ein Unterschied in phonologischen Leistungen zeigte, die Unterschiede jedoch zwischen

diskrepanten und nicht-diskrepanten Kindern irrelevant sind. Sie fanden auch heraus, dass die lese-rechtschreibschwachen Kinder mit höherer Intelligenz Vorteile im orthografischen Wissen hatten. Dieser Vorsprung im orthografischen Wissen geht jedoch mit ansteigender Klassenstufe zurück. In der bereits erwähnten längsschnittlichen Untersuchung zum „Modell Schriftsprach-Moderatoren“ (MSM) von Deimel, Ziegler und Schulte-Körne (2005, S. 6) mit 69 Grundschulklassen konnte gezeigt werden, dass die kognitiven Leistungen und die Werte des Bielefelder Screenings gleichauf prädiktiv auf die Leseleistungen am Ende der zweiten Klasse wirkten. Beim Rechtschreiben stellten sich diese Effekte nicht ein. Eine aktuelle bereits erwähnte Untersuchung von Ennemoser, Marx, Weber und Schneider (2012, S.63ff., vgl. auch Abb. 20, S.161 und Abb. 21, S. 161) konnte mit zwei Studien einen prädiktiven Zusammenhang zwischen dem IQ und den Prädiktoren der phonologischen Informationsverarbeitung sowie der linguistischen Kompetenz, die sich wiederum prädiktiv auf die Lese- und Schreibleistungen der Kinder in der Schule auswirken, nachweisen. In beiden Studien ergibt sich ein indirekter Effekt des nonverbalen IQs auf die schriftsprachlichen Fähigkeiten der Kinder.

Bradley und Bryant (1985, S. 163ff.) konnten in einer Studie zu den Schriftsprachleistungen von Schülern herausfinden, dass der mit Abstand größte Teil der Lese-Rechtschreibvarianz durch die Intelligenz erklärt wird. In der LOGIK-Studie konnte dies ebenfalls bestätigt werden. Hier war es insbesondere die sprachliche Intelligenz, die sich als bedeutsamer Prädiktor für Lese-Rechtschreibleistungen herauskristallisierte. Wie auch schon beim Strukturgleichungsmodell von Schneider und Näslund (1992, S. 268) sichtbar wird, hat der IQ einen indirekten, aber jedoch starken Einfluss auf die Rechtschreibleistung der Kinder. Der Effekt beim Leseverständnis hingegen ist geringer. Stanovich und Siegel (1994, S. 24ff.) teilten die Leser in einer Studie N =250 Kindern mit Leseschwäche und N = 719 normallesenden Kindern in vier IQ-Gruppen auf. Sie konnten feststellen, dass sich in jeder der vier Gruppen sowohl gute als auch schlechte Leser befanden. Eine unterdurchschnittliche Leistung bzgl. der kognitiven Fähigkeiten ließ also nicht zwangsläufig einen Rückschluss auf ebenfalls unterdurchschnittliche Leseleistungen zu, da viele Schüler mit normalem IQ trotzdem schlechte Leseleistungen aufwiesen. Die sprachlichen Leistungen waren bei der Vorhersage der Leseleistung aussagekräftiger als der IQ-Wert des Kindes.

Auch Vellutino, Scanlon und Lyon (2000, S.227) konnten in einer sechsjährigen Longitudinalstudie (N = 1 407 Vorschulkinder) zeigen, dass nur 10 bis 20 % der Varianz von schlechten Leseleistungen durch die Intelligenz erklärt werden konnten. Schwache Leseleistungen sind somit nicht ausschließlich mit den IQ-Werten der Kinder erklärbar. Die Forscher fanden zusätzlich heraus, dass die Intelligenz insbesondere im Bereich des

Schriftspracherwerbs am Anfang der Schulkarriere die Leseleistung nicht gut genug vorhersagen konnte. Durch die Überprüfung des Leseverständnisses konnte gezeigt werden, dass die Vorhersageleistung steigt, je mehr die Kinder sich von den Erstleseleistungen (z.B. dem Lesen von Buchstabe zu Buchstabe) losgelöst haben und ganze Einheiten oder Wörter lesen lernen. Vellutino et al. (2000. S. 236) vermuten auf der Basis der vorhandenen Literatur, dass 10 bis 35% der Leseleistungen durch die Intelligenz (inkl. sprachlicher Intelligenz) erklärt werden können. Die Korrelationen der verbalen Intelligenz waren dabei höher als die der Handlungsintelligenz (vgl. auch Weber 2003, S. 80). Tiu, Thompson und Lewis (2003, S. 424) konnten aufzeigen, dass die Handlungsintelligenz (erhoben durch den WISC-III) sowie die Dekodierfähigkeit und das Hörverstehen in Kombination innerhalb einer Regressionsgleichung einen Beitrag zur Aufklärung der Lesefähigkeit leisten konnten.

Da die Intelligenz selbst nur unzureichend die Varianz der Leseleistungen erklären konnte, untersuchten Forscher nun, ob Teilbereiche der Intelligenz sich bei lese-rechtschreibschwachen Kindern unterscheiden. Klicpera, Gasteiger-Klicpera und Schabmann (1994, S. 87ff.) konnten zeigen, dass die Lese-Rechtschreibleistungen innerhalb der Gruppe der lese-rechtschreibschwachen Kinder (mit und ohne Diskrepanz zwischen IQ und Lese-Rechtschreibleistungen) nicht vom Intelligenzquotienten der Kinder abhängig war. Es konnte nur festgestellt werden, dass die Lesegeschwindigkeit der diskrepanten Kinder deutlich langsamer war und sie mit der Erfassung längerer Wörter größere Schwierigkeiten hatten als die nicht-diskrepanten Kinder.

Die gefundenen Korrelationen zwischen Lese-Rechtschreibleistungen und Intelligenz können nicht ohne weiteres generalisiert werden. Die Korrelationen in den Studien sind im mittleren Bereich angesiedelt, das Konstrukt der Intelligenz lässt jedoch in jeder einzelnen Studie einen Ermessensspielraum des Autors allein in Bezug auf die Wahl des Messinstruments zu, die auf verschiedenen Definitionen der Intelligenz beruhen. Spearman konstatierte schon 1927: „Intelligenz ist ein Wort mit so vielen Bedeutungen, dass es zuletzt überhaupt keine mehr hat.“ (1927, S. 14)

In vielen Intelligenztests spielt der verbale IQ eine Rolle, der teilweise auch einen Teil der spezifischen Vorläuferfertigkeiten des Schriftspracherwerbs abdeckt. Die Studien sind nicht direkt vergleichbar. Zudem haben die Untersuchungen gezeigt, dass die Intelligenz in den unterschiedlichen Phasen des Schriftspracherwerbs eine unterschiedliche Rolle spielt (vgl. zusammenfassend Weber 2003, S. 86). Für die Zeit des Anfangsunterrichts scheinen laut Klicpera, Gasteiger-Klicpera und Schabmann (1994, S. 87ff.) zunehmend die spezifischen Vorläuferfertigkeiten für einen Erfolg im Lesen und Rechtschreiben verantwortlich zu sein,

während Vellutino et al. (2000, S. 236f.) in ihrer Studie nachwiesen, dass die Intelligenz als prädiktiver Wert für die langfristige gute Lese-Rechtschreibleistung umso wichtiger wird. Dies liegt auch an der Komplexität der Sache: für das sinnentnehmende Lesen und die Erfassung des Inhalts eines Textes werden höhere kognitive Leistungen der Leser verlangt als bei basalen Lesefertigkeiten.

Ein weiterer Grund für die genaue Analyse und kritische Betrachtung der Korrelationen sind Hinweise, dass vorhandene Defizite im Lesen und Rechtschreiben Einfluss auf die Intelligenzleistungen der Kinder haben. Die mit fortgeschrittenem Alter der Schüler immer enger werdende Verkettung zwischen Intelligenzleistung und dem Erfolg im Lesen und Schreiben kann auch durch den sogenannten Matthew-Effekt (nach Stanovich 1986, S. 360ff.) erklärt werden. Sie vermuten damit einen „rich-get-richer“ und einen „poor-get-poorer“ Effekt (vgl. ebd.). Vellutino, Scanlon und Spearing (1995, S. 76ff.) konnten dies in ihrer Studie belegen. Es wird vermutet, dass über die Jahre die nachlassende Motivation und auch die Schwierigkeiten in der Aneignung von Wissen und Wortschatz sowie langsamere Gedächtnisprozesse aufgrund mangelnder Übung und schlechte Lesefertigkeiten einen Einfluss auf die kognitiven Leistungen der Schüler haben. Lese-Rechtschreibleistungen und Intelligenzleistungen nähern sich somit zunehmend an. Vellutino et al. (2000, S. 236f.) zeigten einen Schereneffekt betreffs der Intelligenzleistungen zwischen Schülern mit Lese-Rechtschreibschwäche und Schülern ohne Defizite im Lesen und Rechtschreiben auf. Jansen (1994, S. 170) konnte feststellen, dass eine Förderung im Kindergarten, die auf allgemeinen Fähigkeiten und der Motivation beruht, keine Erfolge im Schriftspracherwerb verzeichnen kann. Jansen (a. a. O., S. 173) vermutet, dass die Förderung im Kindergarten nicht spezifisch auf den Lerngegenstand der Schriftsprache ausgerichtet war. Es wurden jedoch nicht die genauen Fördermaßnahmen in den Kindergärten kontrolliert, was eine explizitere Aussage über die Wirkungen nicht möglich macht. Sollte sich dies bestätigen, ist die pädagogische Relevanz einer Diagnostik von Intelligenz aufgrund der Schwierigkeit, diese effizient zu fördern, nicht gegeben. Die Fragestellung, die Walter und Schulz (2013, S. 70) in den Raum stellen, ist inwieweit eine mögliche Varianzaufklärung sich in der praktischen Umsetzung niederschlägt. Bedeutet also eine bessere Prognoseleistung auch, dass durch ein Training kognitiver Kompetenzen das Risiko für Lese- und Rechtschreibschwierigkeiten vermindert werden kann? In Bezug auf das Training induktiven Denkens fassen sie Untersuchungen zusammen, die bereits Trainingseffekte in Bezug auf die sprachfreie Intelligenz als auch auf spezifische sprachliche Bereiche nachweisen (vgl. Marx 2006, 2009, Marx und Keller 2010, Marx, Keller & Beuing 2011, zit. in Walter & Schulz 2013, S. 71). Ein umfassender Überblick der Effekte ist in Walter und Schulz (2013, S. 71, Tab. 4) nachzulesen. Die Effekte sprechen für eine Möglichkeit des Trainings des induktiven Denkens und dessen

Einflüsse auf sprachliche Leistungen. Die Ergebnisse beziehen sich vorrangig auf das Training des induktiven Denkens. Sollten weitere Komponenten der Intelligenz nicht direkt trainierbar sein, könnte die Intelligenz dennoch als Teilkomponente mit anderen spezifischen Fähigkeiten gemeinsam zur Früherkennung einen Beitrag leisten (die wiederum trainierbar sein sollten) und ggf. einen Anteil der Fehlklassifikationen bei den vorgestellten Früherkennungsverfahren erklären.

Die referierten Untersuchungen machen deutlich, dass der Aspekt der Intelligenz bezüglich seines Einflusses auf Lese-Rechtschreibleistungen als Prädiktor nicht vernachlässigt werden darf, jedoch in Bezug auf die Interpretation der Korrelationen Vorsicht geboten ist. Der Intelligenz sollte neben der phonologischen Informationsverarbeitung, dem Wortschatz und dem Sprachverständnis im Rahmen von Untersuchungen im Bereich des Schriftspracherwerbs Beachtung geschenkt werden. Es steht jedoch außer Frage, dass Kinder mit weniger ausgeprägten kognitiven Fähigkeiten aufgrund eines anregenden Bildungsmilieus und der Förderung vorschulischer Kenntnisse und eines ihnen zuträglichen Unterrichts nicht zwangsläufig Misserfolge im Lesen und Rechtschreiben haben müssen, sondern sich in diesem schulischen Bereich durchaus positiv entwickeln können und umgekehrt Kinder mit einem hohen Intelligenzniveau nicht vor einer Lese-Rechtschreibschwäche gefeit sind. Zudem wirken sich kognitive Fähigkeiten der Kinder auf die Schulleistung im Allgemeinen aus und sind kein spezifischer Prädiktor für die Lese-Rechtschreibschwäche. Wichtig ist im Bereich der Diskrepanzforschung hier auch nochmals zu erwähnen, dass die lese-rechtschreibschwachen Schüler nicht aufgrund ihrer Intelligenzleistung in Subgruppen unterteilt werden können, auf deren Grundlage ein unterschiedlicher Zugang zur Förderung basieren kann (vgl. Kap. I.1.3). Eine Trainierbarkeit mit Effekten auf die Schulleistungen und damit auch auf die schriftsprachlichen Leistungen ist nach dem heutigen Forschungsstand keinesfalls auszuschließen.

Im Folgenden werden der Vollständigkeit halber umrisshaft weitere Persönlichkeitsvariablen, die mögliche Schutz- oder Risikofaktoren darstellen können, vorgestellt.

2.1.3 Weitere Persönlichkeitsvariablen

Schon früh konnte bei legasthenen Kindern nachgewiesen werden, dass diese ein Defizit in ihrem Durchhaltevermögen, sowie in ihrer Leistungsbereitschaft und Motivation aufweisen (vgl. Malmquist 1958, zit. nach Zimmermann 1980, S. 33). Ebenfalls konnten eine hohe

Ängstlichkeit sowie ein negatives Selbstbild festgestellt werden (vgl. Linder 1962, S. 34ff.). Es konnte jedoch nicht abschließend geklärt werden, ob diese emotionalen und motivatorischen Schwierigkeiten als Folge oder als Ursache der Lese-Rechtschreibschwäche auftreten. Sie stellen jedoch mögliche Risiko- und Schutzfaktoren einzelner Kinder dar, die trotz risikobehafteter Ausgangsbedingungen dennoch in der Schule gute Lese-Rechtschreibleistungen aufweisen können.

Auch aus der neurologischen Perspektive sind Variablen wie Stress oder Angst zu beachten, da diese das Lernen erschweren. Neuronen werden im Hippocampus gebildet und wandern dann zu den Stellen, an denen neue Vernetzungen entstehen sollen. Stresshormone haben auf verschiedene Zellstrukturen eine wachstumshemmende Wirkung und können eine Neubildung von Neuronen im Hippocampus sogar verhindern. Insbesondere Verlust- und Versagensängste, sowie ein geringes Selbstbewusstsein und wenig Freude an den Dingen, die man lernt, können eine Art Dauerstress für das Kind bewirken (vgl. Hofmann 2006, S. 106). Roth (1999, S. 25f.) merkt an,

> *„[…] , daß die Zusammenhänge keine Aussagen über die kausale Wirkungsrichtung der Faktoren erlauben. Die Persönlichkeitsmerkmale könnten Ursache aber auch Folge der Lese-Rechtschreibschwäche sein. Die Befunde weisen ausschließlich auf ein gleichzeitiges Auftreten spezifischer Persönlichkeitsmerkmale und Legasthenie hin.“*

Im vorherigen Abschnitt wurden die schülerspezifischen Variablen, die Einflüsse auf die Schülerleistungen haben, vorgestellt. Im nächsten Abschnitt sollen nun etwaige Umwelteinflüsse aus der familiären Umgebung der Kinder umrissen werden.

2.2 Variablen der familiären Umwelt

Lese-Rechtschreibstörungen kommen nach Warnke, Hemminger und Plume (2004, S. 2f.) in allen sozialen Schichten vor. Es tritt eine Häufung in sozial schwachen Familien mit höherer Kinderzahl, schlechten Wohnverhältnissen und einem Beruf auf unterem Anforderungsniveau des Vaters auf. Es werden jedoch laut Warnke, Hemminger und Plume (2004, S. 22) nur 9 % der Varianz der Lese-Rechtschreibleistungen durch die soziale Schichtzugehörigkeit erklärt. Valtin (1974a, S. 58ff.) konnte zeigen, dass eine geringe Schulbildung der Eltern lese-rechtschreibschwacher Kinder eine positive Korrelation zu den Lese-Rechtschreibleistungen aufweist. Auch Niemeyer (1975, zit. nach Klicpera, Schabmann & Gasteiger-Klicpera 2010, S. 187) konnte schon feststellen, dass 72 % der Legastheniker aus der unteren sozialen Schicht und nur 28 % aus der mittleren Schicht

stammten. Insgesamt sind positive soziale Einflüsse jedoch nicht nur für den Schriftspracherwerb, sondern für alle Bildungsprozesse und schulische Leistungen relevant. Die Resilienzforschung hat jedoch gezeigt, dass es schützende Faktoren gibt, sodass einige sozial benachteiligte Kinder dennoch gegen die sie umgebenden Umstände resilient sind (vgl. Wustmann 2004, S. 44ff.). Schutzfaktoren können aus personalen, familiären oder sozialen Ressourcen bestehen (vgl. ebd.). Hayiou-Thomas, Harlaar, Dale und Plomin (2006, S. 55f.) konnten beispielsweise anhand von Zwillingsstudien mit 1 270 Kindern zeigen, dass die Umwelteinflüsse eine bedeutende Auswirkung auf die schriftsprachlichen Leistungen der Kinder im Alter von sieben Jahren haben. Olson (2004, S. 30), der ebenfalls an den Zwillingsstudien des Colorado Learning Disabilities Research Center beteiligt war, verbindet genetische Dispositionen und Umwelteinflüsse zu einer Einheit: "A strong genetic influence on some aspect of human behavior or physiology does not mean that it cannot be changed by the environment."

Bezüglich des sozio-ökonomischen Hintergrunds konnten Deimel, Ziegler und Schulte-Körne (2005, S. 7) in einer Untersuchung mit 69 Grundschulklassen herausfinden, dass ein unspezifischer linearer Zusammenhang zwischen dem Bildungsabschluss der Mutter, den Wohnverhältnissen und den späteren Schriftsprachleistungen des Kindes besteht. Caughy (1996, S. 515ff.) konnte in einer Untersuchung mit 867 fünf- und sechsjährigen Kindern eine hohe Korrelation zwischen dem Bildungsgrad der Mutter sowie der häuslichen Umgebung und den Lese- und Mathematikleistungen feststellen. Aufgrund der Auswirkungen auf den mathematischen Bereich kann es sich beim sozio-ökonomischen Hintergrund nicht um einen spezifischen Faktor handeln. Auch Ptok und Dunkelmann (2011, S. 430) konnten anhand von N = 207 Vorschulkindern dagegen keinen korrelativen Zusammenhang zwischen der Schulbildung der Mutter und den Testleistungen zur phonologischen Bewusstheit der Kinder herausfinden.

Ähnlich wie bei den Persönlichkeitsvariablen ist für die Zusammenhänge der familiären und sozialen Einflüsse auf das Kind die Wirkungsrichtung weder geklärt, inwiefern es sich um ein Teil des Wirkungsgefüges handelt, noch wie diese Faktoren auch kompensiert werden können. Dies zeigt sich an Kindern, die aus einem sozial schwachen Elternhaus stammen und trotzdem Erfolg beim Lesen und Rechtschreiben haben.

Schon die PISA-Studie im Jahr 2000 konnte zeigen, dass besonders in Deutschland der Risikofaktor der soziökonomischen Bedingungen der Schüler in der Schule weit weniger ausgeglichen werden kann als in anderen Ländern.

> *„Die Leistungsunterschiede sind in keinem anderen Land so stark durch die soziale Herkunft bedingt wie in Deutschland (S. 39-40): Die Differenz zwischen SchülerInnen aus der oberen und der unteren Sozialschicht liegt bei 111 Punkten – in Korea, Finnland und Japan nur bei 25-50 Punkten (Baumert in: Rubner 2001). In anderen Ländern gleicht die Schule also soziale Benachteiligung wesentlich stärker aus als bei uns.“ (Brügelmann & Heymann 2002, S. 3)*

Im nächsten Kapitel wird auf den Einfluss schulischer und unterrichtlicher Variablen bezüglich der Lese-Rechtschreibleistungen der Kinder eingegangen, die ebenfalls zu einer Erklärung für unterschiedliche Ergebnisse der Untersuchungen zur Prognose von Lese-Rechtschreibschwierigkeiten beitragen können.

2.3 Schulische und unterrichtliche Variablen

In den 80er-Jahren fand ein Paradigmenwechsel statt. Die Wissenschaftler bezogen zunehmend die Ergebnisse aus der entwicklungs- und lernpsychologischen Forschung mit ein (vgl. bspw. Frith 1985, S. 300ff.; Günther 1986, S. 32f.; Scheerer-Neumann 1989b, S. 32f.). Der Fokus wurde immer mehr auf die Probleme beim Erlernen der Schriftsprache von lese- und rechtschreibschwachen Schülern gelegt. Es wurde zunehmend der Verlauf der LRS betrachtet und die Entwicklung der Forschung wurde durch die Phasenmodelle des Schriftspracherwerbs nochmals vorangetrieben (vgl. ebd.). Die Lese-Rechtschreibschwäche wurde nun nicht mehr als Dysfunktion, sondern als Entwicklungsverzögerung betrachtet. Die Förderung lese-rechtschreibschwacher Kinder wurde auf der Grundlage der theoretischen Erkenntnisse immer mehr durch spezifische Trainings geprägt. Insbesondere in den 90er-Jahren wurden dann auch die Vorläuferfertigkeiten für einen erfolgreichen Schriftspracherwerb und die lautlichen Aspekte der Sprache berücksichtigt. Die Förderung der phonologischen Bewusstheit vor und während des Erstlese- und Schreibunterrichts wurde aufgrund dieser Einsicht verstärkt und Präventionsmaßnahmen entwickelt (vgl. Mannhaupt 2006, S. 93ff.).

Der Schriftspracherwerb ist eine erlernte Fähigkeit und muss damit auch vom Lernprozess abhängig sein, der den individuellen Umwelteinflüssen des Lerners ausgesetzt ist. Nach Klicpera und Gasteiger-Klicpera (2010, S. 94ff.) ist der Lernerfolg insbesondere von der Qualität und dem Umfang des erhaltenen Unterrichts abhängig.

Bezüglich der schulischen Einflüsse setzte sich die Forschung mit den Lehrmethoden auseinander. Für den Anfangsunterricht scheint sich die analytisch-synthetische Unterrichts-

methode gegenüber der ganzheitlichen Methode durchzusetzen. Dieser Effekt erweckt den Eindruck über die Jahre abzunehmen und ist am Ende der Grundschule nicht mehr in dem Maße vorhanden (vgl. Ferdinand 1972, S. 105ff.). Die Diskussion über erfolgreiche Unterrichtsmethoden wurde in den Jahren auf Ansätze erweitert, die den Schülern Lust am Lesen und Schreiben vermitteln sollten. Lehrmittel und Lernkonzepte, die selbstgesteuert und strategieorientiert Spaß am Umgang mit Schriftsprache vermittelten, wurden in den Vordergrund gerückt (vgl. Grissemann 1995, S. 25ff.).

Marx (2007, S. 94) fasst die Befunde von Untersuchungen bzgl. der Lehrmethode in Schulen zusammen:

- *„Die meisten Kinder erlernen nach jeder Lehrmethode die Schriftsprache.*
- *Bei jeder Lehrmethode gibt es Kinder, die mehr oder weniger große Schwierigkeiten mit dem Erwerb haben.*
- *Die einzelnen Methoden betonen verschiedene Fertigkeiten, die in der Anfangszeit zu charakteristischen Schriftsprachresultaten führen.*
- *Je nach dem Grad der Vorerfahrung mit Schriftsprache sind die Lehrmethoden unterschiedlich effektiv. Zur Auflösung des Methodenstreits sind empirische Untersuchungen über den Schriftspracherwerb vonnöten und hilfreich.“*

Vellutino, Fletcher, Snowling und Scanlon (2004, S. 28) sammelten Hinweise aus vielfältigen Studien. Sie fassen zusammen, dass ein guter Lese-Unterricht die Kinder vor Defiziten im Lesen bewahren kann.

Die Studien zeigten, dass

> *“[...] comprehensive and well-balanced reading instruction that faciliates the acquisition of phonological awareness and letter-sound decoding skills along with other word-level skills, in addition to oral language and reading comprehension skills, can prevent long-term reading difficulties in children who would otherwise qualify for a diagnosis of reading disability.”*

Morais, Carey, Alegria und Bertelsson (1986, S. 323ff.) fanden heraus, dass die phonologische Bewusstheit aus dem bewussten Umgang mit der Schriftsprache und aus dem Erlernen der alphabetischen Strategie heraus verbessert wird. In ihrer Untersuchung konnten sie zeigen, dass sich phonemanalytische Leistungen nach einigen Monaten im

Anfangsunterricht bei den Schülern mit analytisch-synthetischem Erstunterricht signifikant verbessert haben im Vergleich zu Kindern, die an einem ganzheitlichen Erstunterricht teilgenommen hatten.

May und Okwumo (1999, S. 14) stellten in ihrer Untersuchung einen hohen Lernzuwachs der Kontrollgruppe im ersten Schuljahr fest, der darauf hinweisen könnte, dass für den Lernerfolg im Rechtschreiben der schulische Unterricht einen weit höheren Einfluss haben könnte, als das Training phonologischer Vorläuferfertigkeiten. Auch Richter (1992, S. 281) konnte feststellen, dass der Vorgehensweise im Unterricht ein großer Stellenwert zukommt und dies die Vorhersagekraft der Prädiktoren auf die Rechtschreibleistung beeinflusst.

Hatz (2015, S. 236ff.) verglich in seiner Studie ein reines Training der phonologischen Bewusstheit in der ersten Klasse mit einem Training in einer Kombination aus phonologischer Bewusstheit und Rechtschreibung im Vergleich zu einer Kontrollgruppe. Dafür wurden aus 793 Erstklässlern 113 Risikokinder berücksichtigt, die zusätzlich zum Deutschunterricht ein Kleingruppentraining erhielten. Die Ergebnisse konnten zeigen, dass unmittelbar nach dem Training keine Transfereffekte auf Lese- und Rechtschreibleistungen zu verzeichnen waren. Er konnte jedoch aufzeigen, dass die Kinder, die ein Training erhielten und in einer Klasse unterrichtet wurden, die nach dem Spracherfahrungsansatz die Schriftsprache erwerben, eine zu der Kontrollgruppe aus Fibelklassen vergleichbare Leistung der phonologischen Bewusstheit aufzeigen konnten. Dies wirkte sich jedoch nicht auf die schriftsprachlichen Fähigkeiten aus. Das zusätzliche Rechtschreibtraining konnte nur kurzfristige Effekte bewirken. Hatz (a. a. O., S. 278ff.) vermutet, dass die lehrgangsorientierte Gestaltung des Anfangsunterrichts mehr Effekte bringe, als ein additives Training der phonologischen Bewusstheit in der ersten Klasse.

Ein strukturierter Erstunterricht kann damit einen erheblichen Einfluss auf die Leistungsentwicklung der Schüler haben. Sollten Kinder im vorschulischen Bereich noch phonologische Defizite aufweisen, ist es möglich, dass diese durch einen strukturierten Anfangsunterricht mit einer Integration von phonologischen Elementen oder Trainings ausgeglichen werden kann. Immer mehr Erstlesewerke und Fibeln beziehen die phonologische Bewusstheit in ihre Konzepte mit ein (vgl. Valtin 2010, S. 8). Auch viele Lehrer sind für das Thema der phonologischen Bewusstheit durch die große Popularität auch in Fortbildungsveranstaltungen sensibilisiert und ziehen es häufig vor, eine explizite Phase zur Übung der phonologischen Bewusstheit vorab durchzuführen. Außerdem ist in den Fibeln eine Tendenz vom ganzheitlich orientierten Anfangsunterricht hin zu einem

Trainieren von Graphem-Phonem-Korrespondenzen zu verzeichnen, was auch phonologische Defizite ausgleichen könnte (vgl. Scheerer-Neumann 2004, S. 33ff.).

Die Veränderung des Anfangsunterrichts in deutschen Schulen könnte ein möglicher Grund sein für die schlechten prognostischen Vorhersagen der Früherkennungsverfahren. Diese konzentrieren sich hauptsächlich auf phonologische Fertigkeiten und Defizite in den Vorkenntnissen, die dann zu Beginn der Schulzeit im Unterricht bei einem Großteil der Kinder kompensiert werden können. Schneider und Küspert (2004, S. 220f.) relativieren dies jedoch, indem sie konstatieren, dass der Anfangsunterricht zwar die phonologische Analyseleistung steigert, dies aber nicht bei allen Kindern im gleichen Maße stattfindet.

Im ersten Abschnitt des zweiten Kapitels wurde nun erörtert, welche Inhalte der Früherkennungsverfahren prädiktiv auf die Lese-Rechtschreibleistungen wirken und ob die in den vorliegenden Verfahren genutzten Inhalte ausreichend die Vorläuferfertigkeiten des Lesens und Rechtschreibens abbilden. Die phonologische Informationsverarbeitung und insbesondere die phonologische Bewusstheit stellten sich dabei als die am gründlichsten untersuchten und den größten Erfolg versprechenden Variablen heraus, doch linguistische Vorfertigkeiten könnten einen prädiktiven, wenn auch nicht so spezifischen Charakter haben. Im zweiten Abschnitt wurde der Frage nach weiteren exogenen und endogenen Faktoren nachgegangen, die ebenfalls die hohe Zahl an Fehlklassifikationen in den jüngeren Untersuchungen erklären könnten. Hierbei scheinen sowohl kognitive Variablen, insbesondere die Intelligenz, jedoch vor allem auch sozio-ökonomische Einflüsse und speziell bezüglich der phonologischen Bewusstheit auch der veränderte Anfangsunterricht in deutschen Schulen eine zentrale Rolle zu spielen. Nun soll im dritten Kapitel eine weitere mögliche Ursache für die Fehlklassifikationen der Screening-Verfahren betrachtet werden, die sich auf deren methodische Vorgehensweise bezieht.

3 Methodik der Screening-Verfahren

In diesem dritten Teilbereich der Ursachensuche der Klassifikationsproblematik wird zunächst ein allgemeiner Überblick über die pädagogische Diagnostik gegeben, um dann im Sinne dieses Fundamentes die Vorgehensweise der klassischen Testverfahren und damit auch der vorgestellten Screening-Verfahren zu erläutern. Im Anschluss wird die Methodik der Diagnostik reflektiert und diskutiert.

3.1 Pädagogische Diagnostik

Die pädagogische Diagnostik ist ein Sektor der Erziehungswissenschaften, der hilft, pädagogische Entscheidungen systematisch und methodisch zu unterstützen. Dabei stehen häufig sogenannte Tests im Mittelpunkt der Betrachtung. Der Begriff „Test“ wird durch Lienert (1969, S. 7) folgendermaßen definiert:

> *„Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung.“*

Die Diagnostik ist darüber hinaus auch der Psychologie als Teilbereich zugeordnet:

> *„Diagnostik umfasst die Gesamtheit der Verfahren und Theorien, die dazu dienen, Verhalten und psychische Prozesse einzelner Personen oder auch Gruppen zu erforschen.“ (Bundschuh, Heimlich & Krawietz 2002, S. 50)*

Seit den 70er-Jahren hat sich die Diagnostik im pädagogischen Bereich jedoch immer mehr von medizinischer und psychologischer Diagnostik entfernt. Die pädagogische Diagnostik orientierte sich nun überwiegend an den Lernprozessen und am Kind selbst. Dehn (1994, S.21) formulierte dazu folgende Fragestellungen, die das Kind in den Mittelpunkt des Geschehens rückten:

- Was kann das Kind schon?
- Was muss es noch lernen?
- Was kann das Kind als nächstes lernen?

Die Bedürfnisse der Kinder in Hinblick auf eine adäquate Förderungsmöglichkeit stehen dabei im Mittelpunkt (vgl. Bundschuh, Heimlich & Krawietz 2002, S. 50f.). Die Verknüpfung von Diagnostik und Förderung wurde mit dem Begriff „Förderdiagnostik“ belegt.

> *„Förderdiagnostik gibt auf der Basis von Verhaltensbeobachtung und Verhaltensanalyse im Bereich Erziehung und Unterricht wichtige Impulse für pädagogisch-didaktisches Handeln und begleitet mittels Prozessdiagnostik den Fortgang der Förderung.“ (a. a. O., S. 51).*

Die pädagogische Förderung der Kinder mit Defiziten rückte in den Vordergrund, anstatt eine ausschließliche Selektionsentscheidung (z.B. die Überweisung in die Sonderschule) zu treffen (vgl. Schlee 2008, S. 123). Als kritische Anmerkung zur Förderdiagnostik sei der Artikel von Schlee (2008) „30 Jahre Förderdiagnostik – eine kritische Bilanz“ anzuführen. Schlee (2008, S. 122ff.) zweifelt an der Förderdiagnostik als sinnvolles Diagnostikinstrument. Er unterstreicht, dass Selektion und Förderung von Schülern sich nicht gegenseitig ausschließen. Die pädagogische Diagnostik wird häufig im gleichen Zug mit pädagogischem Handeln erwähnt. Solange pädagogisch gehandelt wird, wollen Erfolge der Anstrengungen gemessen werden. Ingenkamp und Lissmann (2008, S. 13) halten folgende Definition für angemessen:

> *„Pädagogische Diagnostik umfasst alle diagnostischen Tätigkeiten, durch die bei einzelnen Lernern und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren. Zur Pädagogischen Diagnostik gehören ferner die diagnostischen Tätigkeiten, die die Zuweisung zu Lerngruppen oder zu individuellen Förderungsprogrammen ermöglichen sowie die mehr gesellschaftlich verankerten Aufgaben der Steuerung des Bildungsnachwuchses oder der Erteilung von Qualifikationen zum Ziel haben.“*

Damit liegt der Fokus einer pädagogischen Diagnostik darin, die für den Lernenden passenden Förderungs-, Platzierungs- und Selektionsmaßnahmen auszuwählen und aktuelle Lernvorgänge und Lernerfolge zu optimieren (vgl. a. a. O., S. 14ff.).
Schründer-Lenzen (2009, S. 245) konstatiert, dass pädagogische Diagnostik auch immer wieder eine Neufestlegung von Förderzielen und damit auch eine Überprüfung der festgelegten Ziele nach sich zieht: „Pädagogische Diagnostik wird heute als lernwegsbegleitende Aufgabe gesehen, die die Fortschritte der Kinder sichtbar machen soll.“ (vgl. ebd.). Für diagnostische Handlungen können nach Ingenkamp & Lissmann (a. a. O., S. 39ff.) folgende Aspekte von Bedeutung sein:

- Vergleich
- Analyse
- Prognose
- Interpretation
- Mitteilung
- Wirkungskontrolle

Beim *Vergleich* wird das Verhalten eines Menschen im Verhältnis zu sich selbst oder mit dem Verhalten eines anderen Individuums zu gleichen oder unterschiedlichen Zeitpunkten anhand unterschiedlicher Kriterien verglichen. Die *Analyse* dient der Suche nach Gründen für ein Verhalten, also warum ein Verhalten vom Standard abweicht, während die *Prognose* zukünftiges Verhalten vorhersagt. Bei der *Interpretation* werden diagnostische Daten geordnet, bewertet und gewichtet. Als Ergänzung dienen die *Mitteilung* der Diagnose und auch die *Kontrolle der Wirkungen* (ebd.).

Die diagnostische Kompetenz von Lehrkräften und Erziehern beschränkt sich häufig auf subjektive Einschätzungen, die meist unsystematisch erhoben oder im Alltag gewonnen worden sind (vgl. Hesse & Latzko 2009, S. 25). Weinert und Schrader (1986, S. 27) schlagen daher eine zweistufige pädagogische Diagnostik vor. Sie kombinieren eine subjektive, handlungsleitende Diagnostik durch die zuständigen pädagogischen Fachkräfte mit einer standardisierten Diagnostik, die möglichst objektive Urteile hervorbringt. Ein großes Gebiet der pädagogischen Diagnostik im schulischen und vorschulischen Bereich betrifft die Messung von Leistungsverhalten, wie sie auch in der vorliegenden Studie durchgeführt wird.

> *„Leistungsverhalten im weiteren Sinne sind alle Reaktionen eines Individuums auf die Anforderung, ein nach vorgegebenen Kriterien möglichst gutes Ergebnis zu erzielen oder zumindest einem gesetzten Minimalstandard zu genügen." (Fay & Stumpf 1999, S. 380)*

Bei Testverfahren zur Messung von Leistungsdaten, wie beispielsweise Frühdiagnostik der Lese-Rechtschreibschwäche (vgl. Kap. I.2.2), wird das Kind durch eine Instruktion aufgefordert, die gestellte Aufgabe bestmöglich zu erledigen (vgl. Fay & Stumpf 1999, S. 381).

3.2 Klassifikation herkömmlicher Testverfahren

Schon seit längerer Zeit steht zunehmend in der Diskussion, ob Diagnostik und damit einhergehende Normen generell erstellt werden sollten, oder ob für bestimmte Subgruppen differenzierte Normen (Geschlecht, Schulart, Migrationshintergrund,…) genutzt werden müssen. Diese Entscheidung ist von der Grundintention der Diagnostik abhängig. Lese-Rechtschreib-Diagnostik wird vielfältig eingesetzt:

- zur Selektion (Liegt eine Lese-Rechtschreibschwäche vor?)
- zur Förderung (Welche Fördermaßnahmen sind sinnvoll und wo steht das Kind?)
- zur Bewertung der Qualität von Unterricht (Didaktik, Methodik)
- zur Breitenmessung (Wie viele Schüler lesen oder schreiben mit welcher Leistung in welcher Klassenstufe?)

Zur Bestätigung einer Lese-Rechtschreibschwäche dürfen Subgruppenentscheidungen nicht von Relevanz sein, da dadurch finanzierte Fördermaßnahmen eine Bevorzugung bestimmter Gruppen hervorrufen würde und dies unter anderem auch politische Folgen hätte. Mädchen beispielsweise würden durch tendenziell höhere Rechtschreibleistungen aufgrund einer geschlechtsspezifischen Normierung mit insgesamt höheren Leistungen eher eine Lese-Rechtschreib-Schwäche attestiert bekommen als Jungen, die mit gleichem Alter ähnliche Leistungen (in Rohwerten) zeigen. Diese Gründe lassen sich ebenfalls auf die Messungen zur Früherkennung und mögliche Präventionsmaßnahmen einer Lese- und Rechtschreibschwäche übertragen. Zur Förderung oder auch für die Bewertung und Verbesserung von unterrichtlichen Maßnahmen sowie zur Breitenmessung erscheint eine subgruppenspezifische Einteilung als sinnvoll, da die Schüler individuell unterschiedlich sind und zum Beispiel je nach Subgruppe eine andere Methodik zielgerecht verwendet wird (vgl. auch Hasselhorn, Marx & Schneider 2008, S. 4ff.). Die Ziele, Aufgaben und die Qualität von pädagogisch-psychologischer Diagnostik sind von den Handlungs- und Entscheidungssituationen abhängig. Pawlik (1976 zit. nach Wild & Krapp 2001, S. 521) unterscheidet zwischen Modifikations- und Selektionsstrategien (vgl. Tab. 25). Er stellt dabei als idealtypische Unterscheidung vier Strategiemuster in den Vordergrund.

Tab. 26: Idealtypische Handlungs- und Entscheidungsstrategien (nach Wild & Krapp 2001, S. 523)

Objekte der Handlung	Handlungsstrategie	
	Modifikationsstrategie	Selektionsstrategie
Personen	Personenmodifikation	Personenselektion
Umweltbedingungen	Bedingungsmodifikation	Bedingungsselektion

Die Personenmodifikation bezieht sich auf die explizite Förderung des Kindes selbst. Die Bedingungsmodifikation bedeutet eine Anpassung der Umweltbedingungen des Betroffenen. Bei der Selektionsstrategie werden die betroffenen Kinder in Systeme, wie beispielsweise in die verschiedenen Schulformen eingeteilt (Personenselektion) oder sie wählen sich selbst interessengeleitet beispielsweise einen Leistungskurs in der Schule aus (Bedingungsselektion). Die Früherkennungsverfahren lassen sich in diesem Strategiemuster ebenfalls wiederfinden. Sie können zur Personenmodifikation eingesetzt werden. Hier wird das Kind bspw. mit einem Förderprogramm, das sich an seinen Schwächen orientiert, unterstützt. Alternativ ist es auch möglich, dass eine Personenselektion in Subgruppen auftritt, z.B. Risikokinder, die in einem zusätzlichen Förderkurs beim Erwerb der Vorläufer-fertigkeiten des Lesens und Schreibens unterstützt werden. Hier wird nicht institutionell

selektiert, jedoch innerhalb der Einrichtung. Bedingungsmodifikation und Bedingungsselektion sind bisher keine gängigen Handhabungen im vorschulischen Bereich, solange diese sich gezielt auf die hier vorgestellten Früherkennungsverfahren beziehen.

Die klassische Testtheorie besteht aus einer Sammlung von Aussagen. Die Aussagen wurden aus einigen Axiomen (Grundsätze der Psychologie) abgeleitet, die in der Regel mathematisch-statistisch beschrieben sind (vgl. Wild & Krapp 2001, S. 523). Der Zweck der klassischen Testtheorie ist vorwiegend die Erstellung und Bewertung von Messungen zeitlich stabiler Personenmerkmale verpflichtet. Es wird dabei zwischen beobachtetem Wert und wahrem Wert unterschieden.

> *„Ein zentrales Ziel der klassischen Testtheorie besteht darin, Verfahrensweisen und Gütekriterien zu entwickeln, die es erlauben, den nicht direkt beobachtbaren hypothetischen, ‚wahren Wert' möglichst messfehlerfrei zu erfassen und den Fehleranteil der Messwerte möglichst genau, d.h. quantitativ zu beziffern." (Wild & Krapp 2001, S. 524, nach Lienert 1969 und Lienert & Raatz 1998).*

3.3 Kritische Betrachtung der klassischen Testtheorie

Die klassische Testtheorie hat sich in den letzten Jahrzehnten in der psychologischen Diagnostik bei standardisierten Testverfahren etabliert. Die Qualität einer Messung konnte durch die Gütekriterien vermehrt bewiesen werden, wurden diese bei einem Messverfahren nicht angeführt, konnte von der Anwendung des Testverfahrens abgeraten werden (vgl. Wild & Krapp 2001, S. 528).

Standardisierte Tests können bisweilen innerhalb kürzester Zeit verschiedenste Informationen, die bisher noch unbekannt waren, liefern, Hinweise für eine Förderung geben sowie eine inadäquate Beschulung des Kindes durch große Differenzen in IQ und Sprache im Verhältnis zu schulischen Leistungen entdecken. Werden die Kompetenzen eines Kindes falsch eingeschätzt, wäre dies durch ein standardisiertes Verfahren revidierbar. Außerdem kann ein Test auch zu einem besseren Verständnis der Gesamtsituation (Schule, Elternhaus, Kind) beitragen, was besonders im sonderschulpädagogischen Bereich weit verbreitet genutzt wird (vgl. Bundschuh, Heimlich & Krawietz 2002, S. 286f.).

Von verschiedenen Seiten kam jedoch auch Kritik insbesondere im pädagogisch-psychologischen Bereich an der klassischen Testtheorie auf:

(a) Es wird hinterfragt, ob Konstrukte wie Intelligenz, Motivation oder Angst überhaupt gemessen werden können. Dies ist auch bei dem Konstrukt „Lese-Rechtschreibschwäche“ zu beobachten, da hier die Definitionen so weit auseinander klaffen, dass es schwierig ist, den wahren Wert der Lese-Rechtschreibschwäche statistisch zu schätzen. Kritiker sprechen auch von „Hypothetischen Konstrukten“ (vgl. Bundschuh, Heimlich & Krawietz 2002, S. 285).

(b) Die Problematik der „Hypothetischen Konstrukte“ lässt sich auch auf die Messung von Eigenschaften in den Items und Aufgabenstellungen übertragen. Es stellt sich die Frage, ob ausreichend sichergestellt werden kann, dass die Aufgaben eines nicht eindeutig definierten Konstrukts tatsächlich die gewünschten Eigenschaften messen und diese Eigenschaften damit das hypothetische Konstrukt repräsentieren.

(c) Die Festlegung der Grenzwerte bei Klassifikationen ist willkürlich und bei verschiedenen Autoren auch oft different (vgl. Bundschuh, Heimlich & Krawietz 2002, S. 285).

(d) Durch den Vergleich mit einer Norm ergibt sich insbesondere für Kinder mit Schwierigkeiten ein „Negativismus“ von Defizitbeschreibungen, anstatt sich an den positiven Leistungen der Kinder zu orientieren. Dies kann sich auch aufgrund des Prinzips der „self-fulfilling-prophecy“ auf das Verhalten und die Entwicklung der Kinder auswirken.

(e) Mit einer Diagnostik wird Erfolg und Leistung von Kindern gemessen. Diese werden anhand dieser Kriterien klassifiziert und ggf. auch selektiert. Diese Wertung wirkt sich ebenfalls auf die Wahrnehmung gegenüber Kindern mit Defiziten aus (vgl. a. a. O., S. 285f.).

(f) Die Gefahr einer Diagnostik im pädagogisch-psychologischen Bereich besteht zudem in der Messung von Teilprozessen oder bestimmten Funktionsbereichen (vgl. a. a. O., S. 286). Dies hat sich auch in der Diagnostik bei Lese-Rechtschreib-schwierigkeiten gezeigt, indem von einer minimalen kognitiven Dysfunktion oder von legasthenietypischen Fehlern ausgegangen und diese auf eine eingeschränkte teilspezifische Diagnostik zurückgeführt wurde. In der Früherkennungsmessung kann dies ebenfalls eine Problematik darstellen, andere mögliche Prädiktoren oder auch Schutz- und Risikofaktoren wurden bereits in Kapitel II.1 und Kapitel II.2 angeführt.

(g) Die weitere Kritik richtet sich auf die Frage der Lebenswirklichkeit der Tests. Die Aufgaben orientieren sich meist nicht an der Erfahrungswelt der Kinder und liefern

nicht zwingend Hinweise für eine adäquate Förderung. Die Aufgabenstellungen selbst sind vielmals künstlich und von der Realität der Kinder weit entfernt (vgl. ebd.).

(h) Infolge der starken Führung und Kontrolle in den Tests (z.B. auch aufgrund der Durchführungsobjektivität) wird die Interaktion mit dem Kind eingeschränkt (vgl. ebd.). Die Leistungen können dann aufgrund einer fehlenden Beziehung zum Kind und dem entsprechend fehlendem Vertrauen nicht gemessen werden.

(i) Anhand psychologisch-pädagogischer Tests wird häufig nur einmalig die Leistungsfähigkeit eines Kindes gemessen. „Aus pädagogischer Sicht sind vielmehr solche Verfahrensweisen wünschenswert, die Veränderungen eines Merkmals über die Zeit, etwa im Rahmen einer Ausbildungsmaßnahme oder therapeutischen Intervention, […] reliabel erfassen können." (Wild & Krapp 2001, S. 528).

Hinsichtlich der Entstehungsgeschichte der pädagogisch-psychologischen Diagnostik aus dem medizinischen Bereich heraus war die Identifikation von *Unterschieden* zwischen Individuen im Fokus der Aufmerksamkeit (vgl. Wild & Krapp 2002, S. 528):

> *„Das, worauf es der Pädagogischen Psychologie aber oft ankommt, nämlich die Registrierung von lern- und sozialisationsabhängigen Veränderungen kognitiver, motivationaler oder sozialer Merkmale über die Zeit, wird mit den Methoden der klassischen Testtheorie nur unzureichend geleistet."*

Wild und Krapp (ebd.) beschreiben die klassische Testtheorie jedoch nicht als „wertlos", da es auch in pädagogischen Handlungsfeldern in vielen Situationen einer Klassifikation oder Selektion bedarf und damit relativ stabile Aspekte des Verhaltens gemessen werden können. Dies sei insbesondere in der Frühdiagnostik von Vorteil, um bestimmte Risikofaktoren zu bestimmen, die auf ein späteres Defizit schließen lassen. Außerdem hängt die Beurteilung eines Testverfahrens auch von der jeweiligen individuellen Handlungs- und Entscheidungssituation ab.

Insbesondere im Zuge der Veränderungen des Schulsystems hin zu einer inklusiven Schule muss jedoch in Ergänzung zu den bisher häufig verwendeten klassischen Testverfahren, die zur Selektion und Einteilung in bestimmte Fördermaßnahmen dienten, nun weitläufig eine Entwicklung hin zur Evaluation von Lernprozessen des einzelnen Schülers erfolgen (vgl.

auch Kap. I.1.1). Dies können klassische Testverfahren in dem Sinne nicht leisten. Sie sind auf eine einmalige Messung mit einer Klassifikation der Schüler ausgelegt.

Wie die vorangegangene kritische Diskussion der Screening-Verfahren gezeigt hat, besteht durch das Merkmal der „Einmalmessung“ in diesen Verfahren die Gefahr, dass die Einflussvariablen nicht verlässlich erfasst werden. Es ist zu vermuten, dass sich bei wiederholten Messungen bessere Prognosen treffen lassen und damit die Fehl-klassifikationen eingeschränkt werden könnten. Dazu finden wir insbesondere in den USA neuere Ansätze, die den pädagogischen Kontext und insbesondere den Sachverhalt der Mehrfachmessung sowie einer Lernfortschrittsmessung statt einer ausschließlichen Messung zur Selektion und Klassifikation von Kindern miteinbeziehen. Im Folgenden wird das auf diesen Überlegungen aufbauende alternative diagnostische Verfahren der curriculumbasierten Messung vorgestellt, das auch Einfluss auf die Konstruktion des Messverfahrens dieser vorliegenden Arbeit haben wird.

III. Curriculumbasierte Messung (CBM) - eine alternative Diagnostikmethode zur Unterstützung der Prävention

In diesem Kapitel wird ausführlich die curriculumbasierte Messung als alternative Methodik zur Ein- oder Zweifachmessung vorgestellt, da diese Art der Messung das Fundament der in dieser Untersuchung dargestellten curriculumbasierten Messung „Phonolino“ sein wird und damit zur Begründung der ausgewählten Messmethode beiträgt.

Kinder kommen mit den unterschiedlichsten Lernvoraussetzungen in die Schule. Der Anspruch, den Unterricht oder auch die vorschulischen Inventionsmaßnahmen an die individuellen Lernvoraussetzungen der Kinder anzupassen, ist nur auf Basis einer spezifischen und differenzierten Diagnostik möglich. Wie im vorherigen Kapitel II.3.3 beschrieben, können klassisch ausgerichtete normbasierte Testverfahren zwar nützliche Daten liefern, jedoch konnte aufgezeigt werden, dass sie bzgl. vieler Aspekte auch Mängel aufweisen. Sie sind insbesondere für den Einsatz durch die unterrichtende Lehrperson oft aus ökonomischen Gesichtspunkten nicht geeignet und werden häufig ausschließlich von externem Personal durchgeführt. Deno (1985, S. 220) beschreibt das Problem der Anwendung in der Praxis wie folgt:

> *“[…] teachers preference is to use informal observations based on their interactions with students as the data source. Unfortunately, the reliability and validity of this source is at best unknown and at worst inadequate.”*

Dies konnten auch Fuchs, Deno und Mirkin (1984, S. 455) feststellen. Sonderschullehrer nutzen vorwiegend unsystematische Beobachtungsmethoden zur Überprüfung von Fähigkeiten, da sie in herkömmliche Tests kein Vertrauen setzen. Die mangelnde Anwendung der klassischen Testverfahren bedarf damit der Suche nach Alternativen. Walter (2008b, S. 63ff.) fasst die zentralen Kritikpunkte wie folgt zusammen:

- Klassische Leistungstests vernachlässigen die zeitliche Komponente und messen die Veränderungen über die Zeit nicht, da sie häufig nur einmalig durchgeführt werden. Ob eine Förderungsmaßnahme den betroffenen Kindern einen Vorteil verschafft oder gar nicht wirkt, kann somit nicht schnell genug festgestellt werden.
- Klassische Tests sind nicht veränderungssensibel, kurzfristige Leistungsveränderungen von Schülern bleiben dadurch unentdeckt.

Als alternatives Diagnostikkonzept wurde schon in den 80er-Jahren ein Entwurf zur curriculumbasierten Messung (CBM)[16] von Deno (1985, S. 221ff.) vorgestellt. CBM bezieht sich auf curriculare Inhalte und wird hauptsächlich zur Messung basaler Fähigkeiten eingesetzt. „curriculumbasiert" bedeutet dabei nicht, dass es sich auf die Vorgaben der Landesregierungen bezieht, sondern, dass die CBM durch Leistungstest den aktuellen Lernstand und den darauffolgenden Lernzuwachs in Bezug auf ein langfristiges Ziel erfassen soll (vgl. Börnert 2014, S. 4).
Das Verfahren ist auf eine Veränderungsmessung ausgerichtet, d. h. es kann den Lernfortschritt eines Kindes kontinuierlich darstellen. CBM bildet einen Entwicklungsverlauf des Kindes ab, da die Anwendung des Verfahrens nach regelmäßigen Zeitabständen erfolgt (z.B. monatlich oder wöchentlich). Zu mehreren Zeitpunkten werden Stichproben erhoben, die zwar immer wieder die gleiche Fähigkeit messen, jedoch äußerlich immer eine neue Form annehmen müssen. So kann der Leistungsverlauf einer bestimmten Fähigkeit dokumentiert (vgl. auch Klauer 2006, S. 17f.) und Aussagen über die Leistungen in Verbindung mit durchschnittlichen Leistungen von Schülern der gleichen Klassenstufe getroffen werden. Als einen bedeutenden Vorteil der CBM heben Diehl und Hartke (2012, S. 195) deren Sensitivität für Veränderungen, sowie deren formative Eigenschaft in Bezug auf Interventionsmaßnahmen hervor.

In den folgenden Abschnitten wird zunächst kurz die Geschichte der Entstehung und Entwicklung der CBM dargestellt. Im Anschluss werden deren Definition, Aufbau, Einsatzbereiche und Standardisierungsmöglichkeiten sowie deren Testgüte erläutert. Aktuelle Studien aus dem deutschen wie auch aus dem amerikanischen Raum beschreiben die Relevanz der curriculumbasierten Messung. Hierbei wird insbesondere auf den Aspekt der vorschulischen Messungen im Bereich der Früherkennung der Lese-Rechtschreibschwäche eingegangen. Zuletzt wird die praktische Relevanz der CBM verdeutlicht, indem die Interventionsansätze ATI (aptitude-treatment-interaction) und RTI (response-to-intervention) in Beziehung gesetzt werden.

In diesem Kapitel wird eine methodische Alternative, die curriculumbasierte Messung, für die Früherkennung von Lese- und Rechtschreibproblemen vorgestellt. Dazu werden folgende Fragestellungen beleuchtet:

16 Im Amerikanischen auch unter „curriculum-based measurement" bekannt (vgl. Deno 1985).

1. Was ist CBM und welche Möglichkeiten bietet diese?
2. Genügt die CBM den Standards zu Gütekriterien von Testverfahren?
3. Wie ist der methodische Ansatz der CBM mit einer Prävention durch Intervention vereinbar?
4. Welche Vorteile bietet diese Messtechnik gegenüber herkömmlichen Verfahren?

Um die in der vorgelegten Untersuchung konzipierten, curriculumbasierten Messung Phonolino zu evaluieren, werden außerdem Erkenntnisse zur Konzeption und Evaluation von curriculumbasierten Messverfahren beschrieben, die eine Grundlage für die im zweiten Teil dieser Arbeit vorgestellten Methoden und Ergebnisdarstellungen sowie Interpretationen liefern. Zunächst wird nun auf die Geschichte und Entwicklung curriculumbasierter Verfahren eingegangen.

1 Geschichte und Entwicklung

Der Unterschied zwischen formativer und summativer Diagnostik wurde bereits im Jahr 1967 von Scriven (vgl. Klauer 2014a, S. 1) beschrieben. Formative Evaluation bezeichnete er als die wiederholte Evaluation eines noch nicht abgeschlossenen Lernprozesses, während die summative Evaluation einen Lernstand am Ende einer Lernperiode erhebt. Die Begrifflichkeiten wurden über die Jahre weiterentwickelt, hinzu kam die formative, summative, und konfirmative Evaluation oder auch das formative bzw. summative Assessment. Klauer (a. a. O., S. 3) beschreibt, dass „das terminologische Durcheinander (ist) bislang nicht wesentlich besser geworden“ ist. Stanley Deno griff mit seinem Forschungsteam der Universität Minnesota bereits im Jahr 1972 das Konzept auf (vgl. ebd.) und entwickelte die curriculumbasierte Messung in einem über sechs Jahre angelegten Forschungsprogramm im Bundesstaat Minneapolis in den USA (vgl. Deno 1985, S. 230f.). Er beschränkte sich dabei hauptsächlich auf sonderpädagogische Einsatzfelder und entwickelte ein Messverfahren, das Lehrern bei täglicher Messung eine Evaluation des Lernprozesses der Schüler ermöglichen sollte, um die Förderung von Schülern effektiver gestalten zu können. Unter dem Wort „curriculumbasiert“ verstand er jedoch nicht die Testung über allgemeine Schulleistungstests, sondern es sollte sich um Messungen handeln, die Leistungen überprüfen, die im aktuellen Unterricht derzeit behandelt werden (vgl. Klauer 2014a, S. 3f.). Diese Messungen sollten möglichst auch in kürzeren Abständen wiederholt

werden, um den Überblick über die Lernentwicklung über einen längeren Zeitraum zu erhalten.

Denos Ziel war es (vgl. Fuchs & Fuchs 2002, S. 64) ein Messsystem zu entwerfen, das der Lehrer effektiv und ökonomisch einsetzen kann, um aussagekräftige Informationen zum Stand und zum Lernzuwachs der Leistungen der Schüler zu erhalten, sowie um Aussagen über die Effektivität einer Intervention treffen zu können und diese anhand der Daten besser planen zu können:

> *"The primary goal of the research program was to develop measurement and evaluation procedures that teachers could use routinely to make decisions about whether and when to modify a student's instructional program." (Deno 1985, S. 221)*

Es entstand das CBM-Konzept „data-based program modification" (DBPM), in dessen Rahmen eine Reihe von Entwicklungsbeobachtungsverfahren erstellt wurde, die ebenfalls die Gütekriterien herkömmlicher Testverfahren berücksichtigen (vgl. Diehl & Hartke 2007, S. 197).

Deno (1985, S. 221) formulierte in diesem Zusammenhang Kriterien zur Entwicklung von curriculumbasierten Messungen. Die CBM sollten:

- reliabel und valide,
- einfach und effizient,
- einfach zu verstehen und
- kostengünstig sein.

Deno (1985, S. 221ff.) forschte anhand dieser Prämissen nach Aufgabentypen, die täglich erhoben und gemessen werden und die den Lernzuwachs in kurzen Zeitspannen ermitteln können. Deno und Fuchs (1987, S. 2ff., vgl. auch Fuchs & Fuchs 2002, S. 64) hinterlegten ihren theoretischen Ansätzen dazu eine 3 x 3 - Matrix, die zum Ziel hatte, ein systematisches Forschungsprogramm aufzustellen. In den Zeilen dieser Matrix sollten folgende Fragen durch ihre Forschungsarbeiten geklärt werden:

- Was wird gemessen? (What to measure?)
- Wie wird gemessen? (How to measure?)
- Wie werden die daraus resultierenden Daten genutzt? (How to use the resulting database?)

In den Spalten hinterlegten sie die Kriterien, die die Antworten zu diesen Fragen bringen sollten:

- Technische Angemessenheit (technical adequacy)
- Treatment Validität (treatment validity)
- Durchführbarkeit (feasibility)

In den USA wurden von dem Zeitpunkt an zahlreiche Untersuchungen zur CBM auch außerhalb des speziellen sonderpädagogischen Umfeldes durchgeführt (vgl. Deno 2003, S. 185ff.). Espin und Wallace (2004, in Fuchs 2004, S. 188f.) konnten bis 2004 über das „Research Institute on Progress Monitoring“ 585 Dokumente und 141 empirische Studien identifizieren, die sich mit dem Thema der curriculumbasierten Messung auseinandersetzen. Dabei standen vornehmlich die technische Durchführung, die Testgütekriterien, die Validierung der Ergebnisse und auch die Effekte auf das Unterrichtsgeschehen im Mittelpunkt. Im Rahmen dieser Forschung etablierten sich in der Literatur unterschiedliche Begrifflichkeiten, die sich in ihrem Einsatz in Nuancen unterscheiden: „simple measures“, „curriculum-based assessment“ (CBA)[17], „data-based assessment“ und letztendlich auch der hier ins Deutsche übersetzte und genutzte Begriff der „curriculum-based measurement“ (CBM, curriculumbasierte Messung; vgl. Shinn & Bamonto 1998, S. 2ff.). „Curriculum-based measurement“ ist der in den USA am häufigsten verwendete Begriff und wird aus diesem Grund auch in dieser Arbeit mit der Abkürzung CBM oder der deutschen Übersetzung „curriculumbasierte Messung“ verwendet.

Die Forschung zur CBM wurde insbesondere durch die bildungspolitischen Veränderungen in den USA vorangetrieben. In den Vereinigten Staaten von Amerika wird seit den 70er-Jahren im Zuge der „accountability“-Politik zur curriculumbasierten Messung geforscht. Nach dem Bildungsgesetz „No Child Left Behind Act“ (NLCB), das im Januar 2002 durch eine Gesetzesinitiative von George W. Bush verabschiedet wurde, müssen nun 95 % der Schüler die in dem jeweiligen Bundesstaat geltenden Normen im Lesen und in Mathematik erfüllen (vgl. Klauer 2006, S.16; Ardoin, Witt, Suldo, Connell, Koenig, Resetar, Slider & Williams 2004, S. 218f.; Fuchs & Fuchs 2004, S. 26ff.). Die Schulen werden für die Lernfortschritte der Schüler verantwortlich gemacht, d. h. sie sind aufgefordert, besonders den Lernfortschritt schwacher Schüler zu beobachten. Die Schulen sind in der

17 CBM ist beispielsweise eine Form der CBA. Sie wird im Vergleich zur CBA als eine langfristige Messung angelegt, die CBA wird dagegen gewöhnlich als kurzfristige Überprüfung von einzelnen Fähigkeiten angewendet (vgl. Madelaine & Wheldall 1999, S. 73f., Hosp & Hosp 2003, S. 11; VanDerHeyden 2005, S. 28f.).

Verantwortung, dass bis zum Schuljahr 2013/2014 sogar 100 % ihrer Schüler die Normen des jeweiligen Bundesstaates erfüllen (vgl. Klauer 2006, S. 16). Dazu wurden flächendeckend Schulleistungstests in allen Bundesstaaten der USA eingeführt. Schulen, deren Schüler diese Normen nicht erfüllen können, werden sanktioniert.

Fuchs (2004, S. 190ff.) betrachtet CBM nach jahrelanger Forschung als ein validiertes Verfahren. Die Expertenkomitees der USA setzen aufgrund der Forschungsergebnisse das Verfahren für Lernfortschrittsmessungen auf die Liste der anerkannten Verfahren (vgl. Ardoin et al. 2004, S. 219). Die „President's Commission on Excellence in Special Education" empfahlen CBM zur Überprüfung der Effektivität von Interventionen als alternatives Messverfahren zu herkömmlichen standardisierten Tests. Es erwies sich zudem, dass CBM-Messungen die Schüler identifizieren können, die bei den landesweiten Tests in den USA mit oder ohne Erfolg abschneiden (vgl. Shapiro, Keller, Lutz, Santoro & Hintze 2006, S. 19ff., Deno 2003, S. 184).

Die curriculumbasierte Messung (CBM) ist in Deutschland in der Praxis eine bisher wenig bekannte Möglichkeit der Diagnostik, rückt jedoch durch bedeutende Untersuchungen der letzten neun Jahre (vgl. insbesondere Walter 2008b, 2009a, 2010a, 2011b, 2013, sowie das Rügener Inklusionsmodell der Universität Rostock, vgl. bspw. Mahlau, Blumenthal, Diehl, Schöning, Sikora, Voß & Hartke 2014 oder auch Strathmann und Klauer 2008, 2010) und durch schulstrukturelle Veränderungen wie die inklusive Beschulung (vgl. Liebers & Seifert 2012, S. 1) immer weiter in den Mittelpunkt der Betrachtung. Die curriculumbasierte Messung wurde in verschiedenen Forschungsarbeiten bspw. von Klauer (2006, S. 16ff.) und Diehl und Hartke (2007, S. 195ff.) aufgegriffen. Der Ansatz des diagnostischen Verfahrens der Lernfortschrittsmessung war bereits in der Forschung zur Förderdiagnostik zu finden:

> *„Insofern gibt es also gewisse Parallelen zur Förderdiagnostik (z.B. Kornmann, 1982), die allerdings an der Diskrepanz zwischen Wollen und Können gescheitert ist (Schlee, 1985)."* (Klauer 2006, S. 16)

Im US-amerikanischen Bereich wird diese Art der Diagnostik bereits seit mehr als 30 Jahren praktiziert. Das Konzept kann in ähnlicher Form auf den deutschen Kulturbereich übertragen werden, wie bspw. die ersten Studien mit positiven Ergebnissen von Walter (2008b, 2009a, 2010a, 2011b, 2014a) im Bereich des Lesens zeigen konnten. Normen und Entwürfe des oben beschriebenen Messverfahrens (CBM) existieren in Deutschland zum gegenwärtigen Zeitpunkt nur in spezifischen Bereichen. So entwickelte bspw. Walter einschlägige Messverfahren auf Basis der curriculumbasierten Messung (CBM-LL, eine CBM auf der

Basis der Technik des 1-Minute-Lauten-Lesens, Walter 2008; LDL, die Lernfortschrittsdiagnostik Lesen, Walter 2010a und die VSL, die Verlaufsdiagnostik sinnerfassenden Lesens, Walter 2014a; 2011a, S. 106ff.), Diehl und Hartke (2012) ein Inventar zur Erfassung der Lesekompetenz (IEL-1) und die Quop-Werkstatt der Universität Münster unter der Leitung von E. Souvignier Online-Materialien, denen eine positive Wirksamkeit belegt wurde (vgl. www.quop.de).

Details zum curriculumbasierten Messen, etwaigen Forschungsergebnissen und Überlegungen zu möglichen Einsatzgebieten in Bezug auf Fördermaßnahmen werden in den nachfolgenden Kapiteln erläutert.

2 Beschreibung und Aufbau

Im Folgenden wird definiert, was unter einer curriculumbasierten Messung zu verstehen ist, wie diese durchgeführt und wie anhand der Daten der Lernzuwachs dokumentiert wird. Anschließend werden mögliche Einsatzbereiche der curriculumbasierten Messung beleuchtet. Im letzten Teil werden Aufgabenstichproben beschrieben und erläutert, wie diese erstellt werden können.

2.1 Definition

Der Begriff „curriculum-based measurement“ setzt sich aus den Worten „curriculumbasiert“ und „Messung“ zusammen (vgl. Abb. 27, Deno 2003, S. 184f.; Klauer 2006, S. 24).

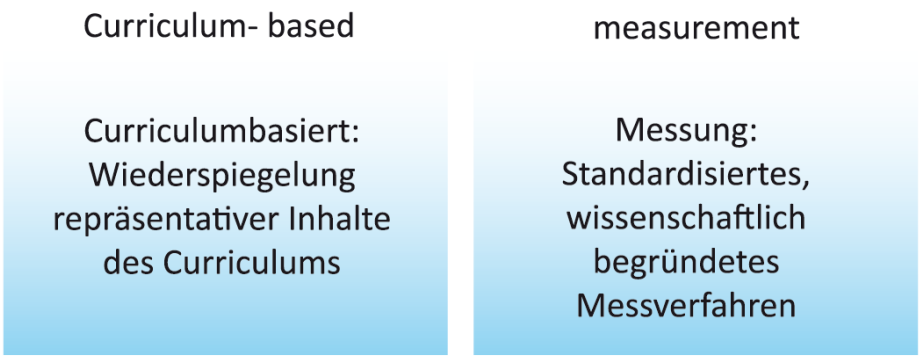


Abb. 27: Darstellung der Begrifflichkeit „curriculum-based measurement“ (CBM)

„Curriculumbasiert“ bedeutet, dass die Messung repräsentative Inhalte des Curriculums beinhaltet. „Messung“ beschreibt, dass es sich um ein wissenschaftlich begründetes

Messverfahren handelt, das im „testpsychologischen Sinn“ (vgl. Klauer 2006, S. 17) angewendet wird und dass es sich nicht um eine bloße Einschätzungsmethode handelt.

> *“Curriculum-Based Measurement (CBM) is a set of standardized and specific measurement procedures that can be used to quantify student performance in the basic academic skill areas of reading, spelling, mathematics computation, and written expression.” (Hintze & Shapiro 1997, S. 351)*

CBM ist ein formatives Messverfahren (vgl. Klauer 2006, S. 17; Graney & Shinn 2005, S. 184ff.; Ardoin, Witt, Suldo, Connell, Koenig et al. 2004, S. 220), das sowohl die Leistungsveränderung als auch den Lernfortschritt eines Kindes über einen bestimmten Zeitraum hinweg dokumentiert. Durch die über einen längeren Zeitraum angestrebte Evaluation können Schülerkompetenzen sensibler erfasst werden. In bestimmten Zeitabständen werden Messungen zu dem gleichen Lernziel durchgeführt, so dass die Leistungsentwicklung in dem gemessenen Zeitraum dokumentiert und transparent gemacht wird. Anhand der Evaluation kann entschieden werden, ob der Lernfortschritt ausreichend ist oder ob eine Modifikation der Methode oder Förderung von Nöten ist (vgl. Klauer 2006, S. 17).
CBM ist als valides Messverfahren somit auch ein Teil einer systematisch formativen Evaluation von Unterricht (vgl. Walter 2008b, S. 63f.), die dazu beitragen soll, die Unterrichtsqualität zu verbessern. CBM ermöglicht, eine schlechte Leistung eines Schülers insbesondere auch auf ein nicht angemessenes Unterrichtsangebot oder eine nicht angemessene Interventionsmaßnahme zurückzuführen. Bei einem formativen Messverfahren wie die CBM werden Lernfortschritte fortlaufend dokumentiert, während bei summativen Messverfahren erst am Ende eines Lernprozesses der gesamte Lernerfolg gemessen wird (vgl. Abb. 28, zur summativen Evaluation s. auch Kap. II.3 - Methodik der Screening-Verfahren).



Abb. 28: Summative im Vergleich zu formativen Messverfahren

Howell und Nolet (2001, S. 112) beschreiben summative und formative Evaluation wie folgt:

- *"Summative evaluation occurs after teaching and learning have taken place. […] Summative evaluation is useful for deciding what to teach.*
- *Formative evaluation is used to illustrate the process of learning. Formative evaluation is useful for making decisions about how to teach."*

CBM werden den „general outcome measures" (GOMs) (vgl. Shapiro et al. 2006, S. 19) oder den „Dynamic Indicators of Basic Skills" (DIBS) (vgl. Shinn & Bamonto 1998, S. 1) übergeordnet (vgl. Abb. 29). DIBS sind Indikatoren des sich verändernden Ausprägungsgrades von Basisfertigkeiten. GOMs dienen als Mittel zur Beobachtung der Lernleistungen und zugleich als Möglichkeit, Instruktionen oder Interventionen zu modifizieren. Es handelt sich um standardisierte Messverfahren, die die Beobachtung des Lernfortschritts eines Kindes über einen längeren Zeitraum ermöglichen. Im Deutschen wird daher zunehmend auch der Begriff „Lernverlaufsdiagnostik" (vgl. Strathmann, Klauer, Greisbach 2010, S. 64) oder „prozessbegleitende Diagnostik" (vgl. Walter 2009b, S. 163ff.) verwendet. Der Einfachheit halber wird in dieser Arbeit weiterhin auf die curriculumbasierte Messung bzw. deren Abkürzung „CBM" zurückgegriffen.

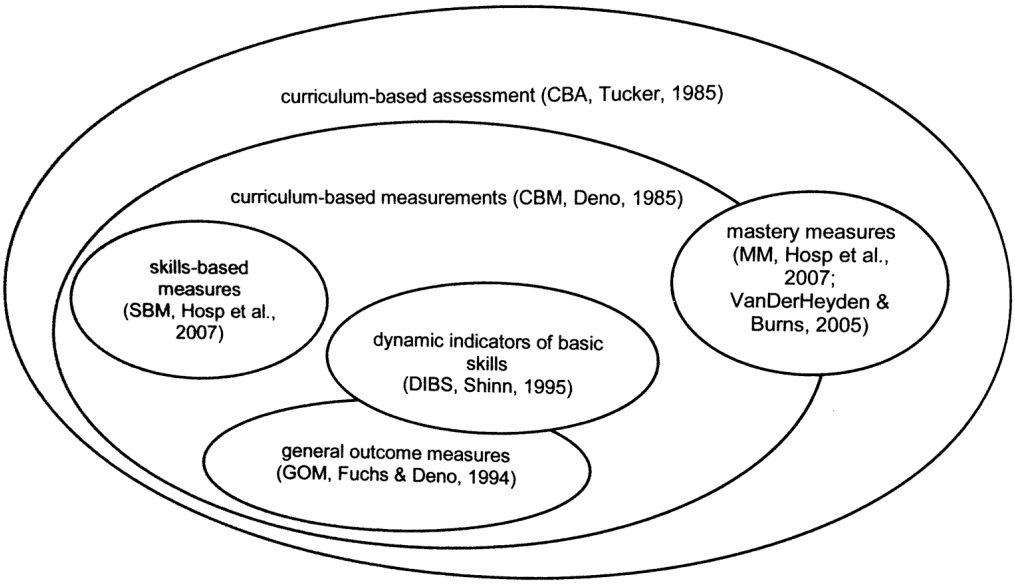


Abb. 29: Schematische Darstellung des Zusammenhangs von CBA, CBM, SBM, GOM, DIBS und MM in Anlehnung an Tucker (1985), Deno (1985), Hosp et al. (2007), Fuchs und Deno (1994), Shinn (1995) sowie VanDerHeyden und Burns (2005), zit. nach Voß & Hartke 2014, S. 86

2.2 Durchführung und Dokumentation

Klauer (2006, S. 17) beschreibt CBM als regelmäßige Erhebung von repräsentativen Aufgabenstichproben zu einem definierten Lernziel. In festgelegten Zeitabständen werden Leistungen anhand von Aufgabenstichproben gemessen. Die Aufgaben müssen immer wieder neu sein, jedoch in allen Belangen die gleiche Leistung messen (vgl. ebd.). Der Lernverlauf der Kinder kann so innerhalb eines bestimmten Zeitverlaufs dokumentiert und auch veranschaulicht werden, wie es auch bei Deno (1985, S. 226, vgl. Abb. 30) zu sehen ist. Deno stellt den Lernverlauf in einem einfach verständlichen Graphen dar. Weitere Beispiele sind in Deno, Lembke und Anderson (2007) zu finden. Die sogenannten „academic thermometers“ (vgl. Shinn & Bamonto 1998, S. 1) lassen eine präzisere Einschätzung des individuellen Entwicklungsstandes des einzelnen Kindes zu.

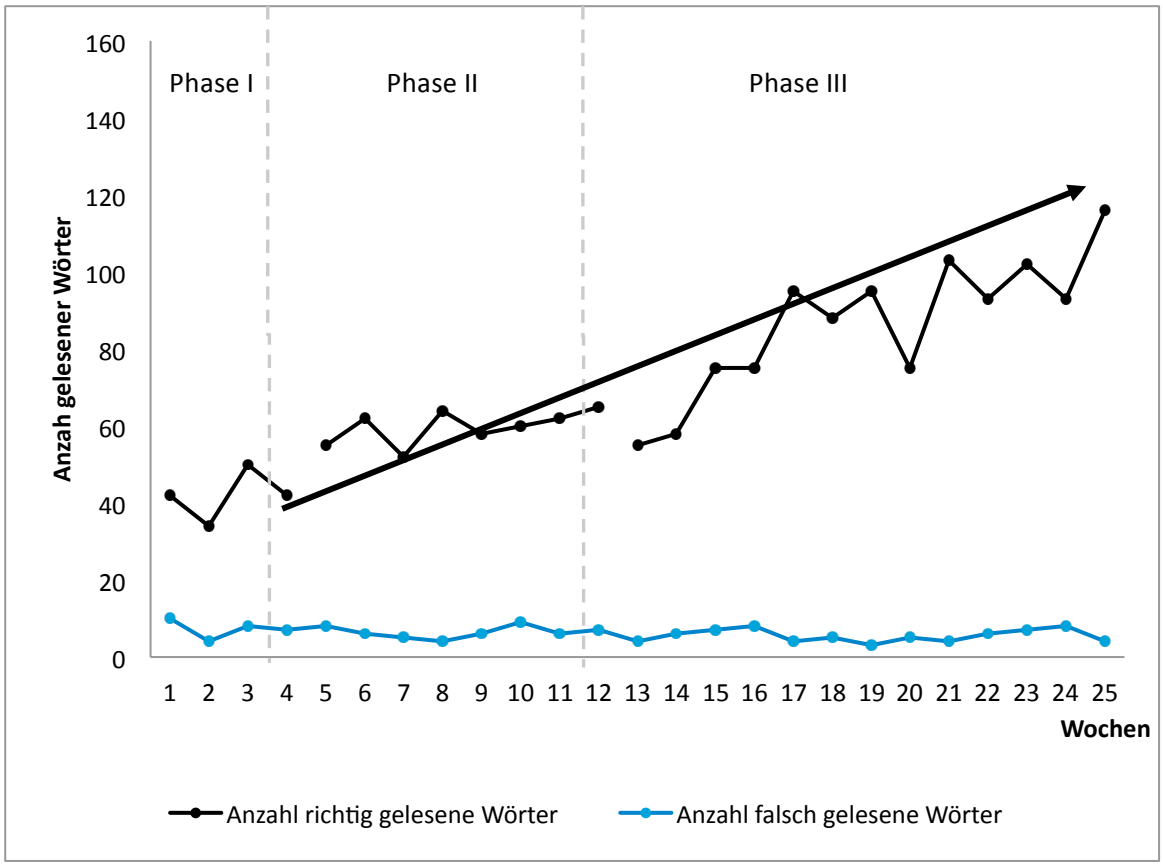


Abb. 30: CBM: Graphische Darstellung der Lernentwicklung (übersetzt aus dem Englischen, verändert nach Deno 2003, S. 186)

Über die hier beispielhaft dargestellte Lernkurve (vgl. Abb. 30) können die Lernfortschritte der Kinder dokumentiert werden und auch Ergebnisse anderer Kinder der gleichen Altersklasse für eine bessere Übersicht eingefügt werden. Dadurch werden auch Vergleiche zu Altersgenossen evident.

Fuchs, Fuchs, Hamlett und Germann (1993, S. 1ff.) konnten in einer Untersuchung mit ca. 3 000 Grundschulkindern der Klassen 1 bis 6 für die Bereiche Mathematik, Lesen und Rechtschreibung die durchschnittlichen Anstiege pro Woche ermitteln, mit denen die Werte der Kinder verglichen werden können. Die ermittelten Werte geben somit einen Hinweis, welche Zuwächse der Kinder in den einzelnen Zeitabschnitten erwartungsgemäß auftreten.

Unbefriedigende Fortschritte können frühzeitig entdeckt und Interventionen auf dieser Basis modifiziert werden (vgl. Klauer 2006, S. 17). Dies ist in Abb. 30 insbesondere auch durch die verschiedenen Phasen sichtbar. Die Lehrkraft konnte aufgrund der sichtbaren nur sehr geringen Fortentwicklung des Schülers ihre didaktische Strategie ändern. Sie verknüpft ihre didaktische Vorgehensweise mit den Ergebnissen zum Lernerfolg aus der CBM (vgl. Klauer 2006, S. 19).

Für ältere Schüler hat sich zudem herausgestellt, dass eine transparente Zielsetzung und Evaluation sich in Bezug auf den Lernprozess als hilfreich erweist (vgl. Fuchs, Fuchs, Karns, Hamlett, Dutka & Katzaroff 2000, S. 83ff.). Schüler in kleineren Klassen dagegen werden durch die Hintergrundinformationen der CBM nicht beeinflusst, sodass dieser psychologische Effekt in den folgenden Ausführungen außer Acht gelassen werden kann.

2.3 Einsatzbereich

Ebenso wie in der Förderdiagnostik sind formative Messverfahren, wie die CBM, nicht für die Klassifikation von Kindern in Untergruppen bzw. für die Selektion von Untergruppen zur Differenzierung geeignet, sondern sie werden eingesetzt, wenn die Leistungsveränderung eines Schülers gemessen werden soll. Die Lernfortschritte des Kindes werden über einen bestimmten Zeitraum dokumentiert, um frühzeitig in den Lernprozess, z.B. durch Änderung der Intervention eingreifen zu können (vgl. Klauer 2006, S. 17).

Ursprünglich wurde die CBM für den sonderpädagogischen Bereich als Messverfahren entwickelt (vgl. Deno 1985, S. 230f.). Fuchs, Fuchs, Hamlett und Germann (1993, S. 13f.) beschreiben, dass der Einsatzbereich der CBM sowohl in sonderpädagogischen Einsatzfeldern, als auch im normalen Unterricht mit allen Schülern eingesetzt werden kann. Durch eine regelmäßige Überprüfung können Kinder identifiziert werden, deren Lernverlauf stark abweichend ist, um daraus resultierend ggf. Interventionen anzupassen.

CBM wird im ursprünglichen Sinne (vgl. Deno 1989, S. 15f.) als Datenbasis für den Problemlöseprozess im sonderpädagogischen Bereich genutzt. Die Messungen dienen der Identifikation des Problems (als Screening), der Problemdefinition (die Differenz zwischen aktueller und erwarteter Leistung) und auch als Entscheidungshilfe und Überprüfung bei der Implementierung eines Interventionsprogramms (vgl. Deno 1989, S. 13).

In der amerikanischen Literatur werden CBM vorrangig beim Lesen (Anzahl richtig gelesener Wörter pro Minute), beim Rechtschreiben (Anzahl richtig geschriebener Buchstabenfolgen in zwei Minuten), beim schriftlichen Ausdruck (Anzahl der richtig geschriebenen Wörter und der richtigen Wortsequenzen) sowie in Mathematik (Anzahl der richtigen Ergebnisse in 2 - 5 Minuten; vgl. Fuchs, Fuchs, Karns, Hamlett & Katzaroff 1999, S. 609ff.) eingesetzt (vgl. Shinn & Bamonto 1998, S. 1ff.).

Deno (2003, S. 185ff.) formuliert außerdem 14 verschiedene Möglichkeiten[18] für den Einsatz von curriculumbasierten Messverfahren. Klauer (2006, S. 23) fasst dazu die wichtigsten Aspekte zum Einsatz von CBM zusammen:

- Verbesserung der Interventionsmaßnahmen für den einzelnen Schüler oder die gesamte Klasse,
- Vorhersage der Leistung in späteren Jahren,
- Entwicklung von Normen, die Schüler einhalten müssten, um ein bestimmtes Lernziel zu erreichen,
- zur Kommunikation mit Eltern und Schülern durch die transparente Darstellungsform,
- zur Früherkennung von Risikokindern mit dem Ziel einer frühzeitigen Förderung

An dieser Stelle sei in erster Linie auf die Vorhersage der Leistung, sowie auf das Screening näher eingegangen. Deno (2003, S. 187) beschreibt die Möglichkeit der Vorhersage von Leistungen durch CBM: "[…] CBM data can be used not only to evaluate instruction but also to classify age and grade developmental status, […]".

Shinn (1989, S. 119ff.) beschreibt ebenfalls einen Einsatz als Screening und zur Entscheidungsfindung für spezielle Interventionen im sonderpädagogischen Bereich, er unterstreicht:

18 Auf alle Ansätze wird hier im Detail nicht eingegangen, für weitere Informationen sei auf Deno (2003, S.185ff.) und Klauer (2006, S. 23) verwiesen.

> *"CBM can provide „simple", reliable, and understandable procedures for screening and eligibility determination but cannot provide simple answers to these complex questions." (Shinn 1989, S. 126)*

US-amerikanische Untersuchungen konnten zeigen, dass die CBM sich als Screening eignet (vgl. VanDerHeyden, Witt, Naquin & Noell 2001, S. 363ff.). Die Autoren um VanDerHeyden (vgl. ebd.) zeigten auf, dass die Identifikation der Schulreife der Kinder im Kindergarten ebenfalls über CBM angestrebt werden kann. Als Screening-Verfahren eröffnet die CBM insbesondere im Vorschulbereich die Möglichkeit frühzeitig Kinder zu erkennen, die Gefahr laufen in einer schulischen Fähigkeit Schwierigkeiten zu entwickeln. Auch Wright (o. J. S. 1) vermerkte ebenfalls, dass CBM sehr gut geeignet ist, um vorschulische Fähigkeiten zu beobachten. Durch die Standardisierung einer CBM ist es damit möglich, die Werte der Kinder mit Normen oder auch mit der Gruppe gleichaltriger Kinder zu vergleichen. Deno, Reschly-Anderson, Lembke, Zorka und Callender (2002, zit. nach Deno 2003, S. 184ff.) konnten nachweisen, dass bei einer CBM-Messung bei Grundschülern die Kinder, die bei der CBM zu den untersten 20 % gehörten, besonders risikogefährdet waren. Die phonologische Bewusstheit zur Prädiktion späterer Lesefähigkeiten spielt laut der Studie von Kaminski und Good (1996, S. 215ff.) dabei eine entscheidende Rolle.

Auch Hosp, Hosp und Howell (2006, S. 26) beschreiben, dass die CBM für ein Screening gut geeignet erscheint. Dabei kann CBM zwei verschiedene Indikatoren eines Problems aufdecken:

(a) niedrige Leistung (low performance) in bestimmten Fähigkeiten
(b) langsames Vorwärtskommen im Lernprozess (low progress) beim Erwerb von Fähigkeiten

Eine Reihe von Autoren bemerken, dass beim Einsatz von sensitiven Messwerkzeugen ein Screening bzw. Benchmarking durch CBM über einen bestimmten Zeitraum hinweg Vorteile für die Förderung verschafft (vgl. Hosp, Hosp & Howell 2006, S. 26f.; Ardoin & Christ 2008, S. 109f.; VanDerHeyden, Witt, Naquin & Noell 2003, S. 204ff.). Hierdurch können insbesondere die Kinder identifiziert werden, die zwar eine durchschnittliche Leistung in den gemessenen Fähigkeiten aufzeigen, sich jedoch nicht weiterentwickeln und die zusätzliche Hilfestellungen oder eine andere Form der Instruktion benötigen.

2.4 Aufgabenstichproben

Bei der Entwicklung von Aufgaben wird häufig ein sogenannter „robuster Indikator" gesucht, der das Gesamtmaß von bestimmten Fähigkeiten und Fertigkeiten des Schülers aus dem jeweiligen Curriculum darstellt, wie beispielsweise das flüssige Lesen von Textpassagen eine Aussage über die allgemeine Lesekompetenz zulässt (vgl. Walter 2008b, S. 65f.). Im Bereich Mathematik konnte bisher kein einzelner allgemeiner Indikator gefunden werden (vgl. zusammenfassend Foegen, Jiban & Deno 2007, S. 121ff.), sodass hierfür systematisch Aufgaben in Gruppen organisiert werden.

Fuchs (2004, S. 191; sowie in Klauer 2006, S. 17, Klauer 2014a, S. 5f.) definiert deshalb zwei Wege zur Entwicklung von Aufgabentypen. Einzelne, eng begrenzte Lernziele, bei denen es sich in der Regel um Fertigkeiten handelt, werden durch den „ersten Weg" bestimmt, umfassendere Lernziele umfassen meistens mehrere Fertigkeiten auf einmal und werden mit dem „zweiten Weg" erstellt (vgl. Klauer 2006, S. 24).

Erster Weg:
Der erste Weg, den Fuchs (2004, S. 191) vorschlägt, eignet sich besonders zur Messung von Leseleistungen. Es werden für die jeweilige Leistung repräsentative Aufgaben gesucht, die möglichst bereits die Teiloperationen umfassen, die bzgl. des Lernziels erfasst werden sollen. Es handelt sich damit um eine multidimensionale Messung. Die Aufgaben sind geeignet, wenn die Stichproben robust mit den geforderten Leistungen korrelieren. So ist beispielsweise die Leseflüssigkeit beim lauten Vorlesen nach Deno, Mirkin und Chiang (1982, S. 36ff.) ein guter Indikator für die allgemeine Leseleistung des Kindes. Die Leseleistung kann in der CBM beispielsweise durch das laute Vorlesen eines Textes (vgl. Abb. 31) oder auch durch das Ausfüllen eines Lückentextes gemessen werden. Im Bereich der Rechtschreibleistung dagegen kann der Lernerfolg durch Wortdiktate erfasst werden.

Text	Anzahl der Wörter
One hundred years ago in Paris, when theaters and music halls	11
drew travelling players from all over the world, the best place to	23
stay was at the widow Gateau's, a boardinghouse on English	33
Street. Acrobats, jugglers, actors, and mimes from as far away […]	43

Abb. 31: Beispiel für eine CBM für die Messung von Leseflüssigkeit (Ausschnitt aus Wright o. J., S. 6)

Zweiter Weg:

Der zweite Weg ist besonders für Messungen im mathematischen Bereich tauglich (vgl. Abb. 32). Es wird zunächst die geforderte Leistung bestimmt, die zum Ende des Zeitraums von den Kindern gezeigt werden soll. Im Anschluss wird durch einen Plan festgelegt, welche Teilmengen der Aufgaben diese Leistung repräsentieren. Für diese sogenannten „single-skill-Messungen“ (vgl. Fuchs 2004, S. 191) wird dann noch die jeweilige Anzahl an Aufgaben bestimmt, um lernzielvalide Stichproben zu erhalten (vgl. Klauer 2006, S. 18). Anhand der Dokumentation der einzelnen Teilleistungen kann eine Intervention gezielter geplant werden, da festgestellt werden kann, in welchen Bereichen das Kind noch Schwierigkeiten hat, z. B. in den vier Grundrechenarten im Fach Mathematik (vgl. a. a. O., S. 21f.).

Addition				
1 +3	9 +3	3 +8	1 +6	1 +8
2 +1	4 +7	1 +8	5 +2	6 +8
8 +8	2 +7	2 +6	3 +4	3 +3
5 +2	1 +1	9 +1	8 +7	8 +1
1 +8	8 +2	6 +5	1 +5	2 +3

Abb. 32: Beispiel für eine CBM für die Messung mathematischer Leistungen (angelehnt an Wright o. J., S. 13)

Ein bedeutender Unterschied zwischen den beiden Wegen ist, dass bei der Messung von komplexeren Fähigkeiten über den zweiten Weg auch Aufgabentypen in der CBM vorkommen, die die Kinder zu Beginn der Messungen voraussichtlich noch nicht bearbeiten können. Es kann damit jedoch der Lernverlauf auf einer Skala über einen längeren Zeitraum abgebildet werden (vgl. a. a. O., S. 24).

Im nächsten Abschnitt werden aktuelle Studien der letzten Jahre zu curriculumbasierten Messungen ausführlich dargestellt. Sie begründen die Auswahl dieser Methodik für das curriculumbasierte Messverfahren Phonolino. Curriculumbasierte Messungen erfuhren zunächst im us-amerikanischen Bereich Beachtung und wurden dort umfassend

wissenschaftlich fundiert. Aus diesem Grund werden zunächst us-amerikanische Studien in Bezug auf ihre Güte und der Vor- und Nachteile curriculumbasierter Verfahren erläutert. Im Anschluss daran werden erste deutsche Verfahren vorgestellt. Zuletzt sollen die curriculumbasierten Verfahren im Vorschulalter dargelegt werden, da die in dieser Studie vorgestellte CBM Phonolino zur Prognose von Leistungen im Schriftspracherwerb ebenfalls als Verfahren zum Einsatz im letzten Kindergartenjahr gedacht ist.

3 Aktuelle Studien

Die Forschung zur CBM ist in Deutschland noch nicht so weit vorangeschritten wie in den USA. Lynn S. Fuchs (2004, S. 189) beschreibt drei wichtige Stadien, die bei der Forschung Berücksichtigung finden sollten:

1) Technical features of the static score
2) Technical features of slope
3) Instructional utility

Sie beschreibt, dass die Möglichkeit bestehe, dass die statistischen Daten positiv ausfallen (1 und 2), das Messverfahren in der Praxis jedoch nicht nutzbar ist (3). Der Fokus bisheriger Untersuchungen lag auf der ersten Stufe, in dem statistische Daten bzgl. der Validität und Reliabilität der Verfahren erhoben worden sind. Fuchs (2004, S. 191) betont jedoch, dass für zukünftige Messungen ebenfalls die forschungsaufwändigeren Stufen 2 und 3 betrachtet werden müssen.

Im Folgenden werden aktuelle Studien zur CBM aus den USA (Kap. III.3.1) sowie aus Deutschland (Kap. III.3.2) vorgestellt, um darzulegen, dass die CBM in den USA sowie ansatzweise auch schon in Deutschland als valides Messverfahren angesehen werden kann. Es werden auch Studien aus dem Vorschulbereich (Kap. III.3.3) insbesondere um die Forschergruppe von Amanda VanDerHeyden aufgezeigt. Die Forschung im Bereich der CBM in Bezug auf die Vorläuferfertigkeiten des Schriftspracherwerbs ist bisher noch nicht sehr ausgeprägt, so wird in der vorliegenden Studie exemplarisch auf einige Forschungsergebnisse hingewiesen, die für die Fragestellungen dieser Untersuchung und zur Erklärung der Konzeption der in dieser Studie evaluierten CBM Phonolino von Relevanz sind.

3.1 Darstellung der Untersuchungsergebnisse aus dem US-amerikanischen Bereich

Literatur und Studien zur curriculumbasierten Messung stammen vorrangig aus dem US-amerikanischen Raum. Dabei wurde der Bereich CBM und Lesen am umfassendsten erforscht. Im Überblick über die Studien konnte festgestellt werden, dass diese die klassischen Testgütekriterien Objektivität, Reliabilität und Validität von CBM bestätigen konnten (vgl. Fuchs, Deno & Mirkin 1984; Deno 1985, 2003; Fuchs, Fuchs, Hamlett & Germann 1993; Madelaine & Wheldall 1999; Fuchs; VanDerHeyden, Witt, Naquin & Noell 2001; Ardoin et al. 2004, zusammenfassend Waymann et al. 2007).

Die Objektivität wird meistens durch die einfache Entscheidung „richtig-falsch“ erbracht und ist dadurch bei der CBM oftmals hoch (vgl. z.B. VanDerHeyden, Witt, Naquin & Noell 2002, S. 363ff.). Die CBM ist überwiegend durch eine hohe Reliabilität gekennzeichnet, da durch die häufigen Wiederholungen der Messungen die Retest-Reliabilitäten berechnet werden können (vgl. z.B. Deno, Mirkin & Chiang 1982, S. 36ff.). Bezüglich der Validität konnten besonders hohe Korrelationen mit Schulleistungstests festgestellt werden (vgl. Klauer 2006, S. 22; Marston 1989, S. 34ff.; Deno, Reschly-Anderson, Lembke, Zorka & Callender 2002, S. 1ff.; Good, Simmons & Kame’enui 2001, S. 257; VanDerHeyden, Witt, Naquin & Noell 2002, S. 363). Für eine zusammenfassende Übersicht der Studien zur Validität und Reliabilität im Lesen, Schreiben, Buchstabieren und Mathematik sei auf die Zusammenschau von Hosp, Hosp und Howell (2006, S. 141f.) verwiesen.

Beispielhaft werden an dieser Stelle die Ergebnisse zum Lesen (vgl. Fuchs & Fuchs 2002, S. 67f.) zusammengefasst, da dieser Bereich im US-amerikanischen Umfeld am besten erforscht ist.
Beim Lesen konnte sowohl für die CBM des lauten Lesens über eine Minute, als auch für die sogenannten MAZE-Messung in 2,5 Minuten eine hohe Kriteriumsvalidität zu herkömmlichen Lesetests (vgl. Fuchs & Fuchs 1992, Marston 1989), sowie zu Lehrerurteilen betreffs der Lesekompetenz (vgl. Fuchs & Fuchs 1992, Fuchs, Fuchs & Deno 1982) und zu informellen Lesemessungen (vgl. Fuchs & Fuchs 1992, Fuchs, Fuchs & Maxwell 1988) gefunden werden. Mehrens und Clarizio (1993, S. 248) merken jedoch hinsichtlich der Kriteriumsvalidität an, dass in den Studien, in denen die Experimentalgruppen mit Einsatz von CBM besser abgeschnitten haben, als die Kontrollgruppen ohne CBM, die Lehrer die Zielwörter der CBM explizit geübt haben könnten, sodass die Ergebnisse dadurch verfälscht wurden.

Auch die prädiktive Validität (vgl. Fuchs, Fuchs, Thompson, Al-Otaiba, Yen, Yang, Braun & O'Connor 2001) und die Konstruktvalidität (vgl. Shinn, Good, Knutson, Tilly & Collins 1992), sowie die diskriminative Validität (vgl. Deno, Mirkin & Chiang 1982, Shinn, Tindal, Spira & Marston 1987, Deno 1985, Fuchs & Deno 1992, Fuchs, Fuchs, Hamlett, Walz & Germann 1993) konnten mit angemessenen Werten nachgewiesen werden. Hosp und Fuchs (2001 in Fuchs & Fuchs 2002, S. 67) konnten zwischen CBM und Wortidentifikation eine Kriterienvalidität zwischen r = .79 und r = .89, zwischen CBM und Wortgewandtheit zwischen r = .70 und r = .82, sowie zwischen CBM und Textverständnis Korrelationen zwischen r = .76 und r = .83 feststellen. Die Retest-Reliabilität lag zwischen r = .93 und r = .97. Die prädiktive Validität zu dem herkömmlichen Testverfahren TerraNova (vgl. Fuchs & Hamlett 1997 in Fuchs & Fuchs 2002, S. 67f.) lag in einer anderen Studie nach 12 Wochen bei r = .86, nach 30 Wochen bei Erstklässlern bei r = .79 und bei Zweitklässlern bei r = .72. Fuchs und Fuchs (2002, S. 81) beschreiben den Erfolg der zahlreichen Studien zu CBM wie folgt: "At present, CBM appears useful for describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders."

Fuchs, Deno und Mirkin (1984, S. 456ff.) konnten außerdem zeigen, dass die Implementierung von CBM in den Unterricht im Zusammenhang mit dem Aufstellen von Langzeitzielen im Rahmen des Lesens sich positiv auf den Lernfortschritt der Schüler auswirkt und auch die Entscheidungen der Lehrer in Bezug auf die Lernziele und etwaigen Interventionsmaßnahmen beeinflusst. In der Studie wurden im experimentellen Versuchsaufbau in einer Experimental- und einer Kontrollgruppe 39 Sonderschullehrer und 141 Schüler untersucht. Dabei wurden eine CBM im Lesen, sowie zwei standardisierte herkömmliche Lesetests, ein Fragebogen für die Lehrer und ein Schülerinterview eingesetzt. Fuchs et al. (1984, S. 458ff.) konnten zeigen, dass die Lehrer aus der Experimentalgruppe (Einsatz von CBM) systematisch die Struktur ihres Unterrichts anhand der Daten verbesserten, während die Lehrer der Kontrollgruppe weniger systematisch vorgingen.

Stecker und Fuchs (2000, S. 134ff.) konnten diese Ergebnisse auch für den Bereich Mathematik bestätigen. Der Einsatz von CBM im Unterricht war effektiver als die Nutzung von anderen Möglichkeiten der Überprüfung von Lernfortschritten.

Nach den umfassenden Untersuchungen zu curriculumbasierten Messverfahren in den USA werden nun im folgenden Abschnitt Untersuchungsergebnisse aus Deutschland dargelegt.

3.2 Darstellung der Untersuchungsergebnisse aus dem deutschsprachigen Bereich

Im deutschsprachigen Bereich ist eine Form der wiederholten Leistungsmessung bereits langläufig durch Klauer (1978) bekannt geworden. Auch der Unterschied zwischen formativer und summativer Evaluation wurde nach dem Artikel von Bloom, Hastings und Madaus (1971) in der Pädagogischen Psychologie in Deutschland aufgegriffen und diskutiert (vgl. Klauer 2014a, S. 7). Als erste Publikation zur CBM veröffentlichte Klauer (2006) den Artikel „Erfassung des Lernfortschritts durch curriculumbasierte Messung". Die Form der curriculumbasierten Messung ist in den letzten Jahren in Deutschland populär geworden und wurde durch unterschiedliche Forschungsarbeiten (vgl. Diehl & Hartke, 2007, 2012; Huber & Grosche 2012; Klauer, 2006, 2011; Knopp & Hartke, 2010; Müller & Hartmann, 2014; Souvignier, Förster & Salaschek, 2014; Strathmann & Klauer, 2008, 2010; Strathmann, Klauer & Greisbach, 2010; Voß, Blumenthal, Sikora, Mahlau, Diehl & Hartke 2014; Voß, Mahlau, Sikora, Blumenthal, Diehl & Hartke 2015; Walter, 2008a, 2008b, 2009a, 2009b, 2010a, 2010b, 2011a, 2011b, 2014a, 2014b) vorangetrieben. Auch in der Erwachsenenbildung im Rahmen der Methodenausbildung von Sonderpädagoginnen und Sonderpädagogen wird das Response-to-Intervention-Modell über kurze CBM-Messungen im Rahmen des Studiums eingesetzt (vgl. Rietz, Koll & Franke 2013).

Walter (2008b, 2009a, 2009b, 2010) konnte in umfangreichen Untersuchungen im Bereich des sinnverstehenden Lesens zeigen, dass die CBM den wichtigsten Gütekriterien entspricht. Mit dem curriculumbasierten Verfahren LDL (Lernfortschrittsdiagnostik Lesen) entwickelte er als erster im deutschsprachigen Raum ein normiertes Verfahren, mit dem die Lesekompetenz von Kindern für eine Prozessdiagnostik innerhalb von kurzen Zeitabständen erfasst werden kann. Das Kompetenzmaß ist hierbei die Anzahl der richtig gelesenen Wörter pro Minute. Der Test wurde mit fast 700 Grundschülern aller Klassenstufen, 570 Hauptschülern und 380 Förderschülern normiert. Dabei wurden hohe Reliabilitäten evaluiert: Paralleltest-Reliabilität r =.90 (Grundschule) und r = .80 (Hauptschule).

Gleichermaßen konnte das CBM-MAZE-Verfahren (Walter 2009a, S. 72ff.) den klassischen Testgütekriterien genügen. Hierbei müssen die Schüler in einem Text jedes siebte Wort (bis auf den ersten Satz) per Multiple Choice mit drei angebotenen Lösungsvarianten ergänzen. Die Anzahl der richtig gefüllten Lücken wird als Indikator für die Lesekompetenz gesehen. Dieses Verfahren ist äußerst ökonomisch einsetzbar, da die Messungen in der Gruppe oder elektronisch durchgeführt werden können. Walter konnte diesbezüglich Paralleltest-Reliabilitäten zwischen .52 und .75 ausmachen. Er stellte außerdem eine zufriedenstellende

Validität sowie Änderungssensibilität des Verfahrens fest. Strathmann, Klauer und Greisbach (2010, S. 69ff.) konnten dagegen keine derartig guten Werte für die Reliabilität ausmachen. Sie überprüften das Verfahren der Lernverlaufsdiagnostik im Bereich der Rechtschreibung mit der Verwendung von Aufgabenstichproben, die durch einen Zufallsgenerator zusammengestellt worden waren. Sie erzielten jedoch nur in zwei der sechs teilnehmenden Klassen mittlere Reliabilitäten. Die Autoren stellten fest, dass unmittelbar aufeinanderfolgende Messungen bessere Paralleltest-Reliabilitäten aufwiesen. Die Variation der Leistungen scheint über die Zeit hinweg stärker zu sein als ursprünglich vermutet. Die Validität des Verfahrens konnte bestätigt werden.

Strathmann und Klauer (2010, 2012) entwickelten die Lernverlaufsdiagnostik LVD-M 2-4 für den Bereich Mathematik für zweite bis vierte Klassen. Die Besonderheit des Verfahrens liegt in der Erstellung der Aufgabenblätter, die mittels einer CD in Form eines Arbeitsblattes jeweils neu generiert werden. Jedes Kind bekommt eine individuelle Zufallsstichprobe von 24 Aufgaben. Durch das PC-Programm können individuelle Lernverläufe der Schüler auch grafisch dargestellt werden. Die LVD-M 2-4 wurde normiert und weist individuelle, soziale und Lehrzielnormen aus (vgl. Strathmann 2014, S. 204 sowie Strathmann & Klauer 2010, S. 114f.). Die Paralleltestreliabilität wies Reliabilitätskoeffizienten zwischen r = .66 und r = .91 und die Split-Half-Reliabilität zwischen r = .56 und r =.88 auf. In gleicher Weise konnten die Validität sowie die Testschwierigkeiten positiv bewertet werden. Die generierten Tests waren insofern in der Lage Lernfortschritte zu erfassen (vgl. Strathmann & Klauer 2010, S. 120). Die Autoren beschreiben ebenfalls die Möglichkeit der Übertragbarkeit der Generierung von Aufgabenstichproben über einen Zufallsgenerator auf andere Fächer (Vokabelwissen in Englisch, Lehrtexte in Sachfächern), verweisen aber auf die Notwendigkeit einer gründlichen Testung der jeweiligen Stichproben.

Voß (2013. S. 517ff.) entwickelte im Rahmen des Rügener Inklusionsmodells, einem Forschungsprojekt der Universität Rostock, curriculumbasierte Messungen für den Mathematikunterricht der Grundschule und konnte zufriedenstellende Befunde in Bezug auf die Gütekriterien Reliabilität, Objektivität, Validität sowie Änderungssensibilität des Verfahrens festhalten. Müller und Hartmann (2014) erstellten die „Lernfortschrittsdiagnose: Addition im ZR 21-99“ für die Grundschule, die als 3-Minuten-Tests konzipiert sind. An der Universität Münster entstand eine Zusammenstellung von curriculumbasierten Messverfahren für den Bereich Lesen (Klasse 3-6), Mathematik (Klasse 1-3) und Englisch (Klasse 5). Diese Verfahren lassen sich internetbasiert auf der Plattform „Quop“ (www.quop.de) auswerten, was für die Praktikabilität und die schnelle Anwendung der Messungen einen enormen Vorteil bietet. Die Qualität der Testreihen wurde in bisher schon

sechs Studien mit 2400 Schülerinnen und Schülern untersucht. Die Messverfahren stellten sich als reliabel und valide sowie veränderungssensitiv heraus. Die Messungen sind dazu in der Lage Leistungen über mehr als ein Schuljahr hinweg vorherzusagen (vgl. Souvignier, Förster & Salaschek 2014, S. 246f.). Außerdem konnte in drei Studien konsistent festgestellt werden, dass die Klassen, dessen Lehrkräfte kontinuierlich diagnostische Informationen über den Lernverlauf ihrer Schülerinnen und Schüler erhalten haben, signifikant höhere Lernzuwächse der Schüler erzielten, als die Vergleichsklassen, die keine Lernverlaufsdiagnostik bekamen (vgl. Souvignier, Förster & Schulte 2014, S. 225ff.).

Kuhlmann (2013, S. 148ff.) konnte bei der Erstuntersuchung der curriculumbasierten Messung FE-RS 2 erste zufriedenstellende Ergebnisse für den Bereich der Rechtschreibung in der zweiten Klassenstufe erlangen. Das Verfahren erwies sich als valide, reliabel und sensitiv. Es kann Schülerleistungen über das gesamte Schuljahr analysieren und differenziert abbilden.

In 2013 wurde von Walter (2011a, S. 106ff.) auch das Verfahren VSL (Verlaufsdiagnostik sinnerfassenden Lesens) veröffentlicht, das auch als ökonomisch einsetzbare PC-Variante (VSL-C) genutzt werden kann. Die CBM wurde mit $N = 3036$ bzw. $N = 2289$ (für die Veränderungsnormen) Schülern evaluiert. Die Normierungen sind für die zweite bis vierte Klasse der Grundschule, sowie für die fünfte und sechste Klasse der Gemeinschaftsschule, Realschule und des Gymnasiums verfügbar. Die Untersuchungen konnten aufzeigen, dass das Verfahren bzgl. der Konstruktvalidität die notwendige Änderungssensibilität mit sich bringt und damit einen essentiellen Beitrag für die formative Messung von Lesekompetenz darstellt (vgl. a. a. O., S. 124).

Diehl und Hartke (2011, 2012) veröffentlichten eine CBM zur Abbildung des Lernfortschritts in der Lesekompetenz für das erste Schuljahr (IEL-1). In dieser CBM lesen die Schüler Wortlisten mit 60 Wörtern vor (Screening mit drei Messzeitpunkten). Wenn ein Kind als Risikokind identifiziert wird, wird danach der Gesamttest mit Aufgaben zur phonologischen Bewusstheit, Graphem-Phonem-Korrespondenz, zum Silbensegmentieren, Wortlesen, Satzlesen und Textlesen durchgeführt. Die Reliabilitätswerte der Retest-Reliabilität lagen für die Messzeitpunkte für den Gesamttest zwischen $.74 \leq r \leq .83$ und für das Screening zwischen $.77 \leq r \leq .86$. Die kriteriumsorientierte Validität liegt für den Gesamttest je nach Kriterium zwischen $.62 \leq r \leq .69$ und für das Screening $.66 \leq r \leq .70$. Kennwerte der prognostischen Validität (vgl. Diehl 2014, S. 161) zeigten, dass die IEL zum Messzeitpunkt

1 mit der WLLP[19] mit r = .58 hochsignifikant korreliert. Dies bestätigte sich auch bei den Korrelationen zum Lehrerurteil (Lesefertigkeit: r = .56 und Leseverständnis: r = .62). In einer Folgeuntersuchung von Blumenthal (2012, zit. in Diehl 2014, S. 161) konnte ein RATZ-Index vom ersten Messzeitpunkt ausgehend von 60 % berechnet werden.

Bäuerlein, Niklas und Schneider (2014, S. 130) konnten ein PC-gestütztes Diagnoseinstrument, das in England bereits weitreichend validiert worden ist, für Deutschland adaptieren. Die Fähigkeitsindikatoren Primarschule (FIPS) basieren auf dem mehrfach evaluierten Messverfahren PIPS (Performance Indicators in Primary Schools), das 1991 in England entworfen worden ist. FIPS erfasst zu Beginn und am Ende des ersten Schuljahrs Leistungen in den Bereichen Wortschatz, Lautbewusstsein, Lesen und Mathematik und wird damit von Lehrkräften zur Dokumentation des Leistungsstandes im Verlauf der ersten Klasse genutzt. Der Leistungsstand der Schüler kann zu den zwei Messzeitpunkten anhand von Diagrammen dargestellt werden, auf dessen Grundlage der Lernzuwachs des einzelnen Schülers beurteilt werden kann. Außerdem ist es für die Lehrkräfte möglich, eine Boxplot-Verteilung für die gesamte Klasse anzeigen zu lassen. Die Messungen werden am Computer durchgeführt, wobei hierbei „adaptiv" getestet wird, das bedeutet, dass der Computer die Items je nach Fähigkeiten des Kindes auswählt. Sobald ein Abbruchkriterium erreicht wird, wird zu der nächsten Aufgabe übergegangen (vgl. a. a. O., S. 133). Dies beschreibt einen bedeutenden Vorteil zu einer „paper-and-pencil"-Diagnostik. Aufgrund der Adaptivität des Verfahrens dauert die Durchführung zwischen 15 und 30 Minuten. Die Autoren konnten für die Reliabilität des Verfahrens gute Ergebnisse jeweils zu Beginn und zum Ende des Schuljahrs erzielen (Lautbewusstheit: .76 und .50, Mathematik: .89 und .88, Lesen: .86 und .79). Für die Konstruktvalidität konnten ebenfalls zufriedenstellende Korrelationen (vgl. a. a. O., S. 135) zu den jeweiligen Außenkriterien festgestellt werden (Mathematik mit dem DEMAT: r = .62, Lesen mit der WLLP: r = .62 und dem DERET: r = -.71 sowie Lautbewusstheit mit dem DERET: -.40). Aufgrund der nur zweimaligen Messung unterscheidet sich das Programm FIPS zur formativen Leistungsdiagnostik im Vergleich zu anderen curriculumbasierten Messverfahren sehr. Es handelt sich eher um einen Prä-Post-Vergleich der Leistungen nach einem Schuljahr und kann nur im weitesten Sinne als formative Evaluation angesehen werden.

Im Raum Hamburg konnte das Landesinstitut für Lehrerbildung unter der Leitung von May sowie in Zusammenarbeit mit der Universität Hamburg und der Universität Erfurt die Testserie KEKS entwickelt und evaluiert werden, die die Kompetenzen der Kinder vom

19 WLLP = Würzburger Leise Leseprobe (Küspert & Schneider 1998)

letzten Kindergartenjahr (ab 4 bzw. 6 Jahre) bis ins jugendliche Alter (12 bzw. 16 Jahre) erfasst und die Entwicklung über zwei Messungen pro Schuljahr darstellen kann. Gemessen werden die Kernkompetenzen Deutsch, Mathematik, Englisch und Herkunftssprachen (vgl. May, Bennöhr & Berger 2014, S. 257ff.). Die Kompetenz „Deutsch“ umfasst dabei auch explizit das Hörverstehen und die phonologische Bewusstheit als Basiskompetenzen sowie sprachliche Mittel wie Wortschatz und Grammatik. Die Testserie kann als punktuelle aber auch als halbjährlicher Test eingesetzt werden. Die Lehrkräfte können die in Einzel – oder Gruppentests durchgeführten Ergebnisse manuell oder online auswerten und können die Ergebnisse mit Normen zur besseren Interpretation abgleichen. Hierbei werden die phonologische Bewusstheit im weiteren Sinne hauptsächlich in der Test-Serie „Mini-KEKS“ und die phonologische Bewusstheit im engeren Sinne größtenteils in der Test-Serie „KEKS 1A“ überprüft (vgl. May & Bennöhr 2013, S. 41ff.).

Die Testentwicklung beruft sich auf bundesweite Daten von bis zu 8,9 Mio. Schülern, die in unterschiedlichen Teilstichproben zu den jeweiligen Fächern und Klassenstufen ausgewertet wurden. Die Autoren konnten für die KEKS-Tests gute Werte für die Reliabilität (Gesamttestwerte mind. r = .90) aufweisen (vgl. May, Bennöhr & Berger 2014, S. 268). Für den Teiltest zur phonologischen Bewusstheit ergab sich ein Reliabilitätskennwert, gemessen durch Cronbachs Alpha, für den Mini-KEKS von .84 und für KEKS-1A von .92 (vgl. May & Bennöhr 2013, S. 122). Die Validität wurde ausschließlich anhand der Übereinstimmung mit dem Lehrerurteil in Bezug auf die Aufgabenschwierigkeit und in Bezug auf die Kompetenzeinschätzung der Schüler überprüft. Hierfür konnten hochsignifikante Korrelationen zwischen den Ergebnissen des KEKS-Tests und der Kompetenzeinschätzung der Lehrkräfte erreicht werden (vgl. May, Bennöhr & Berger 2014, S. 270f. sowie May & Bennöhr 2013, S. 132). Daten für die prognostische Validität, für die Überprüfung der kriteriumsorientierten Validität oder zur genaueren Überprüfung der Änderungssensibilität des Verfahrens liegen nicht vor. KEKS umschreibt damit keine Lernverlaufsdiagnostik im engeren Sinne, da die Messzeitpunkte weit auseinander liegen. Es kann einen Leistungsverlauf der Kinder über die Jahrgangsstufen und Fächer hinweg konsequent abbilden, ist aber nicht als Veränderungsmaß wie die klassischen curriculumbasierten Messungen zu interpretieren, wird hier jedoch als Verfahren erwähnt, dass der CBM nahe ist.

Nachfolgend werden Studien zu curriculumbasierten Messungen für das Vorschulalter vorgestellt, deren Erkenntnisse eine Grundlage für die in dieser Untersuchung evaluierte CBM Phonolino bieten.

3.3 Curriculumbasierte Messung im Vorschulalter

Die curriculumbasierte Messung wird in den USA auch im vorschulischen Bereich eingesetzt. Durch die unterschiedlichen Curricula des amerikanischen bzw. des deutschen Kindergartens wird in den USA häufig in Verbindung mit der CBM bereits das Buchstabenwissen der Kinder überprüft. In Deutschland dagegen wird nicht vorausgesetzt, dass Vorschulkinder schon Buchstabenwissen mit in die Schule bringen.

VanDerHeyden, Witt, Naquin und Noell (2001, S. 363ff.) evaluierten CBM auch bezüglich der Messung unterschiedlicher Vorläuferfertigkeiten („readiness skills“) im Kindergarten: erste mathematische Kenntnisse, Vorläuferfertigkeiten des Lesens und Schreibens, sowie die allgemeinen Fähigkeiten zur Diskrimination wurden erprobt. Ziel der Studie war es, die Kinder zu identifizieren, die in den Fähigkeiten Schwierigkeiten aufwiesen. Die CBM wurde damit als Screening-Verfahren eingesetzt, um die Kinder zu identifizieren, bei denen eine genauere Diagnostik angemessen und eine frühzeitige Förderung eingeleitet werden sollte. Bei den Vorläuferfertigkeiten im Lesen mussten die Kinder im Vorschulalter (N = 107) Anlaute von Wörtern erkennen und den richtigen Buchstaben einkreisen. Gemessen wurde die richtige Anzahl an Buchstaben in 2 Minuten und 10 Sekunden. Bei den Vorläuferfertigkeiten zum Schreiben mussten die Kinder Buchstaben nachschreiben. Hier wurden die richtigen Replikationen pro Minute gemessen. Das geforderte Vorwissen im Lesen und Schreiben beruhte damit mehr auf dem Buchstabenwissen sowie auf motorischen und Diskriminationsfertigkeiten als auf Komponenten der phonologischen Bewusstheit wie es in deutschen Studien üblicherweise der Fall ist.

Schon Shinn (1989, in VanDerHeyden, Witt, Naquin & Noell 2001, S. 375) konnte feststellen, dass sich die CBM sehr gut als Screening in den ersten Schuljahren eignet. Die Autoren um VanDerHeyden bestätigten gleichwohl, dass die curriculumbasierte Messung besonders gut im Kindergarten eingesetzt werden kann, da diese im gewohnten Umfeld der Kinder stattfindet:

> *“Thus, this screening method may contribute to a resistance to intervention model of classification (Vellutino et al., 1998) or a more comprehensive model of problem-solving or functional academic assessment (Good & Kaminski, 1996, Lentz & Shapiro, 1986) with kindergarten students.” (VanDerHeyden et al. 2001, S. 376)*

Die Objektivität konnte durch eine hohe Interrater-Übereinstimmung festgestellt werden. Es wurde bei sechs verschiedenen Messverfahren (45 - 68 Paralleltests) bei 15 - 18 Auswertern

eine Übereinstimmung von 98 – 100 % gemessen, was eine hohe Objektivität des Verfahrens bestätigt. Die Paralleltest-Reliabilitäten lagen zwischen .58 ≤ r ≤ .84 (nur zwei Werte unter .70). Anhand 40 per Zufall ausgewählter Kindergartenkinder wurde die Validität der CBM überprüft. Die 40 Kinder wurden sowohl mit den CBM-Proben, als auch mit zwei standardisierten Vorschultests (DIBELS und CIBS-R[20]) überprüft, parallel dazu wurde ein Lehrerranking erstellt. Die Korrelation zwischen der CBM und den herkömmlichen Tests lag im Bereich .18 ≤ r ≤ .76. Bei der CBM und dem Lehrerranking lagen die Korrelationen zwischen .30 ≤ r ≤ .70, wobei acht von zehn Lehrerrankings signifikant waren. Bzgl. der Klassifikationsvalidität konnten 71,4 % der Risikokinder und 94,4 % der Nicht-Risikokinder richtig klassifiziert werden. Hierbei ist zu beachten, dass es sich mit N = 40 um eine sehr kleine Stichprobe handelte und diese nicht als repräsentativ angesehen werden kann. Der Trend scheint zu zeigen, dass die CBM Aussagen über die Basisfertigkeiten von Kindergartenkindern tätigen kann (vgl. a. a. O., S. 380ff.).

Eine sehr ausdifferenzierte CBM, die sich zur Zeit für das Alter von drei bis fünf Jahren im vorschulischen Bereich auf dem US-amerikanischen Markt etabliert, ist die EL-IGDIs (Early literacy IGDIs, Early Childhood Research Institute on Measuring Growth and Development 1998, 2000), die ökonomisch einsetzbar, einfach durchführbar, reliabel und valide, sowie sensitiv bzgl. des Alters und der jeweiligen Intervention ist (vgl. Priest, McConnell, Walker, Carta, Kaminski, McEvoy, Good, Greenwood & Shinn 2001, S. 163f., McConnell, Priest, Davis & McEvoy 2002, S. 1231ff., Roseth, Missall & McConnell 2012, S. 483ff.). EL-IGDIs umfasst das Benennen von Bildern (picture naming), Reimen (rhyming) und die Alliteration und wird dreimal pro Jahr oder nach Bedarf durchgeführt (vgl. zusammenfassend Missall, Reshly, Betts, McConnell, Heistad, Pickart, Sheran & Marston 2007, S. 436ff.). Jede Messung und Dokumentation dauert im Schnitt zehn Minuten pro Kind. Beim „picture naming“ müssen die Kinder innerhalb einer Minute so viele Bilder benennen, wie sie können. Dazu werden ihnen auf Karten Bilder von Gegenständen aus ihrer direkten Umwelt gezeigt. Das Benennen von Bildern hat sich als valider Indikator für die gesprochene Sprache der Kinder gezeigt. Kinder im Alter von 66 Monaten ohne festgestellte Lernschwierigkeiten benennen im Schnitt 26,9 Bilder pro Minute richtig, Kinder mit Lernschwächen dagegen nur 16,9 Bilder (vgl. Priest, McConnell, Walker, Carta, Kaminski, McEvoy, Good, Greenwood & Shinn 2001, S. 163ff.).

Beim „rhyming“ als ein Bestandteil der phonologischen Bewusstheit im weiteren Sinne (vgl. Kap. II.1.2.1) wird den Kindern eine Karte mit einem Bild des Stimulus gezeigt. Es muss

20 CIBS-R ist die Abkürzung für den standardisierten Test „Comprehensive Inventory of Basic Skills–Revised“.

dann auf einen der drei als Bilder angezeigten Begriffe tippen, der mit dem Stimuluswort einen Reim bildet. Die Anzahl der richtig bestimmten Begriffe in zwei Minuten ist das Ergebnis dieser CBM. Für die Kriterienvalidität konnten zu diesem Untertest moderate bis hohe Korrelationen gefunden werden (.54 ≤ r ≤ .64, Clay 1985, sowie .44 ≤ r ≤ 62 bei Torgesen & Bryant 1994, McConnell, Priest, Davis & McEvoy 2002, S. 1231ff). Für die Konstruktvalidität ermittelte das Forschungsteam einen Korrelationswert von r = .46 (vgl. Missall, Reschly, Betts, McConnell, Heistad, Pickart, Sheran & Marston 2007, S. 437). Die Retest-Reliabilität über drei Wochen lag bei .83 ≤ r ≤ .89 (vgl. Missall & McConnell 2004, S. 6f.). Kinder ohne festgestellte Schwierigkeiten schafften im Alter von 53 Monaten 7,6 richtige Reime, Kinder mit identifizierten Schwierigkeiten dagegen nur 6,5 richtige Reime in der vorgegebenen Zeit (vgl. Priest, Silberglitt, Hall & Estrem 2000, zit. in Missall & McConnell 2004, S. 7ff.). Die Submessung „rhyming" ist außerdem sensitiv zu der monatlichen Wachstumsrate der Kinder (vgl. Missall, McConnell & Cardigan 2006, S. 1ff.).

Die dritte Submessung ist die Alliteration, dessen Inhalt dem Bereich der phonologischen Bewusstheit im engeren Sinne zuzuordnen ist. Den Kindern wird eine Karte mit einem Stimulusbild gezeigt. Das Kind soll auf eines der drei darunter angezeigten Bilder tippen, deren zugehöriger Begriff seiner Meinung nach mit dem gleichen Laut beginnt. Es wird innerhalb von zwei Minuten die Anzahl der richtig bestimmten Wörter gemessen. Die Kriterienvalidität wurde zu drei verschiedenen Kriterien bestimmt, die höchste Korrelation konnte zu dem „Test of phonological awareness" mit .75 ≤ r ≤ .79 festgestellt werden (vgl. McConnell, Priest, Davis & McEvoy 2002, S. 1231ff.). Auch für die Retest-Reliabilität konnten nach drei Wochen moderate bis hohe Korrelationen verzeichnet werden (vgl. Missall et al. 2007, S. 437). Im Alter von 53 Monaten können Kinder ohne identifizierte Risiken im Durchschnitt 5,2 richtige Antworten in zwei Minuten erreichen, Risikokinder dagegen im Durchschnitt nur 4,4 (vgl. Missall et al. 2007, S. 438).

Missall et al. (a. a. O., S. 440ff.) konnten zeigen, dass die EL-IGDI-Ergebnisse signifikant moderat mit herkömmlichen Messungen zur phonologischen Bewusstheit und dem alphabetischen Prinzip (Buchstabenwissen usw.) korrelieren. Außerdem ist EL-IGDI ein signifikanter Prädiktor für das spätere flüssige laute Lesen in der Schule. Die diagnostische Nutzbarkeit wurde als sehr hoch eingeschätzt (vgl. a. a. O., S. 433ff.). Auch für drei- bis vierjährige Kinder hat sich das Messverfahren als hoch änderungssensibel erwiesen, dies konnte von Roseth, Missall und McConnell (2012, S. 483ff.) anhand einer beeindruckenden Stichprobe von N = 7 355 Kindergartenkindern bewiesen werden.

Fuchs und Fuchs (2004, S. 32) fordern umfassende curriculumbasierte Messungen als Begleitung des Schriftspracherwerbs im Bereich Lesen vom Kindergarten bis zur sechsten Klasse. In dem Zuge setzen sie bereits im Kindergarten das erste Mal eine CBM ein. Sie nutzen dazu die Aufgabe zur „letter-sound-fluency", die bereits erste schriftsprachliche Entwicklungen im Kindergarten erfassen kann. Im amerikanischen Kindergartensystem ist das Erlernen von Buchstaben-Laut-Korrespondenzen bereits Teil des Kindergarten-Curriculums. Die Kinder benennen anhand eines Bogens mit 25 Buchstaben die jeweiligen korrespondierenden Laute. Der Erzieher notiert die Anzahl der Laute, die das Kind in einer Minute geschafft hat, im CBM-Graph. Nach Fuchs und Fuchs (2004, S. 34ff.) sagt die Aufgabe „letter-sound" in bemerkenswerter Weise die spätere Lesekompetenz voraus. Die CBM ist hoch reliabel (vgl. a. a. O., 2004, S. 34; Fuchs & Fuchs 2007, S. 38) und ist dazu in der Lage Risikokinder zu identifizieren. Kinder, die es am Ende des Kindergartens nicht schaffen den Standard von 35 erkannten Buchstaben pro Minute zu erreichen, benötigen eine zusätzliche Förderung. Haben die Kinder das Ziel von 35 erkannten Buchstaben pro Minute erreicht, so kann direkt die nächste CBM zur Wortidentifikation angewendet werden. Dies ist in der Regel zu Beginn des ersten Schuljahres der Fall (vgl. Fuchs & Fuchs 2004, S. 34). Es folgen bis zum sechsten Schuljahr noch weitere curriculumbasierte Messungen, wie das Lesen von Textpassagen oder die MAZE-Messung mit jeweils unterschiedlichen Zielsetzungen.

Kaminski und Good (1996, S.215ff.) konnten in einer Studie mit 37 Kindergartenkindern und 41 Erstklässlern feststellen, dass sie mit curriculumbasierten Messungen zu den Bereichen der phonologischen Bewusstheit, Wortschatz und der Buchstabenkenntnis die Effektivität von Interventionen überprüfen konnten. Das Messverfahren stellte sich sowohl als reliabel und valide als auch als änderungssensibel heraus. Für gute Leser der ersten Klasse traten jedoch Deckeneffekten auf.

Good und Kaminski (2002) von der Universität Oregon entwickelten die CBM „Dynamic Indicators of Basic Early Literacy Skills" (DIBELS, vgl. auch Good, Gruba & Kaminski 2001, S. 681ff.) zur Früherkennung von Schwierigkeiten im Lesen und Schreiben in den ersten Schuljahren sowie zur Evaluation der Effektivität von Interventionen. DIBELS integriert inhaltlich unter anderem auch Prozesse der phonologischen Bewusstheit. Es ist eine Summation von Messungen, die gemeinsam interpretiert vorhersagen kann, ob ein Kind in der Schule Schwierigkeiten mit dem Lesen und Schreiben entwickeln wird. DIBELS basieren auf CBM und werden dreimal im Schuljahr durchgeführt. Die Messungen werden in einer Minute pro Submessung durchgeführt. Jede der Messungen wurde evaluiert und ist ein reliabler und valider Indikator zur frühen Identifikation von Kindern, deren Lernverlauf nicht

wie erwartet verläuft. Von den 80er Jahren an wurden Studien zur Reliabilität, Validität und Sensitivität von DIBELS durchgeführt (vgl. DIBELS 2012, Good, Gruba & Kaminski 2001, S. 684). Inhaltlich gehören folgende Messungen zu DIBELS:

1) Messungen zur phonologischen Bewusstheit
    - Anlauterkennung bei Wörtern
    - Lautierung eines Wortes
2) Messungen zum alphabetischen Prinzip und Phonemen
    - Pseudowörter lesen
    - Lautes Lesen
3) Messungen zur Genauigkeit und Leseflüssigkeiten von Texten
    - Lautes Lesen
4) Messungen zum Verständnis
    - Lautes Lesen und Nacherzählen von Texten
5) Messungen zu Wortschatz und gesprochener Sprache
    - Wort im Kontext anwenden

DIBELS misst nicht alle Komponenten der phonologischen Bewusstheit, da die Autoren (vgl. Kaminski, Cummings, Powell-Smith & Good 2008) die Lautierung eines Wortes als robusten Indikator festmachen konnten. Ebenso verhält es sich mit den anderen Submessungen. DIBELS ist darauf ausgelegt die Fähigkeiten der Kinder vom Kindergarten bis zum sechsten Schuljahr zu begleiten (vgl. Abb. 32).

<table>
<tr><td></td><td colspan="5">Retell Fluency</td></tr>
<tr><td></td><td colspan="5">Oral Reading Fluency</td></tr>
<tr><td></td><td colspan="2">Nonsense Word Fluency</td><td colspan="3"></td></tr>
<tr><td></td><td colspan="2">Phoneme Segmentation Fluency</td><td colspan="3"></td></tr>
<tr><td colspan="3">Word Use Fluency</td><td colspan="3"></td></tr>
<tr><td colspan="2">Letter Naming Fluency</td><td colspan="4"></td></tr>
<tr><td>Initial Sound Fluency</td><td colspan="5"></td></tr>
<tr><td>Kindergarten</td><td>First grade</td><td>Second grade</td><td>Third grade</td><td>Fourth grade</td><td>Fifth grade</td><td>Sixth grade</td></tr>
</table>

Abb. 33: Messungssystem DIBELS vom Kindergarten bis zum sechsten Schuljahr (verändert nach DIBELS 2002, S. 1)

Allor, Gansle und Denny (2006, S. 23ff.) dokumentieren mit DIBELS täglich den Lernprozess von sechs Kindergartenkindern, die Schwierigkeiten in der phonologischen Bewusstheit aufwiesen. Sie führten mit den Kindern eine Interventionsmaßnahme durch, die im Durchschnitt 26 Minuten pro Tag an Zeit in Anspruch nahm. Es zeigte sich, dass geschultes Kindergartenpersonal dazu in der Lage ist, Risikokinder anhand von CBM zu identifizieren und ihren Lernprozess mit diesem Mittel zu verfolgen. Durch die sehr geringe Stichprobegröße ist die Studie nicht repräsentativ, zeigt jedoch auf, dass die Möglichkeit der Implementierung von CBM im Kindergarten gegeben sein kann, sowohl als Screening-Verfahren, als auch für das „progressmonitoring", also zur regelmäßigen Überprüfung des Lernverlaufs. In der im zweiten Teil der Arbeit vorgestellten eigenen Studie zur curriculumbasierten Messung „Phonolino" werden geschulte Erzieher die CBM ebenfalls selbstständig im Kindergarten durchführen.

Ein umfassendes curriculumbasiertes Verfahren mit mehr als zwei Messzeitpunkten für den vorschulischen Bereich wurde in Deutschland derzeit noch nicht entwickelt. Dies wird Grundlage der Konzeption von „Phonolino" sein, einem curriculumbasierten Messverfahren, das in der vorliegenden Arbeit evaluiert werden soll.

Im folgenden Kapitel wird zunächst die auf der CBM aufbauenden Interventionsmöglichkeiten verdeutlicht und mit dem herkömmlichen Interventionsparadigma verglichen, um nicht nur den Vorteil in Bezug auf eine Möglichkeit der genaueren Messung darzustellen, sondern auch die im Zusammenhang stehende Intervention und die Aussicht auf eine passgenauere individuelle Förderung einzelner Kinder aufzuzeigen.

4 Interventionsparadigma der CBM im Vergleich zum traditionellen Ansatz

Die CBM ist als vielfältiges Messverfahren verwendbar. Deno (2003) beschreibt (vgl. auch Kap. III.2) allein 14 verschiedene Ansätze zum Einsatz der CBM. Im Folgenden wird der Gebrauch von CBM im Rahmen des RTI-Ansatzes („response-to-intervention") erläutert, der sich insbesondere für den Einsatz im Unterricht etabliert hat. Dabei wird auch der Kontrast des RTI-Ansatzes zum ATI-Ansatz („aptitude-treatment-interaction") dargestellt, der als Folgerung zu herkömmlichen Testverfahren als Förderansatz weit verbreitet ist.

4.1 Der ATI-Ansatz („Aptitude-Treatment-Interaction“)

In den letzten Jahrzehnten ist die Unterrichtspraxis der Sonderpädagogik durch den adaptiven Unterricht geprägt. Im englischsprachigen Raum wird dieser Ansatz auch als „Aptitude-Treatment-Interaction“ (ATI) bezeichnet (vgl. Walter 2008a, S. 202ff.). Der Unterricht und die schulischen Institutionen passen sich an die jeweilige individuelle Ausgangslage der Schüler an, indem sie je nach Merkmal des Schülers methodisch oder didaktisch andere Maßnahmen ergreifen (vgl. Abb. 34).

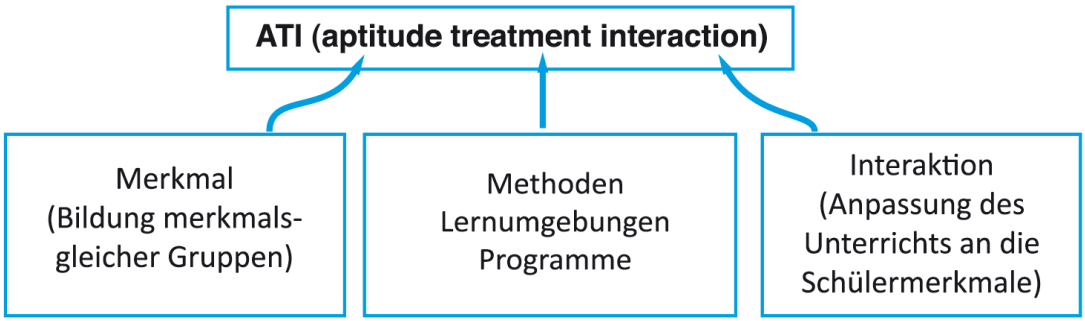


Abb. 34: Begriffsherkunft des traditionellen ATI-Ansatzes

Durch den ATI-Ansatz werden Gruppen je nach Merkmal (Alter, Geschlecht, Leistung, IQ, Sonderschüler, Regelschüler, …) gebildet. Beispielsweise wurde die Gruppe der Legastheniker anhand bestimmter Merkmalsgruppierungen erzeugt, da man vermutete, dass das Merkmal „Legastheniker“ eine bestimmte Methode der Instruktion nach sich ziehen würde. Es wurde damit eine direkte Verbindung vom Merkmal zu Maßnahmen der Förderung gezogen. Diese Zusammenhänge wurden nicht immer wissenschaftlich erforscht, sondern beruhten oft auf „pädagogischer Intuition“ oder gingen bestimmten Trends, wie beispielsweise der Bestimmung von Gruppen nach Lerntypen nach. Bestimmte Merkmalsgruppierungen zum Beispiel im Bereich der Lese-Rechtschreibschwäche konnten durch einschlägige Forschungsergebnisse widerlegt werden. Wie bereits in Kapitel I.1.3.1 beschrieben, konnte nachgewiesen werden, dass Schüler mit Schwierigkeiten im Lesen und Schreiben nicht zusätzlich anhand des Merkmals der Intelligenz unterschieden werden müssen. Beim ATI-Ansatz werden also vorher merkmalsgleiche Gruppen unter anderem durch den Einsatz von herkömmlichen Testverfahren zusammengestellt. Die Schüler dieser merkmalsgleichen Gruppen werden dann mit der gleichen Interventionsmaßnahme gefördert. Es ist somit notwendig vorher zu wissen, welche Merkmale von Unterricht Schülergruppen identifizieren, die auf bestimmte Methoden reagieren, was somit einer präskriptiven Diagnostik gleichkommt. Der Erfolg bzw. die Effektivität ist damit von den Schülermerkmalen und den Merkmalen des Programms abhängig. Der ATI-Ansatz

unterstellt, dass zwischen den Lernvoraussetzungen bzw. Schülermerkmalen und den verschiedenen Unterrichtsarten, Programmen und Fördermöglichkeiten valide Interaktionen bestehen (vgl. Walter 2008a, S. 202f.).

Der ATI-Ansatz bringt auch eine Reihe von Schwierigkeiten mit sich. Die Merkmalsdimensionen sind häufig nur mangelhaft erforscht und die Gruppierungen nach Schülermerkmalen werden oft wahllos gebildet. Die Testverfahren selbst sind meist nur einmalig einsetzbar und nicht curricular valide. Die Tests sind komplex und zeitaufwändig, so dass sie nicht im Klassenraum vom Lehrer selbst eingesetzt werden können. Durch die deduktive Vorgehensweise müsste das Wissen über Schülermerkmale bereits im Vorhinein bekannt sein. Walter (2008a, S. 210) konstatiert:

> *„Möglicherweise ist die Annahme unrealistisch, dass unsere Messverfahren in der Lage sind, auf reliable und valide Art und Weise eine Reihe von Individuen in homogenen Gruppen auf der Basis der Merkmale zusammenzufassen, die relevant mit den Treatments interagieren."*

Es herrscht in der Realität jedoch häufig mangelhaftes Wissen über die Interaktion zwischen den Lehrer- und Schülermerkmalen, den Interventionsprogrammen oder der Klassensituation direkt vor Ort (vgl. Walter 2007c, S. 4). Oft wurden vermutete Zusammenhänge auch widerlegt, wie in dem Fall, dass in merkmalsgleichen Gruppen zum Merkmal (aptitude) „lernbehindert" das Treatment „Sonderbeschulung" größere Effekte mit sich bringt, als das Treatment „Regelbeschulung" (vgl. zusammenfassend Walter 2002b, S. 442ff., 2007b, S. 891f.). Walter (2008a, S. 202) vermutet sogar, dass das Festhalten an dem dominierenden Ansatz den förderpädagogischen Fortschritt blockiert.

Nachfolgend wird der Ansatz „Response-to-intervention", der im Zusammenhang mit der in dieser Arbeit im Mittelpunkt stehenden curriculumbasierten Messung steht und konträr zum herkömmlichen ATI-Ansatz zu betrachten ist.

4.2 Der RTI-Ansatz („Response-To-Intervention")

Der dominierende ATI-Ansatz bringt Probleme mit sich. Die präskriptive Didaktik wird zu stark akzentuiert, durch eine prognostisch-präskriptive Zuweisung werden Förderprogramme mit Kindern durchgeführt, deren Fortgang jedoch nicht weiter verfolgt wird (vgl. Walter 2008a, S. 210). Außerdem scheint die Aussagekraft hinsichtlich sonderpädagogischer

merkmalsgeleiteter Interventionen, die für eine Gruppe gelten, nicht zwingend auf den einzelnen Schüler übertragbar zu sein (vgl. a. a. O., S. 211).

In den USA hat sich seit einigen Jahren im Zuge der Forschung zur CBM ein konkurrierender Ansatz entwickelt, der ein anderes Praxismodell unterstützt. Beim RTI-Ansatz („response-to-intervention", vgl. Abb. 35) werden nicht schon im Vorfeld merkmalsgleiche Gruppen gebildet, um im Anschluss eine Interventionsmaßnahme auszuwählen, sondern es wird eine ständige Anpassung und Überprüfung der Intervention und des Unterrichts gefordert (vgl. Fuchs & Fuchs 1998, S. 204ff.). Der Ansatz ist auch als systematisch-formative Evaluation (vgl. Fuchs & Fuchs 1986, S. 199ff.; Fuchs, Fuchs, Hamlett, Walz & Germann 1993, S. 27ff.) oder Treatment-Validität (vgl. Fuchs & Fuchs 1998, S. 204) bekannt. Die Intervention wird in Abhängigkeit vom Niveau und dem Lernzuwachs der Kinder gestaltet und bei Misserfolg dementsprechend auch kurzfristig angepasst.

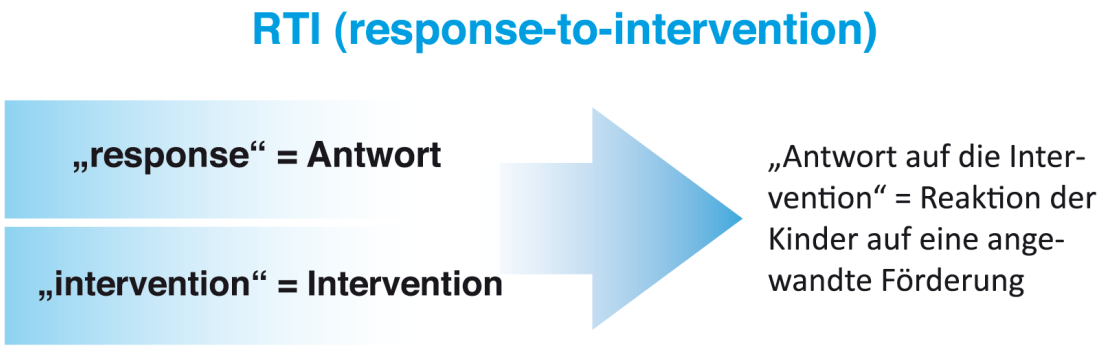


Abb. 35: Begriffsherkunft des neueren RTI-Ansatzes

RTI ist eine Methode, die den Schülern zu erfolgreichem Lernen in der Schule verhelfen soll, indem systematische Hilfen im normalen, wie im speziellen Unterricht geboten werden (vgl. Martinez, Nellis & Prendergast 2006, S. 5).

Der RTI-Prozess wird üblicherweise in drei Stufen geteilt (vgl. ebd. S. 3ff.; vgl. Abb. 36). Stufe 1 (Tier 1) beschreibt eine Art der primären Prävention, in dem die Schüler im Rahmen des Unterrichts individuell gefördert werden und der Lernprozess der Kinder durch den Lehrer oder Erzieher immer wieder evaluiert wird, um Schwierigkeiten frühzeitig erkennen zu können. Stufe 2 (Tier 2) kann mit der sekundären Prävention verglichen werden. Hier werden den Schülern zusätzlich Interventionen zum laufenden Unterricht angeboten. Der Lernverlauf der Schüler wird weiterhin dokumentiert und bei einem ungünstigen Verlauf wird die Interventionsmaßnahme angepasst. Wenn der Lernverlauf der Schüler sich wieder angleicht, können diese zum normalen Unterricht ohne zusätzliche Maßnahmen zurückkehren. Trotz bestmöglicher Unterstützung sind einige Schüler in Stufe 1 und Stufe 2 gescheitert. Es folgt eine intensive Instruktion in Stufe 3 (Tier 3), in der auf diese Schüler noch individueller eingegangen werden kann (z.B. in einer Eins-zu-Eins-Situation).

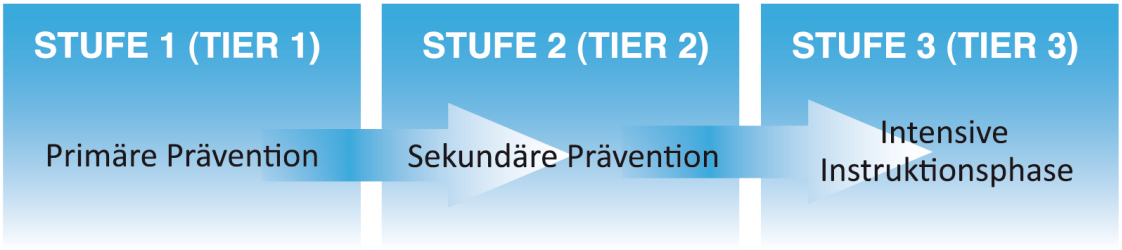


Abb. 36: Prävention nach dem RTI-Modell

Müller und Hartmann (2009, S. 27f.) passten dieses Konzept für den deutschsprachigen Raum in einem Modell an (vgl. Abb. 37). Sie gliedern das RTI-Modell in die drei Ebenen „Regulärer Unterricht“ (1), „Fokussierte Intervention“ (2) und „Spezielle Intervention“ (3).

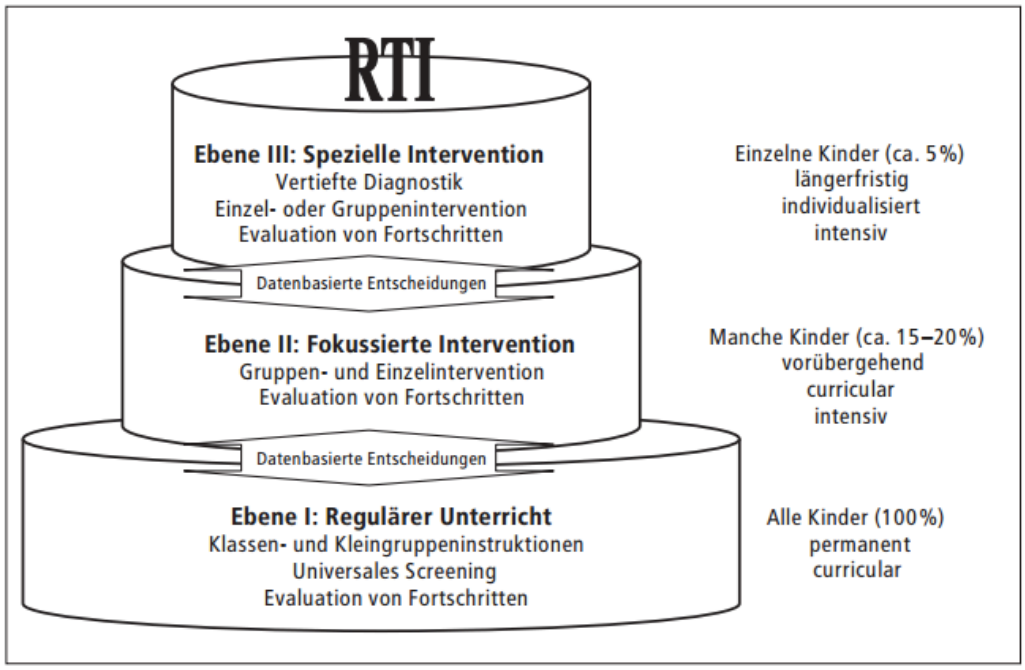


Abb. 37: RTI-Modell (zit. nach Müller & Hartmann 2009, S. 27)

Die unterste Ebene ist der Zuständigkeitsbereich von Regellehrkräften, die Screenings und systematische Fortschrittsmessungen für ihre Schülerinnen und Schüler vorsehen. Hier werden Risiko-Kinder identifiziert, die ggf. auf der zweiten Ebene Interventionen erhalten. Diese „fokussierten Interventionen“ auf der zweiten Ebene werden ergänzend zum täglichen Unterricht durchgeführt. Dabei werden curriculare Lernziele in den Fokus genommen, die von Regelschullehrkräften oder auch Spezialisten durchgeführt und auch evaluiert werden. Bei Erfolg der Förderung, wird diese wieder reduziert und die Kinder erhalten keine zusätzlichen Interventionen auf der Ebene 2. Bei Nicht-Erfolg werden die Maßnahmen modifiziert oder auch verlängert. Bleibt der Erfolg langfristig aus, wird die „spezielle

Intervention“ auf der dritten Ebene aktiviert. Es handelt sich dabei um spezielle sonderpädagogische Interventionsmaßnahmen, die auf einer gründlichen Diagnostik basieren und ausschließlich von speziell ausgebildetem Personal übernommen werden.

Fuchs und Fuchs (1986, S. 203ff.) konnten anhand einer Meta-Analyse von 21 Studien mit insgesamt 3 835 Schülern feststellen, dass die systematisch-formative Evaluation im ausgewogenen Durchschnitt eine Effektstärke von .70 hat, die als moderat[21] zu bewerten ist (vgl. Cohen 1988). Lysakowski und Walberg (1982, S. 559ff.) konnten in einer Studie mit 14 689 Schülern sogar eine Effektstärke von ES = .97 feststellen. Der Lernzuwachs kann weiterhin unterstützt werden, wenn der Lehrer neben der systematisch-formativen Evaluation Techniken der Verhaltensmodifikation anwendet, Regeln der Dateninterpretation einhält und die Daten grafisch darstellt (vgl. auch Walter 2008a, S. 213ff., Walter 2014, S. 167). Es konnte in mehreren Studien gezeigt werden, dass sich ein formatives Vorgehen nicht nur aus bildungsökonomischer Sichtweise als vorteilhaft erweist, sondern sich gleichwohl die Leistungen der Schüler im Durchschnitt von knapp einer dreiviertel Standard-abweichung verbessern (vgl. zusammenfassend Walter 2011b, S. 204ff.).

Lernstörungen oder niedrige Leistungen in verschiedenen Bereichen werden somit als Resultat mangelnder Treatment-Validität betrachtet. Bei jedem Kind wird individuell überprüft, inwieweit es auf bestimmte Fördermaßnahmen anspricht. Die Datenbasis für diesen Ansatz bietet die CBM als systematisch-formatives Messverfahren, die den Lernfortschritt der Schüler dokumentiert. Durch die CBM können sogenannte „Responder“ und „Non-Responder“ klassifiziert werden. „Responder“ reagieren auf eine angewandte Förderung positiv und können einen beständigen Lernzuwachs verzeichnen, während „Non-Responder“ aufgrund der derzeitigen Intervention nicht ausreichenden Lernzuwachs verzeichnen können.

Fuchs und Fuchs (1998, S. 205ff.) beschreiben dazu ein Vier-Phasen-Modell (vgl. Abb. 38). In der ersten Phase wird überprüft, inwieweit in der jeweiligen Klasse angemessene Lernfortschritte in dem gewählten Bereich (z.B. richtig gelesene Wörter pro Minute) möglich sind (z. B. anhand von Klassennormen). In der zweiten Phase werden individuelle Lernniveaus und Lernfortschritte miteinander verglichen. Bei einer doppelten Diskrepanz zwischen Leistungsniveau und Leistungsfortschritt im Vergleich zum Klassendurchschnitt gilt das Kind als „Non-Responder“ und es wird die dritte Phase eingeleitet.

21 Nach Cohen (1988) ist eine Effektstärke von ES = .20 als klein, eine ES = .50 als moderat und eine ES > .80 als groß einzuschätzen.

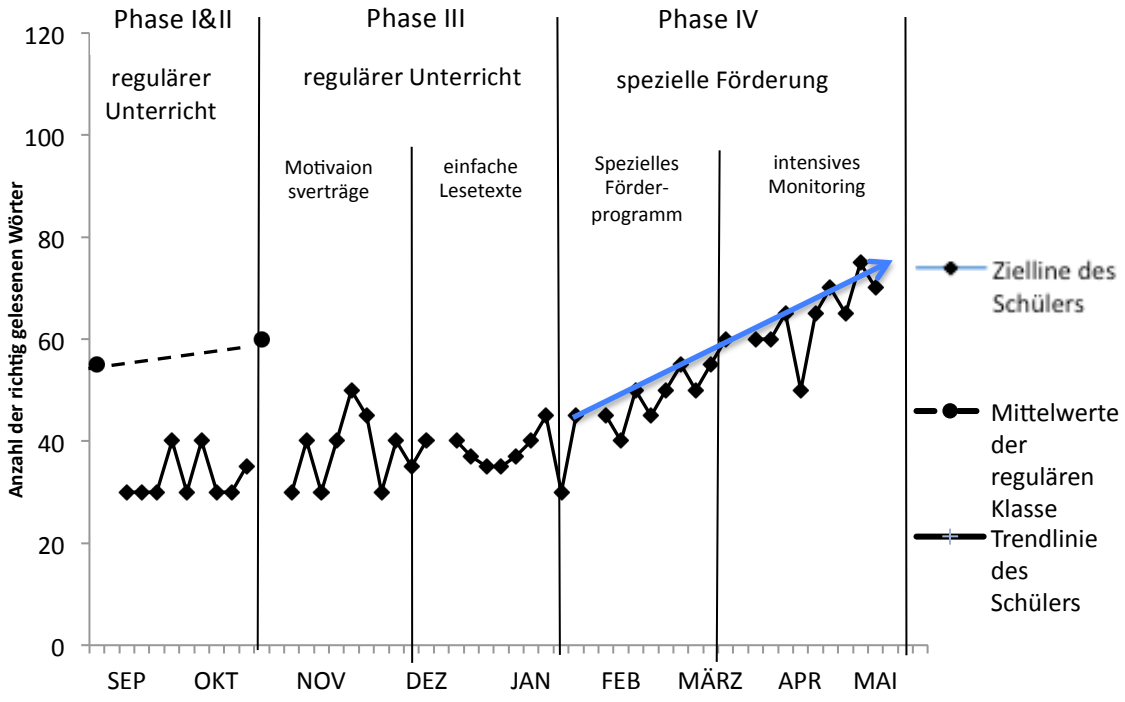


Abb. 38: 4-Phasen-Modell dargestellt an einer exemplarischen Datenerhebung der CBM „richtig gelesene Wörter pro Minute“ (aus dem Englischen übersetzt, angelehnt an Fuchs & Fuchs 1998, S.206)

In der dritten Phase wird die Wirksamkeit von zusätzlichen anderen Fördermaßnahmen gemessen, d.h. die neuen Interventionsmaßnahmen werden für dieses Kind evaluiert. Sollte das Kind trotz der veränderten Intervention weiterhin als „Non-Responder“ eingestuft werden, beginnt die vierte Phase. In der vierten Phase werden sonderpädagogische Maßnahmen angestrebt (z.B. Token-Programme, 1:1 Fördersituation, …) und weiterhin mit der CBM evaluiert. Wenn das Kind in dem Sonderunterricht weiterhin keine Fortschritte macht, dann argumentierten die Autoren 1998 noch, dass es keinen Grund dafür gäbe, das Kind weiterhin in der Sonderbeschulung zu halten (vgl. Fuchs & Fuchs 1998, S. 206ff.). Sie weichen dies in späteren Veröffentlichungen wieder auf, da es sein kann, dass ein Non-Responder in der vierten Phase besonders hochwertige Förderung oder mehr Zeit benötigt und das Curriculum der Regelschule dafür ein Problem darstellt (vgl. Fuchs, Fuchs & Speece 2002, S. 42ff.). Die Autoren Fuchs, Fuchs und Speece (ebd.) definieren 2002 die vierte Phase als erweiterte Diagnostikphase neu, in der auf Grundlage der systematisch-formativen Evaluation im Acht-Wochen-Rhythmus weitere sonderpädagogische Maßnahmen ergriffen werden, sie schließen eine Zurückführung des Kindes zum Regelunterricht jedoch nicht aus (vgl. zusammenfassend Walter 2008a, S. 210ff.):

> *"Given a straightforward implementation of this decision-making process, these students are returned to the regular classroom without additional support when, clearly some action is required." (Fuchs, Fuchs & Speece 2002, S. 43)*

In der praktischen Umsetzung wird in den Phasen auf die einzelnen Werte der CBM zurückgegriffen. Nach Wright (o. J., S. 4) wird nach den ersten drei Messungen zunächst ein Ziel festgelegt, welches das Kind erreichen soll (vgl. Abb. 39). Um ein realistisches Ziel festlegen zu können, wird als erstes die aktuelle Leistung aus den ersten drei Werten der CBM berechnet, indem der Mittelwert gebildet wird. Das Leistungsziel wird anhand der jetzigen Leistung des Kindes bestimmt und beschreibt die nach dem angesetzten Zeitraum erwartete Leistung.

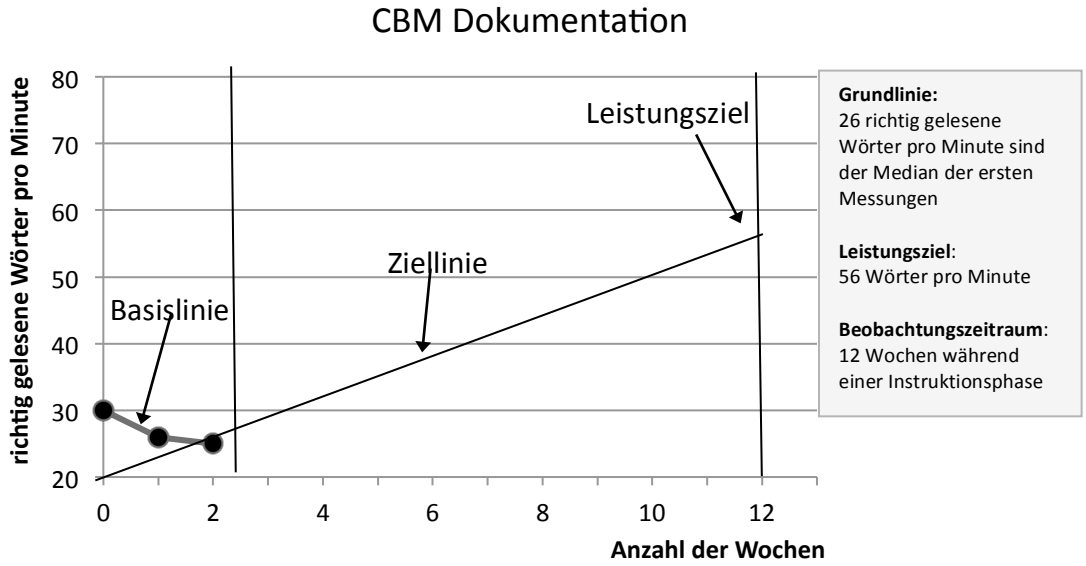


Abb. 39: Beispiel einer Dokumentation zur CBM „richtig gelesene Wörter pro Minute" zur Berechnung einer Ziellinie (aus dem Englischen übersetzt, angelehnt an Wright o. J., S. 4)

Nach Hosp, Hosp und Howell (2006, S. 120ff.) gibt es drei verschiedene Methoden im Graph ein Ziel für den Schüler zu setzen. Das Ziel kann vom Ziel des Curriculums abhängig gemacht werden, sodass das Ziel an der Stelle gesetzt wird, die durch das Jahrescurriculum für die Schüler vorgegeben ist („end-of-year benchmarks"). Dies wird von Leistungsstandards abhängig gemacht, mit Zielen, zu denen die Schüler der jeweiligen Klasse am Ende geführt werden sollen. Dabei wird der niedrigste Wert bei schwächeren Schülern angesetzt, sodass diese dann idealerweise zum Ende des Jahres nicht mehr zu den gefährdeten Schülern des Jahrgangs gerechnet werden müssen. Eine zweite Möglichkeit ist der Vergleich mit einer genormten Vergleichsgruppe. Hierzu ist es sinnvoll nationale

Normen zu verwenden. Die Leistungen der Schüler werden dann mit den Normen verglichen. Dies ist nur dann möglich, wenn diese Daten bereits erhoben worden sind. Eine dritte Möglichkeit ist die Definition von Zielen durch intraindividuelle Vergleiche. Dazu werden von dem Schüler zu mehreren Messzeitpunkten Daten erhoben. Durch Berechnungen anhand der Leistungsdaten des Schülers sowie dessen Lernzuwachsraten wird das Lernziel zu Jahresende bestimmt. Hosp, Hosp und Howell (2006, S. 123) betonen, dass es nur sinnvoll ist, diese intraindividuellen Vergleiche zu nutzen, wenn die vorherigen Leistungen des Schülers dem Durchschnitt entsprachen, da ansonsten ein sehr niedriges Ziel gesetzt wird und damit die tatsächlichen zukünftigen Leistungen des Schülers unterschätzt werden könnten.

Im Anschluss an die Zielsetzung werden wöchentlich die curriculumbasierten Messungen durchgeführt. Anhand der Ziellinie wird dann sichtbar, ob ein Schüler sich durch eine Interventionsmaßnahme weiterentwickelt (Responder) oder nicht (Non-Responder). Dies ist auch durch die Trendlinie sichtbar (vgl. Abb. 40).

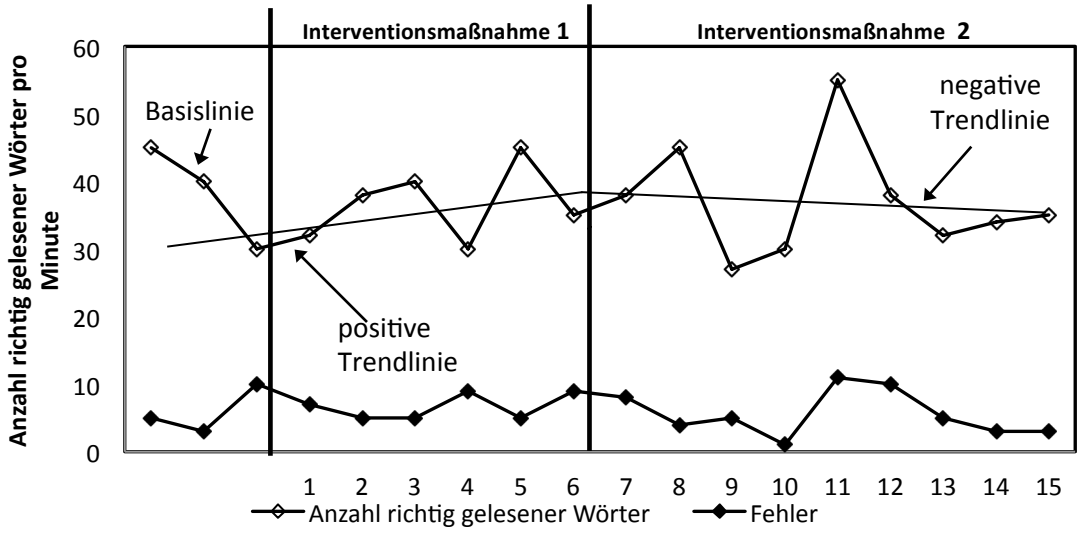


Abb. 40: Beispiel einer CBM zum Lesen (richtige Wörter pro Minute) für eine negative und eine positive Trendlinie innerhalb zweier Interventionsmaßnahmen (nach Wright o. J., S. 10)

Bei der Schülerin in Abbildung 40 konnte in der ersten Interventionsmaßnahme zwar eine positive Trendlinie beobachtet werden, sie erreicht andererseits nicht die erwartete Zunahme an richtig gelesenen Wörtern pro Minute. Die Lehrerin wechselt aus diesem Grund die Interventionsmaßnahme. Dennoch stellte sich heraus, dass die zweite Interventionsmaßnahme noch weniger erfolgreich war als die erste (negative Trendlinie).

Wright (o. J., S. 6) hat die „Drei-Punkte-Regelung“ zur Identifikation eines „Non-Responder“ festgelegt. Sobald die CBM-Werte eines Kindes dreimal hintereinander unterhalb der Ziellinie liegen (vgl. Abb. 41) wird die Intervention angepasst.

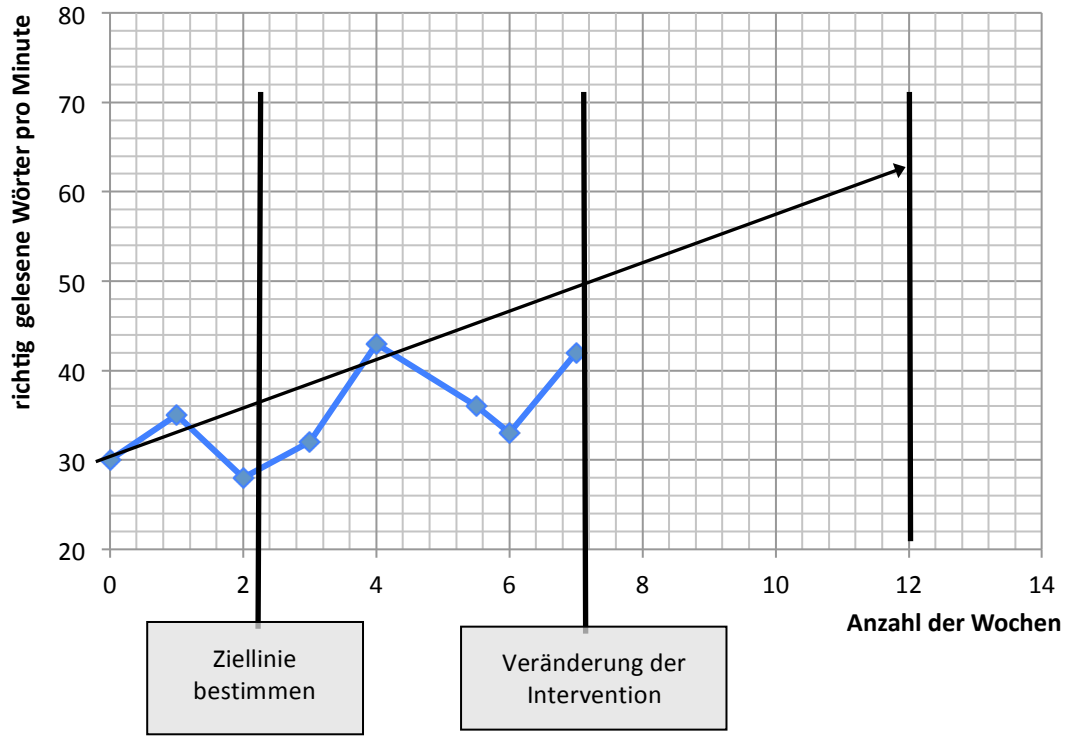


Abb. 41: Beispiel für eine Änderung der Intervention nach der Drei-Punkte-Regel von Wright (angelehnt an Walter 2007c)

Liegen drei Messungen oberhalb der Ziellinie, so wird das Ziel nach oben korrigiert. Fluoreszieren die Werte jedoch um die Ziellinie herum, so wird die aktuelle Förderung weitergeführt (vgl. Abb. 42).

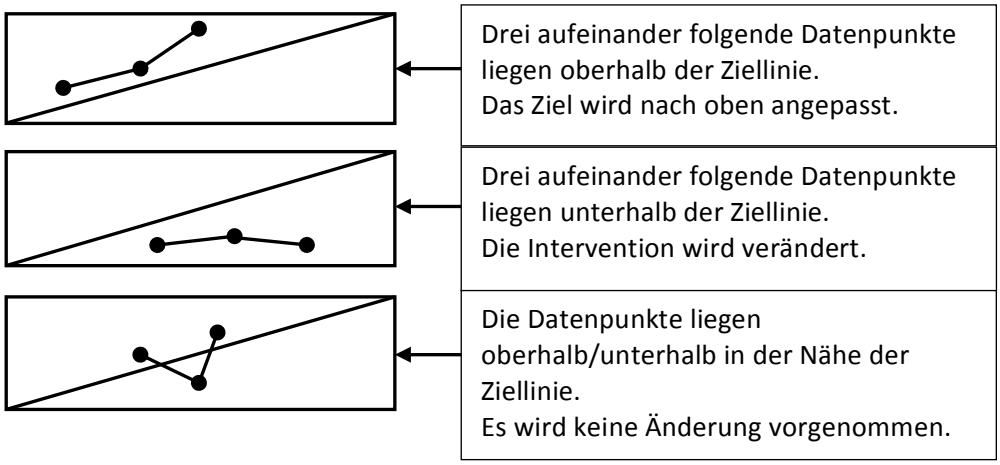


Abb. 42: Drei-Punkte-Regelung (übersetzt aus dem Englischen nach Wright o. J., S. 6)

Das RTI-Modell wird seit dem Schuljahr 2010/2011 auch in Deutschland im Rahmen eines umfangreichen Forschungsprojekts der Universität Rostock, dem „Rügener

Inklusionsmodell“ (RIM), evaluiert. Zielsetzung ist die Etablierung des gemeinsamen Unterrichts von Kindern mit und ohne Förderbedarf. Erste Befunde zu den Lernbereichen Deutsch und Mathematik sowie zu den Förderbereichen emotionale-soziale Entwicklung sowie der sprachlichen Entwicklung liegen vor (vgl. Mahlau, Blumenthal, Diehl, Schöning, Sikora, Voß & Hartke 2014, S. 101). Bisherige Erkenntnisse weisen auf eine positive Entwicklung der inklusiven Schulstruktur hin (vgl. Voß, Blumenthal, Sikora, Mahlau, Diehl & Hartke 2014, S. 127ff.). Der Response-to-Intervention-Ansatz dient in dem Projekt im Rahmen der Inklusionsdebatte der Identifikation und Förderung individueller Entwicklung von Schülern. Ansätze, die individuelle Lernausgangslage im Sinne einer Förderplanung umzusetzen, oder ggf. sogar sonderpädagogische Feststellungsverfahren damit zu ersetzen, werden auf der politischen Ebene bisher nicht diskutiert (vgl. dazu bspw. Ministerium für Schule und Weiterbildung NRW 2014, S. 4).

Im Rahmen der Rostocker Studie konnten die Forscher für den Raum Rügen für die VERA[22]-Ergebnisse feststellen (vgl. Voß, Mahlau, Sikora, Blumenthal, Diehl & Hartke 2015, S. 153):

> *„Das Leistungsniveau der Rügener Kinder entspricht ‚trotz Inklusion’ dem Landesdurchschnitt, was in den Vorjahren nicht immer der Fall war.“*

Außerdem konnte für die Bereiche Mathematik, Lesen und Rechtschreiben nachgewiesen werden, dass die Rügener Schülerinnen und Schüler mit besonderem Förderbedarf erfolgreich gefördert werden konnten (vgl. a. a. O., S. 159). Mahlau (2013, S. 4) hebt hervor, dass besonders Kinder mit einer spezifischen Sprachentwicklungsstörung im Vergleich zur Kontrollgruppe ihre Leistung im Wortschatz und zum Teil in der Grammatik signifikant verbessern konnten. Dennoch beschreiben die Autoren Mahlau et al. (2014, S. 120), dass das Potenzial des RTI-Ansatzes in dem Rügener Inklusionsmodell noch nicht ausreichend ausgeschöpft werden konnte. Gründe könnten die Schwierigkeit der Lehrkräfte bei der Umsetzung des Förderkonzeptes, ggf. mangelnde direkte Instruktionen, korrektive Feedbacks oder tutorielle Unterstützungsmaßnahmen im Rahmen des offenen Unterrichts, sowie die Nutzung der Ergebnisse der diagnostischen Verfahren durch die Lehrkräfte sein.

Souvignier und Förster (2011, S. 243) konnten in einer computergestützten Maßnahme zur Lernverlaufsdiagnostik feststellen, dass nach einer sechsmonatigen Förderung der N = 144 leseschwächsten Schülerinnen und Schüler aus 24 Klassen, in der eine Gruppe der Lehrkräfte Informationen über die Leseentwicklung der Schüler erhielt, während die andere Gruppe diese nicht bekam, die Lernzuwächse in der ersten Gruppe deutlich höher waren als

22 VERA sind zentrale VERgleichsArbeiten, die in den Grundschulen und den weiterführenden allgemeinbildenden Schulen in allen Bundesländern in den Fächern Deutsch, Mathematik und Englisch geschrieben werden.

in der Vergleichsgruppe. Es stellte sich damit heraus, dass die Informationen über die Lernfortschrittsdiagnostik sinnvoll für die Planung von Fördermaßnahmen genutzt werden konnten.

Blumenthal, Kuhlmann und Hartke (2014, S. 61) beschreiben die Vorteile des RTI-Ansatzes mit folgenden Stichpunkten:

- Frühe Förderung lernschwacher Schüler
- Flexibles System der Förderung auf mehreren Ebenen
- Evidenzbasierung von Unterricht und Förderung
- Einsatz formativer Leistungsdiagnostik als Rückmeldung zur Qualität von Unterricht und Förderung
- Selektion pädagogischer Maßnahmen anstelle der Selektion von Schülern

Walter (2007c, S. 12) beschreibt nach Fuchs und Fuchs (1998, S. 208ff.) und Vaughn und Fuchs (2003) fünf zentrale Punkte, die der Ansatz der systematisch formativen Evaluation für die Überwachung der Lernfortschritte der Schüler bietet, um die Reaktion auf bestimmte Fördermaßnahmen zu überprüfen:

> - *„Das induktive Vorgehen stützt sich nicht auf (häufig nicht vorhandenes) A-Priori-Wissen über die Beziehung zwischen Schülermerkmalen und Interventionsformen.*
> - *Die verwendeten psychometrischen Messverfahren (CBM) beziehen sich auf schulisch-curriculare Inhalte (im Gegensatz zu IQ-Tests, Wahrnehmungstests, allgemeine Leistungstests) und besitzen damit eine höhere curriculare Validität.*
> - *Die verwendeten psychometrischen Messverfahren (CBM) werden im Klassenzimmer verwendet und besitzen damit eine hohe ökologische Validität.*
> - *Die verwendeten psychometrischen Messverfahren (CBM) sind auf Veränderungsmessung ausgerichtet und können im Gegensatz zu klassischen Test z.B. wöchentlich während des Schuljahrs wiederholt werden.*
> - *Systematische Formative Evaluation hat empirisch nachgewiesene Vorteile (vgl. Meta-Analyse von Fuchs & Fuchs 1986).“*

Soll dieser Ansatz auf den Kindergarten und die phonologische Informationsverarbeitung übertragen werden, so gilt es darauf hinzuweisen, dass im Kindergarten nicht wie in der Schule angewandte Unterrichtsmodelle überprüft werden können, da eine gezielte Form von Unterricht in deutschen Kindergärten nicht möglich ist. Der RTI-Ansatz eignet sich jedoch

gerade im vorschulischen Umfeld dazu, entwicklungsprozessbegleitende Beobachtung und Dokumentation durchzuführen, um die Lernfortschritte der Kinder im Bereich der phonologischen Informationsverarbeitung zu dokumentieren und zu beobachten und ggf. gezielte Intervention anzubieten (vgl. VanDerHeyden & Snyder 2006, S. 523f.). In vielen Kindertagesstätten in Deutschland werden zudem vermehrt im letzten Kindergartenjahr Interventionsprogramme zur phonologischen Informationsverarbeitung eingesetzt (z.B. „Hören, lauschen, lernen“ von Küspert und Schneider 2006 oder „Wuppis Abenteuer-Reise“ von Christiansen 2007). Auf der Grundlage einer Auswertung von CBM-Daten könnten diese Programme bzgl. ihrer Effektivität beim einzelnen Kind überprüft werden. Bei Bedarf kann dann die Förderung verstärkt, verändert oder durch andere Methoden und Inhalte ergänzt werden. VanDerHeyden und Snyder (2006, S. 526) schlagen, wie in Abbildung 43 dargestellt, vier Kernelemente für die Früherziehung vor.



Abb. 43: Kernelemente des RTI im Kindergarten (aus dem Englischen übersetzt, angelehnt an VanDerHeyden & Snyder 2006, S. 526)

Der RTI-Ansatz bietet damit sowohl für den Einsatz in der Schule, als auch für den Einsatz im Kindergarten die Perspektive einer adäquaten Prävention und Diagnostik. Er bietet Möglichkeiten der Individualisierung, der Begleitung und Beobachtung bzw. Modifikation der Interventionen (vgl. Walter 2008a, S.202ff.), gleichwohl in einigen Bereichen noch ergänzender Klärungs- und Konkretisierungsbedarf besteht (vgl. Walter 2008a, S. 213ff.)

und es besonders im Bereich der Frühpädagogik noch an einschlägigen empirischen Studien im deutschsprachigen Bereich fehlt. Bezüglich des bildungsökonomischen Aspekts konnte bspw. Yeh (2009, S. 7ff.) in den USA anhand von Meta-Analysen nachweisen, dass die schulstrukturelle Einführung eines validen formativen Assessments sich auch finanziell niederschlagen würde. Mit dem Faktor 124:1 könnten derlei Maßnahmen einen Kostenvorteil gegenüber der Reduktion der Klassengrößen darstellen.

Mit Hilfe der Methodik der CBM, die sich anhand der empirischen Befunde als effektives Messverfahren in spezifischen Bereichen erwiesen hat, wird die Grundlage für effektiven und individualisierten Unterricht geschaffen und der Grundstein für eine frühzeitige Erkennung und Frühförderung gelegt, die bestenfalls bereits im Kindergarten stattfinden. Es bleibt nun die Aufgabe, auf der Basis der Forschung zu den existierenden (vgl. bspw. LDL, Walter 2010a) noch weitere Aufgabenstichproben zu entwickeln und zu evaluieren, sowie dieses Modell in der Praxis zu implementieren.

Im nächsten Kapitel werden die curriculumbasierten Messungen unter dem RTI-Paradigma kritisch betrachtet, um für die Konzeption der in dieser Arbeit erstellten CBM Phonolino Vor- und Nachteile, sowie mögliche Schwierigkeiten im Blick zu haben.

5 Kritische Betrachtung der CBM unter dem RTI-Paradigma

CBM wird bereits in den USA als geeignetes Messverfahren anerkannt, das die Lernentwicklung von Kindern über einen längeren Zeitraum beschreibt und mit wenig Zeitaufwand dokumentiert. Vorteile und Nachteile des Verfahrens insbesondere in Bezug auf den Einsatz im Kindergarten werden im folgenden Abschnitt erläutert.

Als größte Schwierigkeit der CBM wird die geringe Augenscheinvalidität (face validity) angesehen. CBM wird als Messverfahren unter Lehrern und Erziehern nicht häufig eingesetzt (vgl. Yell, Deno & Marston 1992, S. 108ff.). In Deutschland gibt es dazu bislang keine Aussagen, es ist doch zu vermuten, dass die Zahl der Lehrer, die von diesem Verfahren schon einmal gehört haben, sehr begrenzt ist, da sich neu erstellte Verfahren (z.B. LDL, Walter 2010a) erst am Markt etablieren müssen. Auch Danielson, Doolittle und Bradley (2007, S. 632ff.) merken kritisch an, dass in den USA die Lehrer verstehen müssen, welche Vorteile ihnen die CBM verschafft. Es bedarf verständlicher und handlicher Messverfahren, wenn dieses Diagnostikkonzept im Zuge des „No Child Left Behind Acts“ flächendeckend für alle Schüler eingesetzt werden soll. Das Vertrauen der Lehrer und Erzieher in dieses

Messverfahren könnte nach Potter und Wamre (1990 in Madelaine & Wheldall 2004, S. 68f.) dadurch gefördert werden, dass es Lehrer und Erzieher unter Anleitung aktiv in der Schule einsetzen und damit dessen Potenzial direkt in der Praxis erleben können. Auch Ardoin, Witt, Suldo, Connell, Koenig, Resetar, Slider und Williams (2004, S. 230) schlussfolgern:

> *"Although teachers may never be convinced of the value of R – CBM in assessing the multiple dimensions of reading, universal screening using R – CBM is cost effective, quick and easy to conduct, and provides results that can be used immediately by schools for identifying students in need of early intervention."*

Die Auswahl der repräsentativen Stichproben ist nur dann möglich, wenn klar definiert worden ist, wofür diese Stichproben repräsentativ stehen sollen und dies wissenschaftlich überprüft worden ist (vgl. Klauer 2006, S. 24). Nach Klauer (ebd.) verzeichnen selbst die amerikanischen Wissenschaftler einen Mangel an einer Grundmenge an Aufgaben mit denen die Lehrziele definiert werden. Hier herrscht in der Forschung noch Nachholbedarf. Somit kann es bei der CBM zu den Schwierigkeiten kommen, dass die Aufgabengrundmengen nicht ausreichend definiert und damit nicht repräsentativ sind. Da sich die englischsprachigen Programme nicht ohne weiteres auf den deutschsprachigen Raum übertragen lassen, muss gerade die Forschung in Deutschland noch geeignete curriculumbasierte Messverfahren für den Einsatz in Schule und Kindergarten konzipieren und evaluieren.

Besonders kritisch wird die Lernverlaufsdiagnostik als Möglichkeit Lernstörungen zu diagnostizieren bewertet. Es werden die mangelnde Validität und Reliabilität, die Schwierigkeiten hinsichtlich der Bezugsnorm oder der Festlegung des Cut-Off-Wertes (zur Unterscheidung zwischen Respondern und Nonrespondern) sowie die nicht vorhandenen standardisierten Richtlinien für die praktische Umsetzung des RTI-Ansatzes (bspw. Verwendung von Methoden und Materialien und Zuweisungskritierien zu den jeweiligen Förderebenen) kritisiert (vgl. Blumenthal, Kuhlmann & Hartke 2014, S. 77).

Eine weitere Schwierigkeit ist, dass bei einer Lernverlaufsdiagnostik von komplexeren Zielen oder Fertigkeiten in der CBM zum Teil schon Aufgaben gestellt werden müssen, die das Kind noch nicht lösen kann (vgl. ebd.). Dies ist für die Motivation der Kinder ein deutlicher Nachteil der CBM, der insbesondere dann auftritt, wenn der Lernverlauf über eine längere Zeit (z. B. über ein Schuljahr hinweg) dokumentiert wird. Der Nachteil bei einer

CBM-Messung von einzelnen Fertigkeiten ist zudem, dass der Lernverlauf oft nur über einen kürzeren Zeitraum abgebildet werden kann und nicht über bspw. ein gesamtes Schuljahr, da viele Teilfertigkeiten im Verlauf des Jahres hinzukommen und Deckeneffekte auftreten können.

Petermann (2006, S. 824) weist zudem auf die Stichprobenproblematik in der Forschung von Veränderungsmessungen hin. Durch die wiederholte Evaluation mit der gleichen Stichprobe können sowohl Probleme bei der Betreuung dieser Stichprobe (Finanzierung, Drop-Out usw.) als auch die oben bereits genannten Motivationsprobleme auftreten.

Trotz der im Vergleich zu klassischen Testverfahren ökonomischen Anwendbarkeit der CBM kritisieren Fuchs und Fuchs (2000, S. 181) dennoch den mit den CBM verbundenen Zeitaufwand, insbesondere, wenn die Messungen wöchentlich stattfinden und mit jedem Schüler einzeln durchgeführt werden müssen, wie es beim 1-Minute-Lesen der Fall ist (vgl. auch Wesson, King & Deno 1984, S. 47ff.). Sie merken an, dass die Durchführung und Dokumentation bei jedem einzelnen Schüler zwar nur 2,5 Minuten dauert (vgl. Fuchs & Fuchs 2000, S. 181, Fuchs & Fuchs 1999, S. 667), dies jedoch wöchentlich hochgerechnet auf 25 Schüler einen Zeitaufwand von über einer Zeitstunde darstellt. Zukünftig könnte an dieser Stelle ein Computersystem Abhilfe verschaffen (vgl. Fuchs, Fuchs, Hamlett & Hasselbring 1987, S. 14ff.).

Ein weiterer Nachteil sei nach Fuchs und Fuchs (2000, S. 182), dass die CBM auf die Messung von Basisfähigkeiten beschränkt ist und weiterführende Fähigkeiten, wie z. B. das problemlösende Denken im Bereich der Mathematik, bisher weitestgehend nicht gemessen werden können. Dieser Nachteil ist für den Vorschulbereich nicht relevant, da in dem Bereich hauptsächlich Basisfähigkeiten für die Messungen von zentraler Bedeutung sind. Speziell für Schüler aus dem sonderpädagogischen Bereich wäre es unerfreulich, wenn diesen die Möglichkeit der systematisch-formativen Evaluation vorenthalten würde, auf der Grundlage dessen, dass man damit nicht alle Bereiche der Schulbildung evaluieren kann. Walter (2008a, S. 202ff.) merkt an, dass es sinnvoll sei, davon auszugehen, dass der Erwerb der weiterführenden Bildung nur auf der Grundlage der Kulturtechniken möglich sei und diese Schlüsselqualifikationen „eine notwendige, jedoch keineswegs hinreichende Bedingung für Bildung darstellen." (Walter 2008a, S. 210). Fuchs und Fuchs (1999, S. 667) kritisieren auch die mangelhafte Anwendbarkeit für entwicklungsverzögerte Schüler. Bei einer CBM im Bereich Lesen kann nicht nach unten differenziert werden, da die Schüler, die noch nicht lesen können, alle den gleichen Stand haben: Null. Dies wird von den Autoren als

zu späte Identifikation erachtet, sodass sie Messungen für den vorschulischen Bereich fordern, damit diese Kinder bereits frühzeitig erkannt und gefördert werden können.

Außerdem sei darauf hingewiesen, dass eine Lernverlaufsdiagnostik mit entsprechender Interventionsmodifikation, wie sie durch Fuchs (vgl. Kap. III.4.2 zum RTI-Paradigma) vorgeschlagen wird, nicht zwangsläufig auf die Ursache hinweist. Wird eine Intervention verändert und das Kind reagiert darauf positiv und verbessert dadurch seine Leistungen, dann heißt es noch nicht, dass die Intervention selbst zu dem positiven Ergebnis geführt hat. Es ist auch möglich, dass sich das Kind durch einen Förderlehrer besser motivieren lässt oder sich im Allgemeinen die Lehrer-Kind-Interaktion verbessert hat (vgl. Klauer 2006, S. 25). Da kein experimenteller Versuchsplan vorliegt, sollte von der Suche nach kausalen Zusammenhängen beim Einsatz von CBM abgeraten werden. Die CBM ist jedoch ein Mittel, um den Kindern eine individuelle Förderung zukommen zu lassen, was insbesondere durch die Veränderung des deutschen Bildungssystems auf dem Weg zur Inklusion in den Bildungsstätten von Vorteil ist. CBM mag für die Forschung in dem Fall ungeeignet sein, doch für die Lehrperson ist die Praktikabilität des Verfahrens wichtig, indem sie anhand der Lernverläufe überprüfen kann, ob das Kind mit der jeweiligen Interventionsmaßnahme Erfolg hat oder ob diese modifiziert werden muss. Nur durch theoriegeleitetes Handeln und lernstandsspezifische Übungen lassen sich Erfolge bei Kindern mit Schwierigkeiten erzielen: „So gesehen fungiert CBM nur als ein Impulsgeber in einem mehr oder minder qualifiziert stattfindenden Vermittlungsprozess." (Diehl & Hartke 2007, S. 202). Dies konnte auch in einer Studie von Graney und Shinn (2005, S. 184ff.) festgestellt werden, die besagt, dass sich die Leseleistung der Schüler von Regelschullehrern, die nur die Information der Lernfortschritte als Leistungskurve erhalten, nicht verbessert. Es ist nicht ausreichend nur den Lernstand des Kindes zur Verfügung zu haben, diese Information sollte parallel dazu mit einer Interpretationshilfe der Daten sowie Empfehlungen zur Änderung der Intervention gepaart sein.

Müller und Hartmann (2009, S. 41) weisen darauf hin, dass die Ergebnisse von curriculumbasierten Messungen kein „(sonder-)pädagogischer Gewinn per se" seien, sondern diese in ein „flexibles System der evidenzbasierten Diagnostik und Intervention" eingebettet sein müssen. Insbesondere die Ableitung von Interventionsansätzen erfordert ein didaktisches oder auch pädagogisch-therapeutisch tief verankertes Vorwissen der handelnden Pädagogen. Dennoch beurteilen sie die CBM als gut geeignetes Mittel zur „transparenteren und objektiveren Einschätzung von individuellen Lernfortschritten in der Schule" (vgl. ebd.).

Neben den aufgezählten Nachteilen der curriculumbasierten Messung und der damit einhergehenden Interventionsmodifikation, bietet die CBM dennoch ein enormes Potenzial für den Einsatz in der Praxis. Gegenüber herkömmlicher standardisierter Verfahren, versprechen curriculumbasierte Messverfahren auch im vorschulischen Bereich folgende Vorteile (vgl. Madelaine und Wheldall 1999, S. 73ff.; Klauer 2006, S. 22ff.; Walter 2008a, S. 202ff.; Wright o. J., S. 3):

- CBM kann *ökonomisch und kostengünstig* durchgeführt werden.
- Ziel der Messung ist die *Dokumentation des Lernfortschritts* und einer damit einhergehenden angemessenen Förderung des Kindes. Es handelt sich nicht, wie häufig bei klassischen Testverfahren, um ein Verfahren zur ausschließlichen Selektion und Klassifikation.
- Die Messungen orientieren sich an den *Inhalten des Curriculums*, auch dies ist bei herkömmlichen Verfahren häufig nicht gewährleistet (vgl. Fuchs & Deno 1994, S. 16f.).
- Die Daten der CBM sind nützlich für intraindividuelle, aber auch interindividuelle *Vergleiche* mit anderen Kindern, mit der gleichen Altersgruppe oder mit gesamten Kindergärten (vgl. Fuchs & Fuchs 1992, S. 45ff).
- *Störgrößen,* wie z.B. Tagesform und Lautstärke im Raum, haben auf die CBM nicht so starken Einfluss wie auf herkömmliche Testverfahren, die häufig nur einmalig durchgeführt werden. Die häufige Wiederholung der Messungen relativiert diese Messfehler.
- CBM kann die *klassischen Gütekriterien* Objektivität, Reliabilität und Validität erfüllen.
- Durch die *solide Datenbasis* der CBM wird die Kommunikation zwischen Eltern und Lehrpersonal sowie zwischen den Kindertagesstätte und Grundschule erleichtert (vgl. Marston, Diment, Allen & Allen 1992, S. 21ff.).
- CBM ist *änderungssensibel*. Bereits kleine Veränderungen im Lernprozess der Kinder können erfasst und dokumentiert werden. CBM kann durch die engmaschigen Messungen präventiv möglichen Lernstörungen entgegenwirken.
- In traditionellen Testverfahren würden *Testwiederholungen* die Testergebnisse aufgrund des Bekanntheitsgrades der Aufgaben verfälschen. Bei der CBM zeigt ein derartiger Übungserfolg den Lernfortschritt des Kindes auf und ist durchaus erwünscht.
- Traditionelle Tests klassifizieren *oft andere Kinder als Risikokinder* als die CBM. Weist ein Kind vorher bereits niedrige Leistungen auf und diese steigen jedoch zum

Ende der Kindergartenzeit stark an, würde dieses Kind durch herkömmliche Verfahren dennoch als Risikokind klassifiziert werden. Die CBM hätte den Lernfortschritt des Kindes erkannt und es als Nicht-Risikokind eingestuft. Sind die durchschnittlichen Leistungen eines anderen Kindes dagegen zum Ende der Kindergartenzeit gleichbleibend oder sinken, so würde ein klassischer Test dieses Kind als Nicht-Risikokind klassifizieren, während die CBM aufgrund des stagnierenden Lernfortschritts das Kind als Risikokind einordnen würde.

- Mit der CBM können die Kinder innerhalb der Einrichtung *durch ihnen bekannte Erzieher (Vertrauenspersonen) überprüft* werden. Dies ist besonders bei Kindern im Kindergartenalter relevant, da diese in Testsituationen oft noch sehr von ihren Bezugspersonen abhängig sind.

Bezüglich einer parallel stattfindenden Förderung bietet die CBM folgende Vorteile (vgl. Madelaine und Wheldall 1999, S. 73ff.; Klauer 2006, S.22ff.; Walter 2008a, S.202ff.; Wright o. J., S.3):

- Die CBM können auf ineffektive Interventionsmaßnahmen *sofort reagieren*, während traditionelle Testverfahren diese häufig erst am Ende der Fördermaßnahme aufdecken können.
- Traditionelle Testverfahren werden besonders in der sonderpädagogischen Diagnostik nur zur Entscheidungshilfe für die Beschulung oder kostenintensiver Fördermaßnahmen eingesetzt. Sie eignen sich nur bedingt dazu Entscheidungen über Fördermaßnahmen zu treffen. Die CBM dagegen zeigen unmittelbar die *Effektivität einer Interventionsmaßnahme* für ein Kind auf.

CBM scheint auch in Zusammenhang mit dem RTI-Ansatz ein sinnvolles Verfahren zur Evaluation von Unterricht und Intervention zu sein und hat sich insbesondere in den USA als Diagnostikkonzept bewährt. Auch im Schriftspracherwerb zeigte sich bereits in Deutschland, dass Tests, aus denen sich keine Interventionsmaßnahmen ableiten lassen, sondern bei denen man lediglich den Prozentrang oder T-Wert ablesen kann, sich für die Ableitung einer Therapie oder Förderung weniger eignen. Förderungsorientierte Tests erscheinen brauchbarer (vgl. Mannhaupt 2003, S. 106f.). Ziel einer Frühförderung ist es Defizite in den Vorläuferfähigkeiten der Kinder auszugleichen, zudem das Interesse am Lesen und Schreiben zu wecken und die Motivation zum eigenständigen Erlernen zu fördern. Die Sinnhaftigkeit des RTI-Modells lässt sich auch in Bezug auf die bildungspolitischen Veränderungen in Deutschland (vgl. Kap. I.1.1 Veränderungen im Bildungssystem) und den zentralen Diskurs zur Inklusion rechtfertigen. Inklusion fordert durch die Vielfalt der

Schülervoraussetzungen und damit heterogenen Lern- und Lehrbedingungen die Schüler frühzeitig zu erfassen, um die individuellen Lernbedürfnisse der Schülerschaft zu erkennen und Interventionen zeitnah anpassen zu können (vgl. Liebers & Seifert 2012, S. 1ff.).

Die empirischen Studien in den USA und die Forschungen in Deutschland weisen darauf hin, dass sich die curriculumbasierte Messung speziell auch für den Einsatz im Vorschulbereich zur Früherkennung von Lese- und Rechtschreibschwierigkeiten eignet. Dadurch, dass das Bildungssystem in den USA anders aufgebaut ist, scheint die Eins-zu-Eins-Übernahme der CBM-Daten aus US-amerikanischen Kindergärten (z.B. eine Messung der Buchstaben-Laut-Korrespondenzen im Kindergarten, vgl. Fuchs & Fuchs 2004, S. 32ff.) aufgrund der differierenden Curricula und der unterschiedlichen Sprachen als nicht sinnvoll. Es müssen anhand der empirischen Grundlage im deutschsprachigen Bereich Aufgabenfelder gefunden werden, die für unsere Kindergartenkinder zu bewältigen sind und inhaltlich bereits empirisch erforscht wurden. Die Kombination der deutschen Erkenntnisse zur phonologischen Informationsverarbeitung und der US-amerikanischen Erkenntnisse zu der Methodik der CBM lassen vermuten, dass das Zusammenführen der Methodik und des Inhalts im deutschen Vorschulbereich von Vorteil sein kann. Somit wird in der hier vorgestellten Studie die Messung des Konstrukts der phonologischen Informationsverarbeitung inhaltlich konstant gehalten, während die Messmethode durch die Nutzung der Methode der CBM eine Variation erfährt.

In der vorliegenden Studie wird anhand eines eigens erstellten Früherkennungsverfahrens untersucht, inwieweit sich die curriculumbasierte Messung als Screening-Verfahren für Fähigkeiten in der phonologischen Informationsverarbeitung eignet und ob sich ein robuster Indikator für die Prädiktoren des Lesens und Schreibens ergibt. Dazu wurde das Messverfahren „Phonolino“ konstruiert, das erstmalig die Komponenten der phonologischen Informationsverarbeitung mit dem Verfahren der curriculumbasierten Messung kombiniert. Um dieses Verfahren evaluieren zu können, werden im folgenden Kapitel Erkenntnisse zur Konzeption sowie zur Evaluation curriculumbasierter Messungen zusammengetragen, die als Fundament für die Konzeption und Evaluation von Phonolino dienen.

6 Erkenntnisse zur Konzeption und Evaluation von CBM

Bei der Lernverlaufsdiagnostik handelt es sich um eine Form von Diagnostikverfahren, die das Lernen diagnostizieren. Wilbert (2014, S. 282) beschreibt dazu:

> *„Um Lernen zu diagnostizieren, bedarf es daher einer wiederholten Messung in mindestens zwei Situationen mit*

> *gleichem Aufforderungscharakter. Im Fokus einer Diagnostik des Lernens steht also die Veränderung, die einen Hinweis darauf geben soll, in welchem Maß eine Person von dem, was zwischen den Messungen stattgefunden hat (z.B. einer Intervention), profitierte."*

Die von Lynn S. Fuchs (2004, S. 189) bereits oben (vgl. Kapitel III.3, S. 217) beschriebenen drei wichtigen Stadien, sollten in zukünftigen Untersuchungen zu curriculumbasierten Messverfahren Berücksichtigung finden (wie bspw. in der Dissertation von Voß 2013, S. 196 bereits umgesetzt):

1) Technical features of the static score
2) Technical features of slope
3) Instructional utility

Unter der ersten Stufe *„technical features of the static score"* fasst L. S. Fuchs (vgl. ebd.) die Überprüfung der allgemeingültigen Testgütekriterien der Objektivität, Reliabilität und Validität zusammen. Die Objektivität der curriculumbasierten Messung soll sicherstellen, dass die Versuchsergebnisse unabhängig von den Randbedingungen, wie den zeitlich-räumlichen Voraussetzungen oder den Untersuchungsleiter, sind. Die Reliabilität der curriculumbasierten Messung gewährleistet, dass die Ergebnisse der Messung nicht durch Messfehler entstanden und die Ergebnisse unter gleichen Bedingungen replizierbar sind. Zusätzlich sollen die Verfahren sich als valide erweisen, und damit die zu messende Größe repräsentieren (vgl. bspw. Klauer 2006, S. 22f.). Ein entscheidender Punkt ist dabei die Schwierigkeit der einzelnen Messungen, die zum einen für eine Lernverlaufsmessung dasselbe erfassen und gleichzeitig gleich schwer sein müssen (vgl. Klauer 2014a, S. 9). Hierfür erprobten Strathmann und Greisbach (2010, zit. nach Klauer 2014a, S. 10) anhand eines Zufallsgenerators Stichproben aus 20 Wörtern zu ziehen. Es stellte sich heraus, dass die Schwierigkeit der Items bei jeder Messung unterschiedlich war, sodass der Schwierigkeit der Messungen eine differenzierte Betrachtung folgen muss. Eine testtheoretische Problematik besteht darin, die homogene Schwierigkeit der Messungen festzustellen, da bei einer Lernverlaufsdiagnostik davon ausgegangen wird, dass sich der „subjektive" Schwierigkeitsgrad für die Probanden verändern muss, weil jede nachfolgende Messung ihnen aufgrund ihrer Lernentwicklung leichter fallen müsste. Aus diesem Grund greifen Strathmann und Klauer (2008, zit. in Klauer 2014a, S. 10) darauf zurück nur nebeneinanderliegende Messungen auf homogene Schwierigkeit zu testen.

Wird die erste Stufe durch das erstellte Verfahren erfüllt, so lässt sich dieses in Bezug auf die klassischen Testgütekriterien als objektives, reliables und valides Screening-Verfahren einsetzen, um die Leistungen der Kinder zu einem festgelegten Zeitpunkt (Statusdiagnostik) zu erfassen.

Sind die Kriterien der ersten Stufe erfüllt, fordert L. S. Fuchs (vgl. 2004, S. 189) auf der zweiten Stufe *„technical features of slope“* eine Überprüfung der CBM in Bezug auf ein sensibles und reliables Verfahren im Sinne einer Lernfortschrittsmessung. Bei einer Lernverlaufsdiagnostik über eine längere Zeit lassen sich Rückschlüsse auf die Lernentwicklung der Probanden schließen. Diese weiteren Analysen sind vonnöten, um aus dem gemessenen Lernverlauf der Kinder Ableitungen für die Förderung zu ziehen. Wilbert (2014, S. 284) beschreibt, dass Verfahren, die es zum Ziel haben Entwicklungsschritte nachzuvollziehen, in höherem Maß Entwicklungen und Veränderungen abbilden müssen, als statusdiagnostische Verfahren (Veränderungssensitivität).

Dies ist nach Klauer (2011, S. 212) bspw. über einen quasi-experimentellen Versuchsplan mit Experimental- und Kontrollgruppe möglich, wobei die Kontrollgruppe auf der Grundlage der Ergebnisse der curriculumbasierten Messung einem Treatment unterzogen wird, das die Kontrollgruppe nicht erhält. Durch ein Prätest-Posttest-Design wäre es damit möglich die erlangten Unterschiede anhand der Mittelwerte zu bestimmen, die Hinweise auf die Änderungssensibilität vermuten lassen. Er kritisiert an der Vorgehensweise ebenfalls:

> *„Vielleicht ist diese Art des Vorgehens dann doch zu unspezifisch, um die Änderungssensibilität zu belegen.“ (Klauer 2011, S. 213)*

Zudem beschreibt Klauer (2011, S. 217), dass der experimentelle Nachweis der Änderungssensibilität noch eine Aufgabe der Zukunft sei. Er schildert die Schwierigkeiten nach Tack 1986 (vgl. a. a. O.): Die klassische Testtheorie scheint nicht geeignet zu sein, Änderungen zu erfassen, da sich einerseits ein Regressionseffekt nicht ausschließen und sich andererseits das Reliabilitäts-Validitäts-Dilemma schwer überwinden lässt.
Der Regressionseffekt lässt sich dadurch beschreiben, dass die Probanden mit stark nach oben oder auch nach unten ausgeprägten Werten bei einer Wiederholungsmessung Werte aufweisen, die zum Mittelwert tendieren. Dies lässt sich durch die Verteilung der Messfehler erklären, sodass die Probanden mit einem hohen Messfehler bei einer Wiederholungsmessung sehr wahrscheinlich einen kleineren Messfehler aufzeigen. Bei der Wiederholung der Messungen über einen längeren Zeitraum ließe sich dieser Regressionseffekt ausgleichen (vgl. ebd., S. 218). Somit sind die einzelnen Messwerte variabler, die langfristige Tendenz zeigt jedoch reelle Werte.

Das Reliabilitäts-Validitäts-Dilemma in Bezug auf eine Lernverlaufsdiagnostik beschreibt, dass die „erlernte“ Veränderung über den gemessenen Zeitraum sich schwer mit geforderten hohen Retest- bzw. Paralleltestreliabilitäten vereinbaren lässt. Klauer (vgl. a. a. O.) schließt daraus, dass änderungssensible Verfahren nur mäßige Retestreliabilitäten aufweisen können. Hinzu kommt, dass durch die Wiederholung der Messungen nicht garantiert ist, dass der Test weiterhin das misst, was er messen soll, und ein valides Messverfahren darstellt.
Somit sagt Klauer folgende Punkte für änderungssensible Verfahren voraus (vgl. a. a. O.):
Änderungssensible Tests weisen ...

1. nur mittelhohe Retest- oder Paralleltestreliabilitäten auf,
2. kleinere Korrelationen zwischen den Messzeitpunkten auf, je weiter diese auseinander liegen. Je länger ein erfasster Lernzeitraum ist, umso niedriger fallen die Korrelationen zwischen den Ergebnissen aus,
3. im Split-Half-Verfahren höhere Reliabilitäten als bei Retest- oder Paralleltestreliabilitäten auf.

Diese von Klauer postulierten Hinweise zur Überprüfung änderungssensibler Verfahren bevor eine weitere experimentelle Studie im Zwei-Gruppen-Design veranlasst werden kann, stellen eine angemessene Grundlage zur ersten Analyse eines Messverfahrens zur Änderungssensibilität dar. Zu ergänzen ist, dass er sich hierbei auf die klassischen eindimensional konstruierten Tests bezieht (z. B. Additionsaufgaben im ZR bis 100).

Walter (2010b, S. 164) konstatiert mindestens drei weitere Voraussetzungen:

- Das Messinstrument sollte Intervallskalen-Niveau besitzen, also gleiche Abstände zwischen den unterschiedlichen Skalenwerten (Merkmalsausprägungen) ausweisen. Es sollte unabhängig von Boden- und Deckeneffekten sein.
- Das Messverfahren muss über die Zeit hinweg das gleiche Merkmal messen und den gleichen Schwierigkeitsgrad aufweisen.
- Es sollten ausreichend Paralleltests mit einer guten Paralleltestreliabilität zur Verfügung stehen, die sowohl durch hohe mittlere Korrelationen als auch über gute Interkorrelationen nachgewiesen werden sollte.

Zusätzlich empfiehlt Walter (2010b, S. 164) einen Kennwert für den Lernzuwachs zu errechnen, der den Lernzuwachs des einzelnen Kindes oder der gesamten Gruppe beschreiben kann und bezieht sich dabei auf die Arbeiten von Fuchs, Fuchs, Hamlett, Walz und Germann (1993), Deno, Fuchs, Marston und Shin (2001) sowie Silberglitt und Hintze (2007) aus dem US-amerikanischen Raum.

Wilbert beschreibt (2014, S. 284f.), dass sich die Erfassung von Veränderungen rein mathematisch auf der Grundlage von Differenzen bildet, dessen Interpretation Schwierigkeiten mit sich bringt. Dies wird besonders durch das sogenannte „Skalierungsproblem“ dargestellt. Es entsteht dadurch, dass sich das Ergebnis einer Messung in verschiedener Form transformieren und damit auch interpretieren lässt. Bei einer Verbesserung von 5 auf 10 korrekte Antworten lässt sich nicht automatisch auf eine verbesserte Kompetenz des Schülers schließen, da diese auch von der Aufgabenschwierigkeit sowie von der Anzahl der Aufgaben abhängig ist. Wilbert beschreibt eindrücklich an verschiedenen Rechnungen, dass sich die Kompetenz des Schülers weder durch einen Vergleich anhand der Rohwerte oder anhand der normgruppenbasierten Interpretation, noch anhand der relativierten Steigerung gut interpretieren lässt. Wilbert folgert (vgl. a. a. O., S. 285), dass eine Messung im engeren Sinne vorliegen muss, dessen Items in gleichem Maß ein Merkmal abbilden. Bei einer Statusdiagnostik durch eine einmalige Messung wäre ein Vergleich mit der Norm möglich, so dass die Problematik der Differenzen gar nicht so sehr ins Gewicht fallen würde. Bei einer Veränderungsmessung ist ein Vergleich der Werte bei nicht vorliegendem gleichen Maß der Items nicht möglich.

Auf der dritten Stufe der *„Instructional utility”* fordert L. S. Fuchs (vgl. 2004, S. 189), dass das erstellte Verfahren in Bezug auf die Tauglichkeit in der praktischen Anwendung untersucht werden sollte. Sie konstatiert, dass Studien diese letzte Stufe zwingend berücksichtigen sollten, damit in Bezug auf die Informationen aus der curriculumbasierten Messung über die Leistungen der Kinder Lehrentscheidungen getroffen werden können.
In der hier vorliegenden Untersuchung wird die Entstehung der curriculumbasierten Messung „Phonolino“ als Screeningverfahren im letzten Kindergartenjahr zur Früherkennung von Lese- und Rechtschreibschwierigkeiten im ersten Schuljahr in Bezug auf einen Teil der Stufen von Fuchs (2004) sowie der Erkenntnisse zur Änderungssensibilität nach Klauer (2011) und Walter (2010b) evaluiert. Das Verfahren wird anhand der Gütekriterien der klassischen Testtheorie (Stufe 1 nach Fuchs, 2004) beleuchtet. Zudem werden Aspekte der Änderungssensibilität (Stufe 2 nach Fuchs, 2004) betrachtet, die in späteren Untersuchungen bspw. durch eine weitere experimentelle Studie im Zwei-Gruppen-Design oder durch die Berechnung eines Indexes für Lernzuwachs noch erweitert werden müssen. Die Stufe 3 in Bezug auf die Praktikabilität kann aufgrund des Ausmaßes der vorliegenden Untersuchung gleichfalls nur im Sinne des subjektiven Augenscheins bewertet werden und bedarf einer weitergehenden Betrachtung in Folge-Studien.

Im Folgenden wird zunächst die Konzeption der curriculumbasierten Messung Phonolino beruhend auf in den vorhergehenden Kapiteln ausführlich dargelegten Erkenntnissen

erläutert. Im Anschluss folgt die empirische Beschreibung, Ausführung und Auswertung der durchgeführten Studie mit Bezugnahme auf die hier vorgestellten Kriterien für curriculumbasierte Messverfahren zur Dokumentation von Lernfortschritten im Bereich der phonologischen Informationsverarbeitung im letzten Kindergartenjahr zur Prävention und Früherkennung von Schwierigkeiten beim Schriftspracherwerb.

IV. Curriculumbasierte Messung „Phonolino“ für die Überprüfung der phonologischen Informationsverarbeitung im letzten Kindergartenjahr

In der vorliegenden Untersuchung wird eine eigens konzipierte curriculumbasierte Messung „Phonolino“ validiert. Im Folgenden werden nach einer kurzen theoretischen Einführung in die notwendigen Grundlagen zur Konzeption einer CBM zur phonologischen Informationsverarbeitung im letzten Kindergartenjahr, die Zielsetzung, die Vorgehensweise bei der Konzeption, sowie Aufbau und Durchführung von Phonolino vorgestellt. Phonolino kombiniert wissenschaftliche Erkenntnisse zur phonologischen Informationsverarbeitung aus dem deutschsprachigen Raum mit Forschungsergebnissen zur Methodik des curriculumbasierten Messens aus den USA in einem Messverfahren, das in dieser Untersuchung hinsichtlich angemessener Gütekriterien sowie seiner prognostischen Eignung validiert werden soll. Die ausführliche Überprüfung der CBM im Sinne einer Veränderungsmessung sowie das Abbilden von individuellen Lernverläufen von Kindern als Grundlage für eine individuelle Förderung sollen in Folgeuntersuchungen im Fokus stehen.

1 Allgemeine Hinweise zur Konzeption und Evaluation von CBM für die Messung phonologischer Informationsverarbeitung

Aus den in Kap. II beschriebenen Studien wird deutlich, dass die phonologische Informationsverarbeitung ein wichtiger Indikator für den Erfolg im Schriftspracherwerb darstellt. Für eine vorschulische Prävention ist es nötig, möglichst frühzeitig Schwierigkeiten in den Vorläuferfertigkeiten aufzuzeigen. Die für diese Studie konzipierte curriculumbasierte Messung „Phonolino“ soll die phonologische Informationsverarbeitung valide und reliabel messen, sowie die schriftsprachlichen Fähigkeiten der Schülerinnen und Schüler ausreichend gut prognostizieren. Hierbei werden die Fähigkeiten der Kinder nicht einmalig erhoben, stattdessen wird die Lernentwicklung der Kinder anhand mehrerer Messzeitpunkte im Verlauf des letzten Kindergartenjahrs beschrieben. Hierzu bedarf es einer ausreichenden Anzahl an Parallelversionen der CBM. Der Anspruch an die Items der CBM ist, dass sie die Breite der Leistungen der Kinder möglichst abdecken und diese sensibel messen. Bei der Konzeption der einzelnen Aufgaben ist zu beachten, dass die Anweisungen für die Erzieher sowie die Instruktionen und die Aufgabenstellung für die Kinder einfach zu verstehen und unmissverständlich auszuführen sind. Um die Objektivität zu gewährleisten müssen Aufgabenstellung und Instruktionsanweisung, sowie das Aufgabenmaterial in allen teilnehmenden Kindergärten, gleich sein. Bezüglich des Inhalts muss sichergestellt werden, dass die Teilbereiche der phonologischen Informationsverarbeitung auf der Grundlage der in Kap. II dargestellten Erkenntnisse theoretisch fundiert abgebildet werden. Zunächst wird

erläutert, aufgrund welcher theoretischen Fundierung die curriculumbasierte Messung „Phonolino“ für den Einsatz im letzten Kindergartenjahr entstanden ist und wie im Detail die einzelnen Items konzipiert wurden.

2 Erhebungsmaterial „Phonolino“

Phonolino besteht aus neun curriculumbasierten Messungen, die im letzten Kindergartenjahr von Erziehern der jeweiligen Einrichtung durchgeführt werden. Die Messungen dauern mit Auf- und Abbau usw. ca. 5 Minuten und sind als Einzelmessung konzipiert. In den folgenden Abschnitten werden die Zielsetzung, sowie die Ergebnisse einer durchgeführten Erhebung vor der eigentlichen Hauptuntersuchung der ersten Version von Phonolino mit etwaigen Anpassungen der Items, der Aufgabe der CBM mit Rückbezug auf die im theoretischen Teil zugrunde gelegten Kriterien für die Erstellung einer CBM und die Erstellung der Parallelmessungen beschrieben.

2.1 Zielsetzung von Phonolino

Zur systematischen Prävention von Lese- und Rechtschreibschwierigkeiten im Vorschulalter wird ein Verfahren benötigt, das Risikokinder im Vergleich zu existierenden Diagnostik-instrumenten zuverlässig identifiziert. Ziel ist es mit diesem Verfahren möglichst hohe Klassifikationsstabilitäten zu erreichen. In dieser Untersuchung stellt sich die Frage, ob die curriculumbasierte Messung Phonolino relevante Voraussagen über die Lese- und Rechtschreibfähigkeiten der Kinder beim Schriftspracherwerb treffen kann und ob sich das Verfahren der curriculumbasierten Messung als Vorteil erweist, oder ob sich das Prognosemodell noch durch weitere Komponenten erweitern lässt.

Phonolino setzt nach dem RTI-Modell an der ersten Stufe an, um als Progress-Monitoring die Entwicklung der Kinder im Kindergarten zu dokumentieren (vgl. Abb. 44).

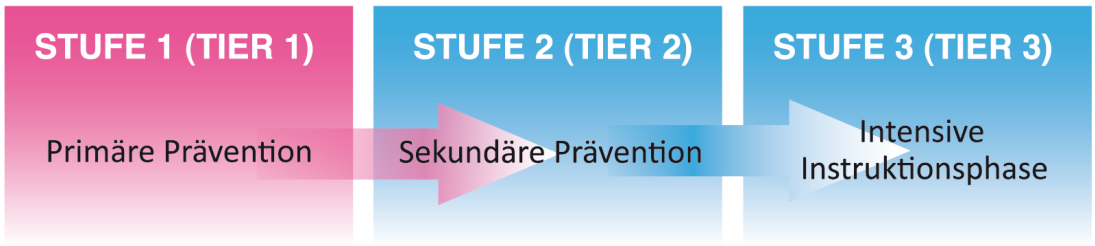


Abb. 44: Einordnung von Phonolino in das RTI-Modell

Phonolino soll ein für Erzieherinnen und Erzieher einfach handhabbares Messverfahren sein, das in der Praxis ökonomisch, mit wenig Material- und Zeitaufwand eingesetzt und ausgewertet werden kann (vgl. auch Stufe 3 nach L. S. Fuchs 2004, S. 189). Ziel ist es, die Risikokinder aufgrund der nah aufeinanderfolgenden Messungen frühzeitig zu identifizieren und Schwierigkeiten beim Schriftspracherwerb präventiv entgegen zu wirken. Die dokumentierten Daten zur phonologischen Informationsverarbeitung können beim Wechsel in die Grundschule eine wichtige Grundlage für nachfolgende Interventionen der Grundschullehrerinnen und -lehrer sein. Sie bieten ihnen eine solide Basis, anhand der sie ihren Unterricht oder eine individuelle Förderung aufbauen können.

In den USA konnte sich die CBM bereits als sinnvolles, standardisiertes Messverfahren auch im Vorschulbereich beweisen (vgl. Fuchs 2004, S. 192). Dies ist mit der Besonderheit des Schwerpunktes auf der phonologischen Informationsverarbeitung als wichtigem Vorhersagewert für die Lese- und Schreibkompetenz in der Schule in Deutschland ebenfalls mit den vorgelegten Messungen zu überprüfen.

Die Zielgruppe der Messung sind Vorschulkinder im Alter von ca. fünf bis sieben Jahren. Phonolino besteht aus neun einzelnen Messungen, die im letzten Kindergartenjahr einmal monatlich (mit Ausnahme der Ferienzeiten) stattfinden. Die erste Messung musste jedoch nachträglich entfernt werden, da es sich nur für wenige Kindergärten einrichten ließ, diese zu dem geforderten Zeitpunkt bereits durchzuführen. Somit wird Phonolino mit acht Messzeitpunkten evaluiert.

Einerseits ist das Ziel der Validierung mit Phonolino ein Screening-Verfahren zu finden, das zur Identifikation von Risikokindern bzw. sogenannten „Nonrespondern“ beiträgt und das prognostiziert, wie sich die schriftsprachlichen Fähigkeiten des Kindes in der Schule ohne Förderung entwickeln könnten. Durch mehrere innerhalb des letzten Kindergartenjahres aufeinanderfolgende Messungen könnten so Risikokinder frühzeitig identifiziert und Schwierigkeiten beim Schriftspracherwerb präventiv entgegengewirkt werden. Andererseits dient Phonolino auch als Dokumentation der phonologischen Fertigkeiten des Kindes, dessen Entwicklungsverlauf dann auch an die pädagogischen Fachkräfte weitergeleitet werden kann, um eine nahtlose Förderung und auch Forderung zu gewährleisten.

Parallel zu den diagnostischen Anforderungen sollte Phonolino auch praktischen Anforderungen genügen (vgl. Stufe 3 nach Fuchs, Kap. III.6). Es soll ein für Erzieherinnen und Erzieher handhabbares Messverfahren darstellen, das im Praxiseinsatz ökonomisch durchgeführt und ausgewertet werden kann. Auch motivatorische Komponenten sollen bei

der im Folgenden beschriebenen Konzeption Berücksichtigung finden. Dazu wird das Schaf „Phonolino“ als vertraute Instanz die Durchführungsanweisungen geben und an allen Messungen teilnehmen.

Im nächsten Abschnitt wird das Vorgehen bei der Konzeption von Phonolino in Bezugnahme auf einen durchgeführten Vortest beschrieben.

2.2 Vortest zu Phonolino

Am 15. und 17. Juli 2008 wurde in einem Kindergarten in Flensburg mit 25 Vorschulkindern kurz vor Eintritt in die Schule ein Vortest mit Phonolino durchgeführt. Die Kinder wurden einmalig mit einer Phonolino-Messung überprüft. Die erste Version von Phonolino umfasste fünf Items (vgl. Tab. 26). Die theoretische Fundierung zur Auswahl der Aufgaben ist unter „Konzeption der Items von Phonolino“ (vgl. i. F. Kap. IV.2.5) nachzulesen.

Zielsetzung des Vortests war keine vorrangig quantitative Auswertung, sondern eine Testung der Praktikabilität im Umgang mit Material, Protokollführung und Auswertung während der Messung, sowie ein Überblick zur Komplexität der Aufgaben und der Dauer sowie der motivatorischen Wirkung auf die Kinder. Nach der Auswertung wurden die Items von Phonolino nochmals angepasst.

Tab. 27: Items der Vorversion von Phonolino, Juli 2008

Items des Vortests	Inhalt
Pseudowörter nachsprechen	Kinder sprechen Nonsenswörter nach, die ihnen von einer CD vorgespielt werden.
Lautsynthese	Kinder synthetisieren Laute zu sinnvollen Wörtern (Bsp.: F-isch), Laute werden von einer CD präsentiert.
Schnelles Benennen von Farben	Kinder sollen Farben von Tieren in möglichst kurzer Zeit benennen.
Silben klopfen	Kinder sollen Wörter nachsprechen, sie anhand ihrer Silben zerlegen und dabei klopfen.
Lautisolation aus Wörtern	Kinder sollen erkennen, an welcher Position des Wortes (vorne, Mitte, hinten) sich der angebotene Laut befindet (Bsp.: „Wo hörst du ein „k“ in Kiste?“).

Folgende Erkenntnisse konnten aus dem Vortest gewonnen werden:

Anzahl der Items

Es zeigte sich, dass sich die Messung auf ca. 15 Minuten erstreckte, was sich als unökonomisch in der Durchführung im Kindergarten erwies. Daraufhin wurde anhand theoretischer Relevanz geprüft, welche Items in der Theorie besonders aussagekräftig bzgl. der Identifikation von Lese-Rechtschreibschwierigkeiten sind und welche Items ggf. entfernt werden können. Der Zeitverlust vor Beginn, zwischen und nach Ende der Messung wurde unterschätzt, sodass die Messung von fünf auf drei Items reduziert wurde (vgl. Tab. 28).

Tab. 28: Verminderung der Items als Resultat des Vortests

Items des Vortests	Inhalt
Pseudowörter nachsprechen	Kinder sprechen Nonsenswörter nach, die ihnen von einer CD vorgespielt werden.
Silben klopfen	Kinder sollen Wörter nachsprechen, sie anhand ihrer Silben zerlegen und dabei klopfen.
Lautisolation aus Wörtern	Kinder sollen erkennen, an welcher Position des Wortes (vorne, Mitte, hinten) sich der angebotene Laut befindet (Bsp.: „Wo hörst du ein „k“ in Kiste?“).

Aufgrund des zu großen Zeitverlusts während der Messung mussten somit zwei der Items aus den Phonolino-Messungen entfernt werden. Auf Basis von wissenschaftlichen Erkenntnissen (vgl. Kap. II.1.2) erklärt die phonologische Bewusstheit 71 % der Varianz der Rechtschreibleistung, während dem phonologischen Arbeitsgedächtnis eine geringere Varianz zugemessen wird. Außerdem wurde die Submessung „schnelles Benennen“ wegen der nicht eindeutig geklärten Wirkung (vgl. Kap. II.1.2.2) entfernt. Es ist in der Forschung noch sehr umstritten, ob diese Fähigkeit einen spezifischen Zusammenhang zur Lese-Rechtschreibschwäche markiert oder es sich doch um einen allgemeinen Geschwindigkeitsfaktor handelt, der sich nur unspezifisch auswirkt. Aufgrund der Ökonomie konnte dieser Teilaspekt also vernachlässigt werden. Das Item zur Lautsynthese wurde ebenfalls entfernt, da sich herausstellte, dass 24 von 25 Kindern (die sich bereits kurz vor der Einschulung befanden) keine Fehler in der vorgegebenen Zeit unterliefen und es sein könnte, dass Deckeneffekte auftreten würden.

Methodik und Praktikabilität

Es war vorgesehen, die Messungen sowohl mit einer Handpuppe als auch mit zusätzlichem Material (Tiere vom Bauernhof, ...) zu begleiten, um die Motivation der Kinder zu erhalten und die Angst vor einer Messung zu verringern. Dieser hohe Materialeinsatz erwies sich jedoch für die durchführende Person als nicht praktikabel, da sie parallel den CD-Player, die Dokumentation auf Papier und die Handpuppe nutzen musste. Somit wurde das Zusatzmaterial entfernt und als aktivierende Elemente Bildkarten mit Fotos der jeweiligen Bauernhoftiere eingebunden.

Inhaltliche Veränderung des Items zur Lautisolation

Bei der Durchführung des Items zur Lautisolation ergab sich, dass keines der 25 Kinder dazu in der Lage war, eine mittlere Position des gesuchten Lautes zu identifizieren. Die Probanden des Vortests standen kurz vor Schuleintritt, während die Kinder, die für die Hauptuntersuchung vorgesehen sind, mit ihrem letzten Kindergartenjahr erst starten. Dadurch ist zu vermuten, dass das Item zu komplex ist und keine ausreichende Differenzierung nach unten bietet. Daraufhin wurde die Erkennung der Mittellaute aus Item 3 entfernt.

Die Kinder des Vortests nahmen nicht an der Phonolino-Evaluation teil, die im Folgenden beschrieben wird. Für den Vortest wurden gezielt Kinder gewählt, die im Sommer 2008 die Schule besuchen würden und damit nicht mehr an den folgenden Phonolino-Messungen des Kindergartens teilnahmen. Ein Unterschied war zudem, dass die curriculumbasierte Messung „Phonolino“ nur einmalig durchgeführt wurde. Durchführende Person der Messungen war die Verfasserin dieser Arbeit anstelle der Erzieherinnen des Kindergartens, sodass eine längere Einführungs- und Kennenlernphase bei jedem Kind stattgefunden hat. Die CBM dauerte im Schnitt pro Kind 15 Minuten (mit Abzug der Kennenlern- und Verabschiedungsphase am Anfang und Ende der Messung, jedoch inklusive der „Umbauarbeiten“ der Materialien während der Messung, da eine möglichst realistische Zeitsituation ausprobiert werden sollte, die später auch von den Erziehern durchgeführt würde).

Der detaillierte Aufbau von Phonolino wird im folgenden Abschnitt erläutert.

2.3 Aufbau von Phonolino

Phonologische Informationsverarbeitung bietet, wie bereits in Kapitel II.1.2 beschrieben, ein spezifisches Konstrukt, auf dessen Basis die Prävention und Intervention bei Schwierigkeiten

im Lesen und Schreiben verläuft. Einmal pro Monat wird eine Phonolino-Messung in ca. fünf bis sieben Minuten (inkl. Aufbau und Umschalten des CD-Players) durchgeführt. Phonolino kann ausschließlich in Einzelmessungen durchgeführt werden. Die Antworten der Kinder werden in einem Heft protokolliert.

Jede Phonolino-Messung hat drei Items, die Teilkomponenten der phonologischen Informationsverarbeitung messen (vgl. Tab. 29). Die Items werden jedes Mal in der gleichen Reihenfolge präsentiert. Sie werden über eine beigelieferte CD abgespielt, damit für alle Kinder eine gleiche Aussprache der dargebotenen Wörter gewährleistet ist.

Tab. 29: Items bei Phonolino

Item	Inhalt	Anzahl möglicher „Wörter“	Rahmenhandlung
1	Pseudowörter erkennen	12	Pferd Helge
2	Silben klopfen	16	Hühner auf dem Hof
3	Lautisolation	12	Kuh Klara

Das Schaf Phonolino in Form einer Handpuppe begleitet die Kinder durch die Messungen.

Item 1 – Pseudowörter erkennen: Die Kinder besuchen Phonolinos Freund Helge, das Pferd. Dieser Freund hat sich eine Spaßsprache ausgedacht und möchte sie den Kindern vorsprechen. Die Pseudowörter auf der CD sollen von den Kindern nachgesprochen werden.

Item 2 – Silben klopfen: Die Kinder lernen die Hühner auf dem Hof kennen, die immer die Anzahl der Silben eines Wortes auf den Boden klopfen. Die Wörter auf der CD sollen von den Kindern in Silben untergliedert werden. Sie sollen die Wörter nachsprechen und dabei bei jeder Silbe mit dem Finger auf dem Tisch „picken“.

Item 3 – Lautisolation: Die Kinder lernen Phonolinos Freundin Klara, die Kuh, kennen. Sie spielt mit ihnen ein Suchspiel – sie versteckt Laute in Wörtern und die Kinder sollen herausfinden, ob sich der Laut am Anfang oder am Ende des Wortes versteckt (vgl. Phonolino-Handbuch, Anhang 1).

Die durchführende Person sollte möglichst eine dem Kind bekannte Person sein, z. B. die Erzieherin aus der Kindertagesstätte. Sie sollte die Instruktionen vor der ersten

Durchführung mehrmals laut durchlesen oder auswendig lernen, um einen reibungslosen Ablauf zu gewährleisten (vgl. Abb. 45). Die Messung wird ausschließlich in der Eins-zu-Eins-Situation durchgeführt. Um die Anspannung einer Testsituation zu reduzieren, sind die Messungen in immer wiederkehrende Rahmenhandlungen eingebettet (vgl. Phonolino-Handbuch, Anhang 1).

Lautisolation aus Wörtern – (im Kuhstall)

Da ist meine Freundin Klara Kuh (Foto). Klara kann schon lesen und schreiben *(schnell nicken). S*ie spielt mit mir auf dem Bauernhof immer ein Suchspiel. In den Wörtern sind die Laute versteckt. Das spielen wir jetzt auch mal, ja?

Also, hör dir mal die erste Suchfrage an (erste Frage auf der CD abspielen).

Bei richtiger Antwort: Oh toll, stimmt ja, jetzt verstehe ich das.
Bei falscher Antwort: Hm, hör nochmal genau hin, ich glaube in ... (gedehnt sprechen) hört man das /…/ (Laut aus der Frage einsetzen) am Anfang. Was sagst du …. (Name der Erzieherin)? (nickt)

Bei den folgenden Wörtern wird nicht mehr korrigiert.

Jetzt hör mal, welche Wörter Klara Kuh auf die CD gesprochen hat. Immer wenn sie ein Wort gesagt hat, sagst du, ob du den Laut am „Anfang“ oder am „Ende“ gehört hast. Mal sehen, wie gut du hinhören kannst und wie schnell du dabei bist.
(Zeit: 1 Minute)

(Nach jedem Wort wird die Pausentaste gedrückt und gewartet, bis das Kind antwortet, dann wird das nächste Wort abgespielt, währenddessen wird protokolliert.)

Protokollierung:
Nur die erste Antwort, die das Kind gibt, wird in der mittleren Spalte protokolliert. Verbessert sich das Kind nochmals wird dies in der zweiten Spalte notiert. Sollte das Kind etwas anderes als „Anfang“ oder „Ende“ sagen, dann dies als Notiz am Rand vermerken.

Abb. 45: Beispiel der Durchführungsanweisung für Item 3, die Erzieherin spricht dabei mit der Phonolino-Handpuppe (vgl. Phonolino-Handbuch, Anhang 1)

Die durchschnittliche Untersuchungsdauer beträgt fünf bis sieben Minuten. Die Netto-Zeit der Messung beträgt drei Minuten. Die Items sollen nicht bis zum Ende durchgeführt werden, sondern werden nach einer Minute abgebrochen, Item 2 nach 45 Sekunden.
Das Kind darf jedoch die letzte Antwort aus motivationalen Gründen noch zu Ende führen. Diese wird aber nicht mehr ausgewertet.

Die Durchführung erfolgt entlang festgelegter Durchführungsanweisungen, die für die durchführende Person auf Karten gedruckt worden sind. Die Anweisungen sollen wortwörtlich wiedergegeben werden. Die jeweiligen Items werden für die Einhaltung der Objektivität von einer CD abgespielt. Die Antworten der Kinder vermerkt der Testleiter in einem Protokollheft (vgl. Abb. 46, Abb. 47 und Anhang 2). Auch die Protokollierung sowie der Umgang mit Unterbrechungen ist im Phonolino-Handbuch für die Erzieher, die die Messung durchführen, hinterlegt (vgl. Anhang 1).

CBM 3 (1.12.2008 – 7.12.2008)

Datum: _____ Uhrzeit: _____

Besondere Vorkommnisse (Konzentration, Störungen während der Messung usw.):

_____

_____

_____

_____

1. Pseudowörter

	Wort	r/f
Übung	ta - mi - sa - lo	
1	ad - mi - so - pas -ti	
2	tu - ri - lo - ma	
3	a - bo - sas - pa - ha - me	
4	a - bi - bi - se - lum	
5	tak - ka - ki - lo	
6	krun - lu - ma - pa - ta - wi	
7	mi - du - pi - ra - so	
8	ro - pam - li - do	
9	pim - bam - pu - la - he - su	
10	pam - pos - sa - me - nuk	
11	ru - li - do - blum	
12	hi - po - li - mi - mi - go	

Abb. 46: Seite 1 eines CBM-Protokolls aus dem Protokollheft am Beispiel der Messung Phono3

2. Silben klopfen

	Wort	r/f	Notizen
Übung	mü - de		
1	Re - gen - bo - gen		
2	Mu - sik		
3	Zahn - bürs - te		
4	Stein		
5	Ba - de - wan - ne		
6	Don - ner		
7	Son - nen - schein		
8	schwarz		
9	Pup - pen - wa - gen		
10	Hei - zung		
11	Zahn - pas - ta		
12	Bett		
13	Fin - ger - na - gel		
14	Hand - tuch		
15	Ap - fel - saft		
16	Kleid		

3. Lautisolation:

	Laut	Wort	A	E	r/f	Notizen
Übung	/o/	Opa	x			
1	/m/	kaum		x		
2	/r/	Rasen	x			
3	/l/	Waffel		x		
4	/n/	Schwein		x		
5	/b/	Bauch	x			
6	/ei/	Eier	x			
7	/f/	Fahne	x			
8	/s/	Gruß		x		
9	/k/	Klammer	x			
10	/t/	Gott		x		
11	/a/	achtzig	x			
12	/v/	Wasser	x			

Abb. 47: Seite 2 eines CBM-Protokolls aus dem Protokollheft am Beispiel der Messung Phono3

2.4 Motivationale Komponenten

Ein großer Nachteil klassischer Testverfahren besteht darin, dass die Diagnostiker fundiertes Wissen benötigen, um Tests in Schule oder Kindergarten durchführen, auswerten und interpretieren zu können. Zu einer Testdurchführung gehört jedoch auch ein Vertrauen in den Diagnostiker (vgl. Quaiser-Pohl 2010, S. 61), das häufig bei kleineren Kindern im

Kindergarten oder der Grundschule in externe Personen nicht so schnell aufgebaut werden kann. Aufgrund einer adäquaten diagnostischen Ausbildung zur Durchführung von psychologischen Tests ist es aber meist von Nöten, dass der Diagnostiker (z. B. ein Sonderpädagoge oder ein Schularzt) von einer externen Einrichtung hinzugezogen wird und dem Kind unbekannt ist. Es wird häufig versucht, eine solche Vertrauenssituation durch eine Art „Warming-up“ herzustellen, doch wird dieser Teil des Tests aus zeitlichen und Kostengründen häufig stiefmütterlich behandelt. Herkömmliche Testverfahren unterliegen zudem tagesrhythmischen Schwankungen bzgl. der Uhrzeit oder auch Gemütsschwankungen der Kinder.

Lernen, Lernerfolg und Motivation stehen in einem wissenschaftlich relevanten Zusammenhang. Die emotionale Befindlichkeit hat eine hohe Bedeutung, sowohl für das Lernen, als auch für die Leistungsfähigkeit zu einem bestimmten Zeitpunkt. Kinder mit sprachlichen Schwierigkeiten im Vorschulalter erleben über einen längeren Zeitraum konstant Misserfolge, die sie ihrer eigenen mangelnden Leistungsfähigkeit zuschreiben (vgl. Schulte-Körne 2006, S. 37f.). Für die Diagnostik ist es somit essentiell, motivational-emotionale Aspekte in die Konstruktion von Diagnostikinstrumenten im Bereich der Lese-Rechtschreibschwäche mit einzubeziehen. Die heute weit verbreiteten Diagnostikmethoden beruhen häufig nur auf einer einzigen Messung, die diese im Vorschulalter doch recht stark beeinflussenden Faktoren nicht berücksichtigen kann.

Um zu gewährleisten, dass die Kinder nicht die Motivation verlieren, begleitet „Phonolino“, eine Handpuppe in Form eines Schafes, die Kinder durch die Messungen. Zudem kann das Schaf Phonolino durch geeignete Spiele, Lieder und Geschichten vor den Messungen in der Kindergartengruppe kennengelernt werden. Das Schaf soll ein vertrautes wiederkehrendes Element darstellen. Durch das Schaf Phonolino wird eine lustige und entspannte Atmosphäre geschaffen, sodass Ängste von vornherein vermieden werden können (vgl. Phonolino-Handbuch, Anhang 1).

Im folgenden Abschnitt wird die genaue Konzeption der einzelnen Items von Phonolino dargestellt und begründet.

2.5 Konzeption der Items von Phonolino

Für die Entwicklung einer CBM stellten Deno und Fuchs (1987, S. 4) folgende Fragestellungen in den Fokus (vgl. Tab. 30 und Tab. 31):

- Was wird gemessen?
- Wie wird gemessen?
- Wie werden die Daten genutzt?

Tab. 30: „Decision-Making Matrix“ (nach Deno und Fuchs 1987, S. 4)

What to Measure	T-1	E-1	L-1
How to Measure	T-2	E-2	L-2
How to Use Data	T-3	E-3	L-3

Anmerkung: Die Nummerierungen der Zellen-werden in Tab. 31 erklärt.

Was wird gemessen?

Der Inhalt der CBM wird bereits durch die in der Theorie mehrfach beschriebene phonologische Informationsverarbeitung geliefert (vgl. Kap. II). Es werden die Fähigkeiten der Kinder im letzten Kindergartenjahr in ausgewählten Komponenten der phonologischen Informationsverarbeitung gemessen. Die Auswahl der Items wird im nächsten Abschnitt dargestellt. Es werden Prädiktoren gewählt, die bereits in deutschsprachigen Studien eine Vorhersage über die schriftsprachlichen Kompetenzen der Kinder in der Schulzeit treffen konnten.

Wie wird gemessen?

Phonolino entspricht einer Kombination der Methodik des curriculumbasierten Messens mit den wissenschaftlichen Erkenntnissen zur phonologischen Informationsverarbeitung aus dem deutschsprachigen Raum. Wie bereits im Kapitel III.2 ersichtlich, gibt es zwei Wege zur Erstellung eines curriculumbasierten Messverfahrens (vgl. Klauer 2006, S. 18). In diesem Fall wurde nicht der erste Weg der multidimensionalen Messung gewählt, in dem das Messverfahren die jeweiligen Teilfähigkeiten innerhalb einer Aufgabe abbildet, sondern der zweite Weg, für den zunächst die geforderte Leistung bestimmt wird, die zum Ende des Zeitraums, also in dem Fall zum Schulbeginn, von den Kindern gezeigt werden soll. Dafür müssen Aufgaben gefunden werden, die diese Teilmengen der Leistungen repräsentieren und in diesem Fall als Prädiktoren für Schwierigkeiten im Schriftspracherwerb gelten. In den Durchführungsanweisungen von Phonolino wurde außerdem die Instruktion mit exaktem Wortlaut festgehalten, den die Testleiter anwenden sollten. Ebenfalls vorgegeben wurden die monatliche Wiederholung der CBM sowie die Zeiteinheiten der Items.

Da die motorischen Fähigkeiten einiger Schüler noch nicht ganz genügend ausgeprägt sind, wurde darauf verzichtet, dass sie selbstständig etwas markieren oder ankreuzen sollen. Die gesamte CBM findet im Interviewstil mit einer spielerischen Komponente (Schaf Phonolino) statt.

Tab. 31: Fragestellungen der „Decision-Making Matrix“ (nach Deno und Fuchs 1987, S. 4)

Technical
T-1. What to measure: What behaviors clearly index growth and are responsive to the effects of instruction? T-2. How to measure: What measurement procedures generate reliable and valid data that are sensitive to short- and long-term growth? T-3. How to use data: How should the data be summarized, analyzed, and interpreted to ensure reliable and valid decision making?
Effectiveness
E-1. What to measure: Does measuring different behaviors relate to different achievement outcomes? E-2. How to measure: Do different measurement procedures result in different rates of student improvement? E-3. How to use data: What data-based decision rules result in more successful programs?
Logistical
L-1. What to measure: What behaviors are readily accessible for repeated measurement? L-2. How to measure: What measurements procedures are most efficient and least intrusive? L-3. How to use data: What data utilization procedures are most efficient and cost-effective?

Wie werden die Daten genutzt?

Die Frage der Datennutzung ist für die Evaluation der CBM nicht relevant (Stufe 3 nach Fuchs 2004, vgl. Kap. IV.6, S. 249). Die Erzieher haben jedoch genaue Hinweise für die Protokollierung erhalten. Die Auswertung wurde nicht von den Erziehern ausgeführt.

Im folgenden Abschnitt werden die Operationalisierung der phonologischen Informationsverarbeitung und die Unterscheidung der Komplexität der Aufgaben diskutiert.

2.5.1 Begründung zur Auswahl der Items

Im Kapitel II.1.2 wurden verschiedenen Untersuchungen vorgestellt, die nachweisen konnten, dass die phonologische Informationsverarbeitung eine bedeutende Komponente für den Schriftspracherwerb darstellt (z.B. Marx 1992b, S. 225). Die dargestellten Korrelations- und Trainingsstudien konnten zeigen, dass Kinder bereits vor Schuleintritt Vorläuferfertigkeiten des Schriftspracherwerbs aufweisen, die in den einzelnen Kategorien der phonologischen Informationsverarbeitung unterschiedlich ausgeprägt sind.

Es kann zusammengefasst werden (vgl. auch Kap. II.1.2.1 - Phonologische Bewusstheit), dass sich die phonologische Bewusstheit im engeren Sinne zeitlich meist nach der phonologischen Bewusstheit im weiteren Sinne entwickelt. Da die Zielgruppe der Phonolino-Messungen im Vorschulalter ist, wird erwartet, dass sie sowohl einige Fähigkeiten der phonologischen Bewusstheit im engeren als auch im weiteren Sinne aufweist. Aus diesem Grund wurde ein Item der phonologischen Bewusstheit im weiteren Sinne (Silben klopfen) sowie ein Item zur phonologischen Bewusstheit im engeren Sinne (Lautisolation) gewählt. Außerdem wurden diese beiden Items durch ein Item zur Überprüfung der Arbeitsgedächtnisprozesse ausgewählt (Pseudowörter nachsprechen).

Eine weitere wichtige Komponente sind auch die jeweiligen sprachlichen Einheiten, die innerhalb der Aufgaben von den Kindern identifiziert werden sollen. Das linguistische Abstraktionsniveau der verwendeten Einheiten stellt unterschiedliche kognitive Anforderungen an die Kinder (vgl. Barth 1999, S. 30). Nation und Hulme (1997, S. 154ff.) fanden heraus, dass das Onset-Reim-Bewusstsein aktiv instruiert werden muss, damit es Leseanfängern einen Vorteil im Schriftspracherwerb verschafft. Goswami (1993, S. 468f.) geht zudem davon aus, dass die Onset-Reim-Bewusstheit keine Vorläuferfertigkeit darstellt, sondern wahrscheinlich eine Folge des Schriftspracherwerbs ist (vgl. Kap. II.1.2.1 - Phonologische Bewusstheit), was Wimmer, Landerl, Schneider (1994, S. 469ff.) bestätigen konnten. Sie vermuten, dass Reime sich nicht direkt auf die alphabetische Strategie auswirken, sondern im späteren Lese- und Schreibprozess relevant werden. Aus diesem Grund wurden Aufgaben zum Onset-Reim-Bewusstsein bei der Erstellung von Phonolino nicht berücksichtigt.

Im Folgenden werden das wissenschaftliche Fundament und die Begründung zur Auswahl der jeweiligen Items dargestellt.

Item 1: „Pseudowörter nachsprechen“

Ein Prädiktor für Schwierigkeiten beim Schriftspracherwerb ist die Fähigkeit des phonetischen Rekodierens im Arbeitsgedächtnis, der unter anderem durch das Wiederholen von Pseudowörtern (z.B. ba-re-mi-fuk) gemessen werden kann. Nach Gathercole und Baddeley (1993a, S. 101ff., vgl. auch Kap. II.1.2.3 - Phonetisches Rekodieren im Arbeitsgedächtnis) begünstigen effiziente Arbeitsgedächtnisprozesse den Schriftsprach-erwerb. Besonders im Alter von sechs bis zehn Jahren konnten sie ein starkes Wachstum bzgl. der Dauer des Präsenthaltens und der Menge an Informationseinheiten beobachten (vgl. ebd.). Wie bereits im Kapitel II.1.2.3 bzgl. des Schriftspracherwerbs erwähnt wurde, ist die Kurzzeitgedächtnisspanne ebenfalls eine wichtige Komponente. Diese Fähigkeit wird durch das Benennen von Pseudowörtern überprüft. Hippmann (2008, S. 45f., vgl. auch Kap. II.1.2 - Phonologische Informationsverarbeitung) weist darauf hin, dass die Systeme der zentralen Exekutive und der artikulatorischen Schleife bei den Aufgabenstellungen nicht vermischt werden. Dies ist beim Einsatz von Pseudowörtern gewährleistet. Nach Baddeley und Hitch (1974, S. 47ff.) ist die Gedächtnisspanne für phonetisch ähnliche Items geringer als für klanglich deutlich verschiedene Items. Sie vermuten, dass ähnliche Items auch artikulatorisch ähnliche Spuren auf der phonologischen Schleife hinterlassen. Dies kann zu einer schlechteren Wiederholungsleistung der Kinder führen.

Item 2: „Silben klopfen“

Es konnte gezeigt werden, dass Vorschulkinder im letzten Kindergartenjahr normalerweise über ausgeprägte Fähigkeiten auf der Silbenebene verfügen (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002, S. 35). 72 % der Schüler konnten im Vorschulalter die Wörter richtig in Silben segmentieren, so dass zu vermuten ist, dass die Aufgabe „Silben klopfen“ für Vorschulkinder ein angemessenes Niveau hat, jedoch eher eine leichtere Aufgabe als Item 1 und 3 darstellt und damit gut nach unten differenzieren kann.

Item 3: „Lautisolation“

Jansen und Marx (1999, S. 11f.) konnten herausfinden, dass Kinder im Vorschulalter bereits lange Anfangsvokale kategorisieren können, die Spracherkennungsleistung aber abhängig von der Position des Vokals im Wort ist. Vokale werden von den Kindern leichter als Konsonanten analysiert. Auch die Analyse von Konsonanten ist von der Position im Wort abhängig und grundsätzlich am Anfang des Wortes leichter zu erkennen. Insgesamt ist festzustellen, dass die Kinder im Mittel schon relativ gut dazu in der Lage sind, Vokale und

größtenteils auch Konsonanten am Wortanfang und Wortende zu identifizieren. Üblicherweise sind sie jedoch noch nicht dazu fähig, diese in Phoneme zu segmentieren (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002, S. 34f.).

2.5.2 Zur Komplexität der Aufgaben

Die Aufgaben zur phonologischen Bewusstheit unterscheiden sich bezüglich ihrer Komplexität und der linguistischen Ebene, auf der sich die Prozesse abspielen. Jansen (1992, zit. nach Schnitzler 2008) stellt vier Kriterien zur Einordnung der Aufgaben zur phonologischen Bewusstheit auf:

(a) Größe der sprachlichen Einheit, die angeboten wird (Silbe, Onset-Reim, Phonem)
(b) Bewusstheit der sprachlichen Prozesse (implizit – phonologische Bewusstheit im weiteren Sinn, explizit – phonologische Bewusstheit im engeren Sinn)
(c) Kognitiver Anspruch (Aufmerksamkeit, Gedächtnis, Input und Output,…)
(d) Bedeutsamkeit der Einheiten für schriftsprachliche Prozesse (alphabetische bzw. orthografische Strategie, prädikative bzw. direkte Bezüge)

Größe der sprachlichen Einheit

In Item 1 werden Wörter dargeboten, die mit Betonung auf Silbenebene vorgesprochen werden. Da es sich jedoch um Pseudowörter handelt, ist nicht vorausgesetzt, dass den Kindern diese aus dem Alltag geläufig sind. Die dargebotene sowie auch die geforderte sprachliche Einheit ist beim Nachsprechen von Pseudowörtern das Wort. In Item 2 ist die Größe der sprachlichen Einheit in der Abfrage die Silbe. Die dargebotene sprachliche Einheit ist dagegen das Wort. Es wird ein Wort geboten, dass das Kind in Silben segmentieren soll. In Item 3 liegt die dargebotene sprachliche Einheit auf Wort- und Phonemebene. Die Kinder sollen herausfinden, ob sich ein Laut am Anfang oder am Ende eines Wortes befindet. D. h. die Segmentierung findet auf der Phonemebene statt.

Bewusstheit der sprachlichen Prozesse

Im Allgemeinen unterscheiden Jansen, Mannhaupt, Marx und Skowronek (2002, S. 9) Aufgaben zur phonologischen Bewusstheit im weiteren und die phonologische Bewusstheit im engeren Sinne (vgl. auch Kap. II.1.2.1, Abb. 48). Dies wirkt sich auch auf die Komplexität der Aufgaben aus. Die *phonologische Bewusstheit im weiteren Sinne* bezieht sich auf größere phonologische Einheiten. Insbesondere unbewusste metakognitiv-sprachliche Prozesse werden dieser Fähigkeit zugeordnet. Die Kinder setzen ihre Fähigkeiten oft intuitiv bei Reimspielen oder beim Silbenklatschen ein, die kognitiven Prozesse sind in der Regel

nicht im Bewusstsein der Kinder. Die *phonologische Bewusstheit im engeren Sinne dagegen geht* auf die Phoneme ein. Sie ist expliziter und die Prozesse sind im Bewusstsein der Kinder häufig genau ausgeprägt. Es handelt sich um eine bewusste Analyse der phonologischen Struktur der Sprache und erfolgt in der Regel nicht im Zusammenhang mit der Semantik oder dem Sprachrhythmus (vgl. Jansen und Marx 1999, S. 13f.).

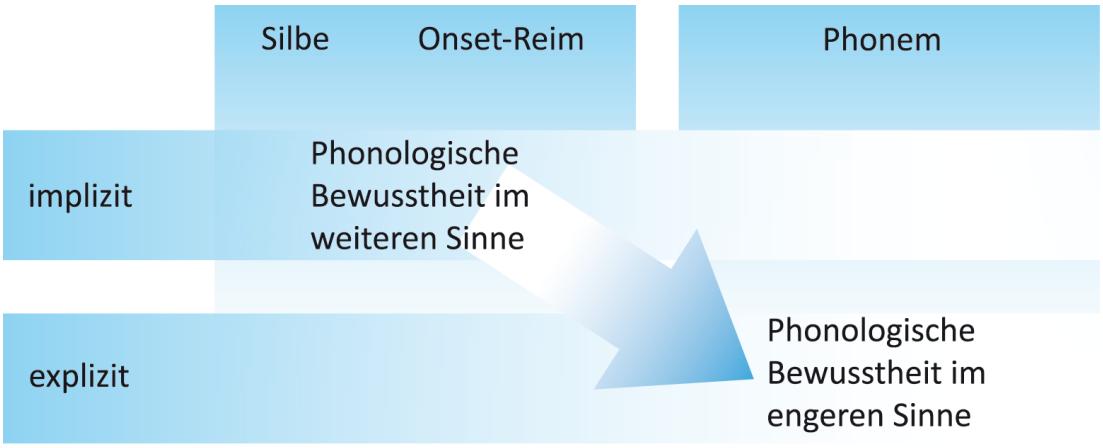


Abb. 48: Zusammenhang der phonologischen Bewusstheit im engeren und im weiteren Sinne (nach Schnitzler 2008, S. 20)

Aufgrund dieser grundsätzlichen Unterscheidung wurden auch die Items von Phonolino entwickelt. Item 1 und 3 werden der phonologischen Bewusstheit im engeren Sinne, Item 2 der phonologischen Bewusstheit im weiteren Sinne zugeordnet. Alle drei stellen wichtige Variablen zur Vorhersage von schulischen Schwierigkeiten im Schriftspracherwerb dar.

Kognitiver Anspruch

In Item 1 ist das Ausmaß der kognitiven Anforderung bzgl. des Gedächtnisses sehr hoch, da die Kinder sich die einzelnen Silben der Pseudowörter merken müssen, um diese dann in der richtigen Reihenfolge wieder abzurufen. Beim Segmentieren der Silben (Item 2) ist das Ausmaß der geforderten Konzentration nicht ganz so hoch, da die Kinder bekannte Wörter erhalten, die sie in Silben segmentieren sollen. Dies ist oft ein unbewusster Prozess und erfordert deshalb auch keine hohe Gedächtnis- oder Konzentrationsleistung. Bei der Erkennung der Position der Laute (Item 3) ist die Aufmerksamkeitsleistung etwas höher, da die Kinder zwei Informationen erhalten: das Wort und das Phonem. Sie müssen diese beiden Informationen erfassen, analysieren und kombinieren.

Bedeutsamkeit der Einheiten für schriftsprachliche Prozesse

Da es sich bei den drei Items ausschließlich um Messungen für den Vorschulbereich handeln soll, die vor dem Anfangsunterricht stattfinden, beziehen sie sich auf Vorläuferfertigkeiten

des Schriftspracherwerbs. Orthografische Kenntnisse werden dabei aufgrund des Alters gänzlich außer Acht gelassen. Bei Item 1 und 2 handelt es sich um prädikative Bezüge zum Schriftspracherwerb, während sich Item 3 zur Erkennung der Position von Lauten bereits direkt auf die Schriftsprache und auf die Erkennung der Laute bezieht. Anhand der Einteilung der Aufgaben zeigt sich, dass die Aufgaben der drei Items in ihrer Komplexität unterschiedlich angesiedelt sind, wobei alle Inhalte am Anfang des Schriftspracherwerbs einzuordnen und damit für den Vorschulbereich gut geeignet sind (vgl. Tab. 32).

Tab. 32: Phonologische Informationsverarbeitung bei den Items von Phonolino (Einteilung nach Schnitzler 2008, S. 29f.)

Sub-messung	Inhalt	Sprachliche Einheit	Phonologische Bewusstheit …	Kognitiver Anspruch	Bedeutsamkeit für den schriftsprachlichen Prozess
1	Pseudo-wörter erkennen	Wort → Wort	im engeren Sinne	Hoch (Konzentration + Gedächtnis)	Alphabetische Strategie/prädikative Bezüge
2	Silben klopfen	Wort → Silbe	im weiteren Sinne	Niedrig	Alphabetische Strategie/prädikative Bezüge
3	Lautisolation	Wort → Phonem	im engeren Sinne	Mäßig (Konzentration)	Alphabetische Strategie/direkte Bezüge

2.5.3 Zweidimensionales Konstrukt zur Erstellung der Aufgaben

Eine weitere Möglichkeit zur Unterscheidung von Aufgaben basiert auf dem zweidimensionalen Konstrukt, das auf dem psycholinguistischen Sprachverarbeitungsmodell von Stackhouse und Wells (1997) und der englischsprachigen Testtheorie beruht (vgl. Schnitzler 2008, S. 21f.). Dieses Konstrukt zur Differenzierung der Komplexität von Aufgaben zur phonologischen Bewusstheit wird in zwei Dimensionen unterteilt, die jeweils in verschiedene Stufen eingeteilt sind. Die beiden Dimensionen stehen anders als beim Konstrukt aus der Bielefelder Studie senkrecht aufeinander, sodass alle Komplexitätsebenen berücksichtigt werden können, die in dem Bielefelder Konstrukt zum Teil unbeachtet gelassen wurden (vgl. Schnitzler 2008, S. 21f.).

Es wird zwischen der Dimension der phonologischen Einheit und der Dimension der Operation unterschieden (vgl. Abb. 49).

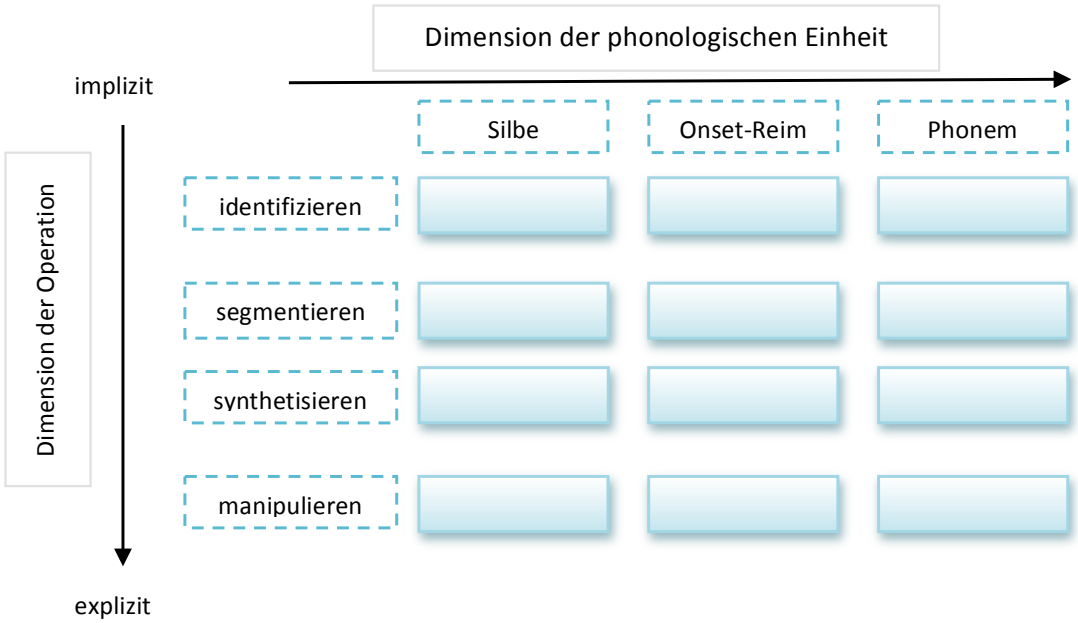


Abb. 49: Zweidimensionales Konstrukt der phonologischen Bewusstheit (nach Schnitzler 2008, S. 29).

Die Komplexität bzgl. der *Dimension der phonologischen Einheit* bezieht sich insbesondere auf die Größe der Einheit. Es wird in Silben, Reime, Onsets und Phoneme in aufsteigender Komplexität unterschieden. Die Silben können im Sprechfluss erkannt werden und sind zum Teil aus Sprachspielen, Kinderliedern und weiteren bekannten Komponenten aus dem Kindergarten bekannt. Sie bieten einen ersten Zugang zu den formalen Aspekten der Lautsprache. Die Onsets und Reime werden laut Bryant MacLean und Crossland (1989, S. 407ff.) durch das Umfeld der Kinder angeregt, da Kinderreime und Sprachspiele häufig zum Kindergartenalltag dazugehören. Um Phoneme zu erkennen und größere Spracheinheiten in Phoneme zu durchgliedern, ist ein größeres Abstraktionsvermögen vonnöten. Phoneme werden häufig nicht durch Erfahrung gelernt, sondern sind abstrakte Informationseinheiten bei der Verarbeitung von Sprache (vgl. Schnitzler 2008, S. 22ff.). Die Komplexitätsstufe ist jedoch nicht nur vom Ergebnis der Aufgabenstellung abhängig, sondern auch von einfacher erkennbaren Komponenten. So geht es beispielsweise bei der Lautisolation, also bei der Angabe des Anfangsphonems im Ergebnis darum, dass die Kinder das Phonem erkennen. Innerhalb dieser Aufgabenstellung lassen sich aber noch einfachere und schwierigere Aufgaben unterscheiden. Ist das zu erkennende Phonem gleichzeitig eine Silbe (z.B. O-fen), so ist dies leichter zu erkennen als wenn das Phonem ein Onset ist (B-all). Die schwierigste Stufe wäre das Erkennen eines Phonems, das nicht Onset oder Silbe ist, z. B. F – lug (vgl. Schnitzler 2008, S. 23). Existieren nun ausschließlich Items auf Silbenebene in einem Test und es wird jedoch die Fähigkeit des Erkennens von Anfangslauten auf Phonemebene abgeleitet, dann könnte dies zu einer Fehlanalyse des Ist-Zustandes des

Kindes führen. Interpretationen sollten mit Bedacht und unter genauer vorheriger Analyse der Wörter durchgeführt werden. Der Schwierigkeitsgrad macht sich somit anhand des größtmöglichen Elements in der Aufgabe fest, nicht am geforderten Ergebnis der Überprüfung (vgl. Schnitzler 2008, S. 23).

Die *Dimension der Operation* bezieht sich auf vier Stufen, die sich auf die Explizitheit der Operation bzw. der Bewusstheit der Analyse berufen. Die kognitive Anforderung zur Lösung einer Aufgabe sowie der Grad der Bewusstheit einer Operation werden hier berücksichtigt. Die Bewusstheit der Analyse bzw. die Explizitheit der Operation beeinflusst den Grad der Abstraktheit beim Abruf der phonologischen Repräsentationen im mentalen Lexikon. Je expliziter eine Operation, umso mehr werden die Gedächtniskomponenten der Kinder in Anspruch genommen. Die Dimension der Operation wird in aufsteigender Reihenfolge am Identifizieren, Synthetisieren, Segmentieren oder Manipulieren in der Aufgabe festgemacht (vgl. Tab. 33).

Tab. 33: Einheiten der Dimension der Operation eines zweidimensionalen Modells (angelehnt an Schnitzler 2008, S. 21ff.)

Identifizieren	Vorgegebene Einheiten müssen erkannt werden	Hörst du ein /o/ in Ofen?
Segmentieren/Synthetisieren	Erkennen phonologischer Einheiten	Wie viele Phoneme hat das Quatsch-Wort „Rasek“?
Manipulation	Phonologische Einheiten müssen verändert werden	Vertausche die ersten beiden Laute des Wortes Amsel. (/masel/)

Gathercole und Baddeley (1993a, S. 101ff.) beschreiben die Unterschiede als explizite und implizite phonologische Bewusstheit in Bezug auf die Bewusstheit bzw. aktiven Realisierung der Handlung. Jansen und Marx (1999) ermittelten diesbezüglich, dass 96 % der Vorschulkinder vier Monate vor Schulbeginn zwar ein Wort in Silben unterteilen konnten, jedoch nur 34 % der Kinder dazu in der Lage waren, eine der identifizierten Silben (z. B. die mittlere) zu benennen.

Dies zeigt, dass Kinder für Aufgaben mit wenigen Verarbeitungsschritten auf kognitiver Ebene ihr implizites sprachliches Wissen anwenden können und eine direkte Bewusstheit der Handlung nicht notwendig ist, während bei vielen kognitiven Verarbeitungsschritten die Nutzung von explizitem sprachlichem Wissen unumgänglich ist (vgl. Schnitzler 2008,

S. 25). Im Vergleich zum Konstrukt der phonlogischen Bewusstheit im weiteren und im engeren Sinne können Aufgaben anhand des zweidimensionalen Modells noch genauer klassifiziert werden. Es ergibt sich eine Matrix als eine Art Kompetenzraster, anhand derer Aufgaben gefunden werden können, die den Anforderungen im Vorschulbereich entsprechen.

Bezüglich des zweidimensionalen Modells können nun die einzelnen Wörter der Items von „Phonolino“ nochmals genauer betrachtet werden. Eine grobe Einordnung konnte schon durch das Bielefelder Modell vorgenommen werden. Nun erfolgt eine Einordnung in die Dimensionen der Operation und der phonologischen Einheit im zweidimensionalen Konstrukt der phonologischen Bewusstheit.
In Item 1 werden die Pseudowörter als Wörter geboten, die jedoch als Verarbeitungseinheit für die Kinder als Silben wahrgenommen und im Arbeitsgedächtnis gespeichert werden. Die Wörter werden identifiziert, in Silben segmentiert, anschließend synthetisiert und dann artikuliert. Item 1 betrifft damit die einfachste Ebene der Dimension der phonologischen Einheit, bewegt sich aber bezüglich der Operation auf den ersten drei Ebenen der Identifikation, Segmentierung und Synthese (vgl. Abb. 50). Bei Item 2 werden Silben geklopft. Dazu werden die dargebotenen (wahrscheinlich bekannten) Wörter zunächst identifiziert und anschließend in Silben segmentiert. Auch Item 2 bewegt sich ausschließlich auf der untersten Ebene der Dimension der phonologischen Einheit, während es bezüglich der Komplexität in der Dimension der Operation auf den ersten beiden Ebenen anzusiedeln ist. Item 3 gehört auf der Dimension der Operation zum Bereich „Identifizieren“, da es sich um eine Laut-zu-Wort-Operation handelt, in der ein Laut identifiziert werden muss. Die Kinder benötigen hier nur ein rudimentäres Konzept zu den Einheiten und müssen nur wenig auf abstrakte Einheiten zurückgreifen. Die Identifikation tritt je nach Item auf der Onset-Reim- oder auf der Phonemebene auf. Bei der Erstellung der Aufgaben von Phonolino wurden insbesondere die Konsonantendopplung sowie die Anzahl der Silben bei der Erstellung der Parallelmessungen berücksichtigt. Ein weiteres relevantes Kriterium war der Bekanntheitsgrad der Wörter. Es wurde in Item 3 explizit darauf geachtet, dass die genutzten Items dem Wortschatz von Kindergartenkindern entsprechen, sie phonetisch möglichst einfach zu produzieren sind und es sich nicht um Fremdwörter handelt.

Die Messungen sind für den Vorschulbereich erstellt worden. Wie anhand des zweidimensionalen Kompetenzrasters sichtbar gemacht wurde, bilden die Aufgaben die basalen Kenntnisse der Kinder im Bereich des Schriftspracherwerbs ab. Bei den drei Items der Messungen von Phonolino wurden explizit keine Aufgaben zur Manipulation dargeboten, da

dies insbesondere den Fähigkeiten von Schulkindern entspricht und die Kinder im Vorschulbereich noch überfordern würde (vgl. zusammenfassend Schnitzler 2008).

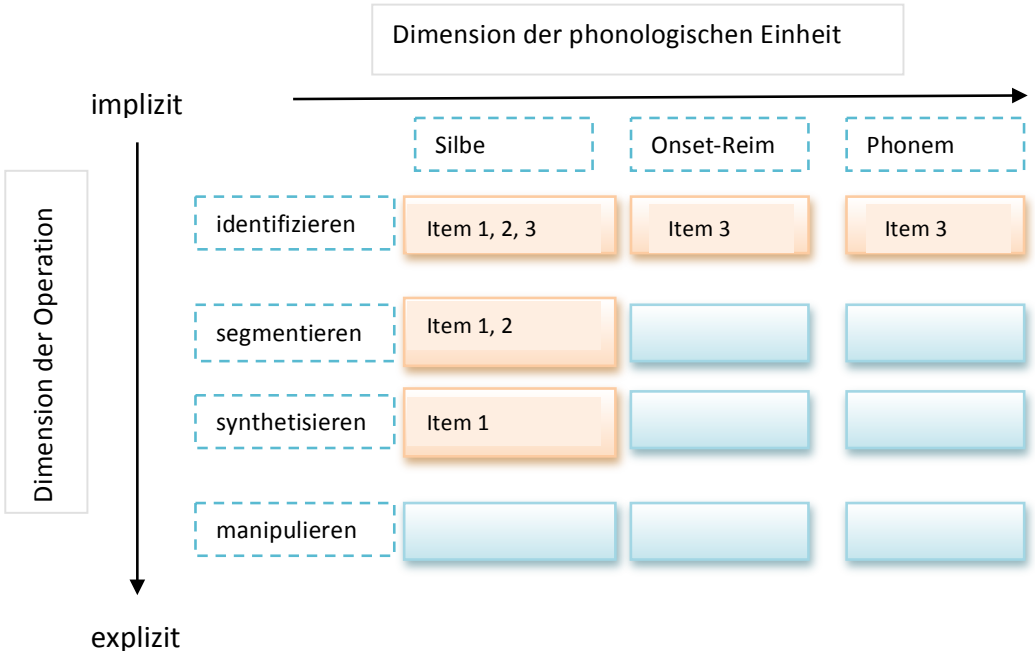


Abb. 50: Einordnung der Messungen von Phonolino am zweidimensionalen Konstrukt der phonologischen Bewusstheit (verändert nach Schnitzler 2008, S. 29)

2.5.4 Erstellung der Parallelmessungen

Zusätzlich ist es sinnvoll auch etwaige Nebenfaktoren zu betrachten, die die Komplexität des Items beeinflussen können (vgl. Schnitzler 2008, S. 30ff.). Dies ist für die Erstellung der Parallelmessungen vonnöten, um eine annähernde Ähnlichkeit in der Komplexität der Items von Messung zu Messung zu gewährleisten.

Es wurde darauf verzichtet Pseudoparallelmessungen, also eine ausschließliche Vertauschung der Items bei jeder Messung, zu verwenden. Dies war mit der Methodik der curriculumbasierten Messung nicht zu vereinen, da nicht gewährleistet werden konnte, dass bei einer Durchmischung der Items alle Items bei jedem Schüler angewendet werden können, da die CBM zeitabhängig als Speedtest strukturiert ist. Also wurden anhand bestimmter Kriterien Items mit möglichst gleichem Schwierigkeitsgrad erstellt.

Nebenfaktoren, die die Komplexität des Items beeinflussen können sind (vgl. Schnitzler 2008, S. 30ff.):

(a) Sonorität der phonologischen Einheit
(b) Mit Vokalen, Liquiden oder Nasalen können Kinder einfacher umgehen als mit Obstruenten. Innerhalb der Vokale können lange Vokale besser erkannt werden als kurze.
(c) Betonung der phonologischen Einheit
(d) Phonologische Einheiten in einer betonten Silbe sind einfacher zu bearbeiten als diejenigen in einer unbetonten Silbe.
(e) Position der phonologischen Einheit
(f) Die initiale Position der phonologischen Einheit ist einfacher zu bearbeiten als die finale Position und die finale Position ist einfacher als die mediale Position.
(g) Komplexität der phonologischen Umgebung der phonologischen Einheit
(h) Die phonologische Umgebung ist hinsichtlich der Länge der Einheit und der phonologischen Silbenstruktur von Interesse. Je länger die Items sind, umso mehr wird das Arbeitsgedächtnis in Anspruch genommen und umso mehr ist die selektive Aufmerksamkeit aufgrund der erhöhten Anzahl irrelevanter Merkmale gefragt. Bezüglich der phonologischen Silbenstruktur sind beispielsweise Konsonantenhäufungen im Vergleich zu einfachen KVK[23]-Strukturen schwieriger (vgl. auch Hacker 2002, S. 20f.).
(i) Lexikalität
(j) Bei der Darbietung von Wörtern hat die Vorkommenshäufigkeit der Wörter im jeweiligen Sprachraum Relevanz. Bekannte Wörter können durch lexikalische Effekte einfacher erkannt werden als beispielsweise Pseudowörter. Jedoch auch bei der Darbietung von Pseudowörtern existieren Wörter, die bekannten Wörtern eher ähneln und damit Klangassoziationen und Top-Down-Effekte hervorrufen.
(k) Input- und Output-Anforderungen
(l) In- und Outputleistungen: Inputleistungen beschreiben beispielsweise das Erkennen von Lauten, während eine Outputleistung zum Beispiel die Realisation von Lauten ist. Es wird vermutet, dass Inputleistungen vor den Outputleistungen erworben werden.

Der Unterpunkt „Lexikalität“ erschwerte die Beachtung aller Nebenfaktoren, da es für die Erstellung der Parallelmessungen vonnöten war, eine große Anzahl an Wörtern zu produzieren. Die Kinder sollten für Item 2 und Item 3 möglichst alle Wörter kennen, damit sie nicht durch die unbekannte Wortbedeutung abgelenkt werden. So mussten Wörter gefunden werden, die dem Lebenshorizont der meisten deutschsprachigen Kinder im

23 K ist die Abkürzung für Konsonant, V für Vokal, KVK ist eine Verbindung von Konsonant-Vokal-Konsonant.

Vorschulalter entsprechen. Der Lexikalität wurde vor den anderen Nebenfaktoren eine höhere Bedeutung zugemessen, da diese die Leistung der Kinder stark beeinflusst und gleichzeitig die Durchführung durch Nachfragen hemmen könnte. Zudem wäre dann anderenfalls nicht sicherzustellen, ob die Kinder die sprachlichen Einheiten nicht identifizieren oder ob sie das für sie unbekannte Wort nur nicht so lange speichern konnten bzw. ob Nebenschauplätze (z. B. die Suche nach der Bedeutung des Wortes) höhere Priorität als die Lösung der Aufgabe hatte. Eine Übersicht zu den Parallelmessungen findet sich in Anhang 3.

Nebenfaktoren für Item 1 „Pseudowörter nachsprechen“

Bei der Erstellung der Parallelmessungen der Items standen insbesondere die Komplexität der phonologischen Umgebung und die Länge der Einheiten im Vordergrund. Um überprüfen zu können, wie viele Einheiten die Kinder im Kurzzeitgedächtnis präsent halten können, war es wichtig, die Anzahl der Einheiten der Pseudowörter zu variieren. Die Ähnlichkeit der Items wurde über die phonologische Umgebung bestimmt. Es wurde ein Muster von KVK-Strukturen entwickelt mit einer vorgegebenen Anzahl von Silben pro Pseudowort, das für die Parallelmessungen umgesetzt worden ist (vgl. Tab. 34).

Tab. 34: Beispiel für das KV-Muster aus den Parallelmessungen Phono5 und Phono6

4-silbiges Pseudowort					
KV	KV	KV	KV		
ba	ni	ka	to		
la	ri	fa	ro		
6-silbiges Pseudowort					
KKVK	KV	KV	KV	KV	KV
kron	lo	ma	ta	ka	wi
krun	lu	na	pa	ta	si

Bei der Erstellung der Parallelmessungen wurde die Silbenlänge des Pseudowortes konstant gehalten. Bei KV-Strukturen war zudem der Vokal gleichbleibend und nur der Konsonant wurde verändert, um möglichst gleiche Items in den Parallelmessungen zu erstellen, da es den Kindern aufgrund der Sonorität leichter fällt, Vokale wieder zu erkennen, als Konsonanten. Bei Konsonantendopplungen dagegen steht die Kombination der beiden

Konsonanten als komplexe Einheit im Vordergrund. Bei Kombinationen wie KKVK wurde deshalb der Vokal variiert. Da von Kindern lange Vokale besser als kurze erkannt werden können, wurden für die Pseudowörter nur lange Vokale verwendet, damit die Konzentration der Schüler auf das Behalten der Einheiten gelenkt werden kann.

Nebenfaktoren für Item 2 „Silben klopfen“

Beim Silbenklopfen wurde besonders auf die Anzahl der Silben geachtet, die sich in den einzelnen Parallelmessungen widerspiegelt (vgl. Tab. 35), sowie auf die Lexikalität, damit die Kinder die sprachlichen Einheiten besser präsent halten konnten.

Tab. 35: Beispiel aus den Parallelmessungen Phono5 und Phono6

	4-silbiges Wort			
Phono5	Klei-	der-	ha-	ken
Phono6	Erd-	beer-	ku-	chen
	2-silbiges Wort			
Phono5	Spiel-	platz		
Phono6	Stra-	ße		
	1-silbiges Wort			
Phono5	Huhn			
Phono6	Nuss			

Als Methodik wurde absichtlich auf das Klatschen der Silben verzichtet und alternativ das Klopfen zur Unterteilung der Silben genutzt, damit Kinder im Kindergarten auswendig gelerntes Wissen nicht anwenden, sondern sich allein auf die Silbentrennung konzentrieren können. Durch die Berücksichtigung der Lexikalität und der Anzahl der Silben konnten aufgrund der benötigten neun CBM-Parallelversionen weitere Komponenten kaum Berücksichtigung finden.

Nebenfaktoren für Item 3 „Lautisolation“

Bei der Erstellung von Item 3 zur Lautisolation stand besonders die Position der sprachlichen Einheit im Vordergrund. Wie sich im Vortest (vgl. Kap. IV.2.2) erwiesen hatte, konnten die Kinder im letzten Kindergartenjahr die mittlere Position eines Lautes noch nicht erkennen. Somit wurden nur die beiden etwas leichteren Einheiten an initialer und finaler Position in

dem Item berücksichtigt. Die Lexikalität hatte ebenfalls Priorität, da die Kinder nicht von der Unbekanntheit eines Wortes irritiert werden sollten. Es wurde zudem versucht die Länge der Worte möglichst kurz zu halten, damit die Kinder nicht aus einer Vielzahl an Lauten differenzieren müssen. Außerdem kann das kürzere Wort länger präsent gehalten werden.
Bei der Erstellung der Parallelmessungen wurden gleiche Items erstellt, indem der gleiche Vokal oder Konsonant an gleicher Position nur in einem anderen Wort erkannt werden sollte (vgl. Tab. 36). Dabei wurden Sonoritäten der Vokale und Konsonanten weitestgehend konstant gehalten. Wenn ein langer Vokal erkannt werden sollte, wurde auch in den Parallelmessungen der gleiche lange Vokal in einem anderen Wort gefordert.

Tab. 36: Beispiel aus den Parallelmessungen CBM5 und CBM6

Erkennung des Vokals /o/	
Phono5	ohne
Phono6	oder
Erkennung des Kononanten /k/ mit Berücksichtigung der Konsonantendopplung „kl“	
Phono5	Klappe
Phono6	Kleid
Erkennung des Konsonanten /n/	
Phono5	Zaun
Phono6	Bein

Diejenigen Wörter, bei denen initiale bzw. finale Laute erkannt werden sollten, wurden an unterschiedlicher Stelle angeboten, sodass die Kinder keine Regel ableiten konnten. Die Position wurde so wenig wie möglich variiert. Somit soll verhindert werden, dass die Kinder Regelmäßigkeiten entdecken und beginnen, die Position des Lautes zu erraten.

Nach einer ausführlichen Darstellung der theoretischen Grundlagen zu Erstellung einer curriculumbasierten Messung im Vorschulalter sowie nach der Erläuterung der Konzeption von Phonolino werden im nun folgenden zweiten Teil dieser Arbeit Aufbau und Konzeption der CBM „Phonolino“ dargelegt sowie die längsschnittliche Evaluation mit N = 256 Kindern in Schleswig-Holstein beschrieben und im Anschluss diskutiert. Dazu werden zunächst

Forschungsfragen aufgestellt, die im folgenden in Hypothesen und Annahmen operationalisiert werden.

V. Forschungsfragen der empirischen Untersuchung

Nach einer theoretischen Betrachtung der Notwendigkeit der Forschung im Bereich der Früherkennung von Lese- und Rechtschreibschwierigkeiten im ersten Kapitel dieser Arbeit sowie einer darauf folgenden Bestandsaufnahme zu den Folgen der Lese- und Rechtschreibschwäche, wurden schließlich die existierenden Testinstrumentarien und deren Problematik bei der Klassifikation beschrieben. Im zweiten Kapitel wurden drei mögliche Einflussfaktoren, die eine Fehlklassifikation verursachen können, vorgestellt. Zum einen wurden die Inhalte und mögliche Vorläuferfertigkeiten begutachtet, zum anderen aber auch exogene und endogene Faktoren, die die Leistungen der Kinder als Risiko- bzw. Schutzfaktoren verändern können. Als dritte Komponente wurde die Variation der Methodik in Anbetracht gezogen, die möglicherweise die Vorläuferfertigkeiten der Kinder nicht genau genug messen kann und ebenfalls dadurch Fehlklassifikationen verändert wird. Die aus dem amerikanischen Sprachraum stammende Methodik der curriculumbasierten Messung scheint ein sinnvolles Diagnostikkonzept zu sein, mit dessen Hilfe Kinder und Schüler mit Schwierigkeiten im Lesen, Rechtschreiben und in Mathematik in der Praxis effizient identifiziert werden können. Die vorliegenden empirischen Befunde zu curriculumbasierten Messungen im amerikanischen wie teilweise schon im deutschsprachigen Raum lassen vermuten, dass eine CBM im Vorschulalter mit den Inhalten zur Überprüfung der Vorläuferfertigkeiten, insbesondere der phonologischen Informationsverarbeitung der Kinder eine gute Prognose über die Lese- und Rechtschreibfähigkeiten in der Schule ermöglichen kann. Die Vorläuferfertigkeiten der phonologischen Informationsverarbeitung können anzeigen, so die Vermutung, ob im Verlauf der kindlichen Entwicklung im letzten Kindergartenjahr Schwierigkeiten in der Schriftsprache zu erwarten sind. Bei einer frühzeitigen Erkennung können rechtzeitig Maßnahmen zur Förderung bereits vor Schuleintritt ergriffen werden, sodass die Kinder sich ihren Fähigkeiten entsprechend mit angemessener Förderung und Unterstützung entwickeln können.

Die oben beschriebenen und entwickelten theoretischen Befunde bilden die Grundlage für Konzept und Design der folgenden empirischen Untersuchung. Es wurde ein Messverfahren zur Früherkennung von Lese-Rechtschreibschwierigkeiten im Vorschulalter mit Namen „Phonolino“ entwickelt (vgl. Kap. IV), das die Methodik der CBM und die theoretischen Erkenntnisse zur phonologischen Informationsverarbeitung zusammenführt. In der zweiten Hälfte der vorliegenden Arbeit wird nun zunächst die Fragestellung abgeleitet, im Anschluss werden das Untersuchungsdesign, Untersuchungsinstrumentarien sowie die eingesetzten statistischen Verfahren erläutert (vgl. Kap. VI). Nachfolgend werden die Ergebnisse der über zwei Schuljahre durchgeführten Längsschnittstudie vorgestellt (vgl. Kap. VII), in den theore-

tischen Rahmen eingeordnet und diskutiert (vgl. Kap. VIII). Im Ausblick werden dann die Grenzen und Möglichkeiten beschrieben, wie diese Ergebnisse in der Forschung einzuordnen sind und wie eine mögliche Erweiterung und Vertiefung des Forschungsgebiets durch nachfolgende Studien sinnvoll ist (vgl. Kap. IX).

Erfolg im Schriftspracherwerb ist von vielen unterschiedlichen Faktoren abhängig. Aus pädagogischer Sicht spielen dabei die gezielte Diagnostik und individuelle Förderung eine zentrale Rolle. Eine möglichst frühe Betrachtung der Schwierigkeiten anhand der Vorkenntnisse der Kindergartenkinder vor dem schulischen Anfangsunterricht ist essentiell. Eine der Schwachstellen stellt dabei der Übergang vom Kindergarten in die Schule dar (vgl. Knauf & Schubert 2008, S. 1). Es gibt trotz einiger Reformen zur Ermöglichung der Weitergabe von Daten häufig Schwierigkeiten bei der Übermittlung des Leistungsstandes der Kinder vom Kindergarten in die schulische Einrichtung. Diese missliche Datenlage kann im Anfangsunterricht ohne weitere Diagnostik und Analyse zu einer unzureichenden Passung der Unterrichtsinhalte und des Leistungsvermögens des Kindes führen. Es können Lese- und Rechtschreibproblematiken entstehen, die durch eine frühzeitige Diagnostik (und ggf. Förderung) hätten ausgeschlossen werden können. Diese Lese- und Rechtschreibproblematiken bleiben oftmals in den ersten Schuljahren unentdeckt (vgl. Kap. I.1.3.2 - Verlauf und Stabilität der Lese-Rechtschreibschwäche). Dies erweist sich vor dem Hintergrund der Bedeutsamkeit der Lese- und Schreibfähigkeit in unserer Gesellschaft als fatal, da Menschen, die in ihren schriftsprachlichen Kenntnissen eingeschränkt sind, die Teilhabe am gesellschaftlichen Alltags- und Berufsleben dadurch verwehrt bleibt.

Durch die Einführung von Curricula in Kindertagesstätten stellt sich gleichermaßen die Frage nach einer adäquaten Förderung der einzelnen Kinder sowie einer ökonomischen und effizienten Art der Datenerhebung zum Leistungsstand des Kindes (vgl. Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein 2004). In der Praxis scheinen jedoch weder geeignete, den Bedürfnissen der Kindertagesstätten angepasste, valide Diagnostikinstrumente zu existieren (vgl. Kap. I.2), noch sind ausreichend pädagogische Fachkräfte vorhanden, die eine diagnostische Ausbildung absolviert haben. Die Schulung der phonologischen Bewusstheit war in den vergangenen Jahren zunehmend Gegenstand eines Fachdiskurses. Sie wurde als wichtigster Faktor zur Vorhersage der Lateralität der Schüler im Schulleben erkannt. Erzieherinnen und Erzieher wurden vermehrt in Verbänden und Organisationen sowie öffentlichen Einrichtungen zu dem Thema geschult und von einigen Institutionen bereits zur expliziten Förderung der phonologischen Bewusstheit in der Kindertagesstätte verpflichtet. Der diagnostische Bereich wurde jedoch oft aufgrund fehlender Mittel, Zeit oder geeigneter Diagnostikinstrumente vernachlässigt

oder auf informeller Ebene durchgeführt. Vorhandene standardisierte Testverfahren, die auf dem Markt zur Verfügung stehen, lassen sich meist nicht sinnvoll in den Alltag der Kindertagesstätten integrieren (vgl. Stufe 3 nach Fuchs 2004, s. Kap. III.6). Entweder nehmen sie als Einzeltest zu viel Zeit in Anspruch oder die Auswertung und Interpretation sind ohne testdiagnostische Ausbildung zur direkten Anwendung in der Einrichtung zu komplex. Herkömmliche Testverfahren können zudem individuelle Veränderungen der Leistung im Laufe des letzten Kindergartenjahres nicht erfassen. Sie zeigen nur einmalige Momentaufnahmen an, die von der Motivations- und Leistungsfähigkeit des Kindes in der Testsituation tagesabhängig sind.

In der Metaanalyse von Bus und van Ijzendoorn (1999, S. 406ff.) konnte gezeigt werden, dass Vorschulkinder deutlich mehr von einer Förderung im Bereich der phonologischen Bewusstheit profitieren als Kinder, die bereits die ersten Klassen der Grundschule besuchen. Einsiedler et al. (2002, S. 199ff.) konnten diese Befunde bestätigen. Sie stellten fest, dass sich zwar die Trainingsgruppe signifikant in der phonologischen Bewusstheit verbesserte, die Transferleistungen auf das Lesen und Rechtschreiben dagegen irrelevant waren. Der Übergang zwischen Kindergarten und Grundschule stellt zudem

> *„[...] ein aus psychologischer und sozialisationstheoretischer Perspektive kritisches Lebensereignis dar, dessen Bewältigung oder aber Nichtbewältigung entscheidend zum Ge- oder Misslingen des weiteren kindlichen Entwicklungs- und Bildungsprozesses beiträgt.“ (Schumacher 2007, S. 9)*

Somit ist es sinnvoll so früh wie möglich die diagnostischen Bemühungen voran zu treiben, um damit auch frühzeitig eine Förderung einleiten zu können. Ein übergeordnetes Ziel dieser Studie besteht darin, die Verbesserung der Grundvoraussetzungen der Schüler bei Schuleintritt durch rechtzeitige Erkennung und folgende frühzeitige Förderung der Kinder zu ermöglichen und nachzuweisen.

Die dargestellten Gründe legen nahe, dass eine curriculumbasierte Messung, wie sie bereits in Kapitel III dargestellt wurde, auch für den vorschulischen Bereich einen enormen Vorteil mit sich bringt. Die Verbindung der curriculumbasierten Messung mit dem RTI-Ansatz kann insofern ein großer Gewinn für die frühkindliche Förderung sein. Zudem ist es aus wissenschaftlicher Perspektive insbesondere nach den eher schlechten Untersuchungsergebnissen der Replikationsstudie zum Bielefelder Screening interessant (vgl. Marx & Weber 2006), ob weitere beeinflussende Faktoren das Risiko erhöhen, Schwierigkeiten im Lesen und Schreiben in der Schule zu entwickeln und ob die

phonologische Informationsverarbeitung tatsächlich einen prädiktiven Charakter hat. Hinzu kommen Ergebnisse, die anzeigen, dass zwar die spezifischen Leistungen, also das Vorwissen der Kinder zum direkten maßgeblichen Erfolg beitragen, jedoch die allgemein-kognitiven Leistungen ebenfalls einen indirekten Einfluss auf den Erwerb schulischer Kompetenzen haben (vgl. Walter 2014b S. 8f.).

Tröster (2009, S. 71f.) beschreibt die wichtigsten Voraussetzungen für die Früherkennung mit drei Grundelementen:

- *Die Störung muss durch Frühindikatoren identifizierbar sein*. In dieser Studie wird anhand der theoretischen Grundlagen vermutet, dass die phonologische Informationsverarbeitung einen der Frühindikatoren darstellt (vgl. Kap. II.1, S. 96).
- *Die Störung muss beeinflussbar sein.* Anhand der Erkenntnisse zur Förderung der phonologischen Informationsverarbeitung ist diese veränderbar (vgl. Kap. II.1.2.1).
- *Ein früher Interventionsbeginn sollte die Prognose verbessern.* Die phonologische Informationsverarbeitung kann bereits im Kindergarten trainiert werden (vgl. Kapitel II.1.2.1, sowie Walter 2002a, S. 39ff.).

Es stellt sich nun noch die Frage nach einem zeitökonomisch handhabbaren Diagnostikinstrument, das dem Kindergartenalltag gerecht werden kann, zugleich aber auch die Gütekriterien eines standardisierten Testverfahrens erfüllen kann. Ein wesentlicher Anspruch an ein solches Instrument wäre, dass mit seiner Hilfe Leistungen im Bereich der phonologischen Informationsverarbeitung eingeschätzt, Risiko- und Nicht-Risikokinder identifiziert und auf Grundlage der Ergebnisse dann Fördermaßnahmen abgeleitet werden könnten.
Eine curriculumbasierte Messung kann diesen Forderungen genügen. Sie dient, wie oben bereits geschrieben, als methodisches Fundament für das hier vorgestellte Diagnostikverfahren „Phonolino – eine curriculumbasierte Messung zur Prognose und Prävention von Schwierigkeiten im Schriftspracherwerb“. Phonolino hat das Ziel Risikokinder zeitökonomisch und spielerisch durch monatliche Messungen zu identifizieren und soll in dieser Studie zunächst auf seine Brauchbarkeit bzgl. der Prädiktoren und der Methodik mit einer Stichprobe von Kindern im Vorschulalter in Schleswig-Holstein validiert werden. Des Weiteren soll noch analysiert werden, ob das hier dargestellte Prognosemodell mit der phonologischen Informationsverarbeitung als Prädiktor einer Vorhersage von Lese- und Rechtschreibleistungen im ersten Schuljahr genügt, oder ob der weiter oben dargestellte indirekte Prädiktor der nonverbalen kognitiven Fähigkeiten in einem weiteren Prognosemodell eine bessere Vorhersageleistung treffen kann.

In den hier dargestellten Ableitungen der Forschungsfragen, Annahmen und Hypothesen wird auf die von Fuchs (2004, S. 189, vgl. auch Kap. III.6) beschriebenen Stufen der Evaluation einer curriculumbasierten Messung, ergänzt durch die Hinweise von Klauer (2011, S. 218, vgl. auch Kap. III.6), zurückgegriffen.
Die Forschungsfragen und Annahmen, die der ersten Stufe und somit den Gütekriterien für eine klassische Testtheorie zuzuordnen sind, werden in Kapitel V.1 beschrieben. Kapitel V.2 ergänzt Fragen und Hypothesen zur Änderungssensibilität nach Klauer (2011, S. 218), kann jedoch in dieser vorliegenden Untersuchung in Bezug auf die Stufe 2 von Fuchs (2004, S. 189) nur in Ansätzen geklärt werden. Somit kann die Beantwortung dieser Fragestellung nur einen Ausblick darstellen, hierzu bedarf es Folgeuntersuchungen, wie bspw. durch Klauer (2011, S. 218) beschrieben. Die Stufe 3 nach Fuchs (a. a. O.) kann in dieser Studie nicht hinreichend überprüft werden, wurde aber dennoch bei der Konzeption des CBM Phonolino in Kapitel IV. berücksichtigt. Eine Evaluation dieser Bemühungen muss in einer Folgestudie geleistet werden. Zusätzlich zu der beschriebenen Klassifikationsdiagnostik sollen in Kapitel V.3 die Fragestellungen und Hypothesen zu einer Optimierung der Vorhersage von schriftsprachlichen Leistungen durch Vorläuferfertigkeiten nachgegangen werden, um sicherzustellen, dass das vorgelegte Messverfahren ggf. durch die vermutete Komponente der nonverbalen kognitiven Leistungen der Kinder für bessere Vorhersagen ergänzt werden muss.

1 Ableitung der Forschungsfragen zur Überprüfung des Messverfahrens nach den Gütekriterien der klassischen Testtheorie

In den Vereinigten Staaten von Amerika sind die curriculumbasierten Messungen (CBM) im Zuge des „No Child Left Behind Acts“ als systematisch formatives Messverfahren erprobt worden und gelten als ein reliables und valides Messverfahren. Amerikanische Forschungen konnten schon wiederholt zeigen, dass die CBM sich im Bereich Lese- und Lesefähigkeit sowie Rechtschreibung als reliables und valides Messverfahren erwiesen haben (vgl. hierzu Kap. III.3). In deutschsprachigen Untersuchungen konnten zudem Komponenten der phonologischen Informationsverarbeitung als Indikatoren im Vorschulalter für spätere Schwierigkeiten im Lesen und Schreiben in der Schulzeit ermittelt werden (vgl. Kap. I.2.2).

In dieser Studie gilt es, diese Forschungserkenntnisse zur curriculumbasierten Messung sowie zur phonologischen Informationsverarbeitung zusammen zu führen und zu überprüfen, ob sich die curriculumbasierte Messung „Phonolino“ mit Inhalten der phonologischen

Informationsverarbeitung als stabiler Indikator, also als valides und reliables Messverfahren erweist. Dies entspricht der Überprüfung der Gütekriterien in der klassischen Testtheorie wie es von Fuchs (2004, S. 189) in der Stufe 1 zur Evaluation von curriculumbasierten Messungen gefordert wird. Hierbei geht es um die Eignung des Verfahrens als Screening für Risiken im Erwerb der Schriftsprache. Die Erstellung und Evaluation curriculumbasierter Messungen wie Phonolino sind besonderen Ansprüchen ausgesetzt. Die Evaluation von Lernfortschrittsdiagnostik wird in der Theorie jedoch noch offen diskutiert (vgl. z.B. Klauer 2011, S. 209ff. oder Wilbert & Linnemann 2011, S. 226).

Es werden die klassischen Gütekriterien der Objektivität, Reliabilität, Validität (vgl. Kap. I.2.1, S. 54) sowie der Differenzierungsfähigkeit zwischen guten und schlechten Leistungen der Kinder überprüft. Insbesondere hervorzuheben ist dabei die Parallelität der Messungen, die eine wesentliche Voraussetzung ist, Veränderungen der Kinder im Vorschulalter zu erkennen (vgl. Klauer 2011, S. 209 und Wilbert & Linnemann 2011, S. 228).
Zusätzlich muss das Verfahren in Bezug auf seine Validität geprüft werden. Hierbei wird auf Prüfung der konvergenten Validität oder auch Übereinstimmungsvalidität zurückgegriffen, um zu erwägen, ob die Messungen das gleiche Konstrukt messen, wie das etablierte Verfahren „Bielefelder Screening“ (vgl. Kap. I.2.2.1, S. 71). Da die curriculumbasierte Messung Phonolino die Form eines Screenings zur Vorhersage schriftsprachlicher Leistungen im ersten Schuljahr annehmen soll, ist eine Prüfung der prognostischen Qualität des Verfahrens von Nöten. Hierzu wird zum einen die prognostische Validität über Korrelationen bestimmt, besonders zentral für die Interpretation, wie in Kapitel I.2.1.4 (vgl. S. 69) bereits beschrieben, sind jedoch die klassifikatorischen Analysen, die eine Bewertung der Messungen und einen Vergleich zu anderen Verfahren zulässt. Zur Begründung der hier in den Fragestellungen verwendeten Kennwerte wird für eine detaillierte Erläuterung auf das Kapitel VI.5 – Statistische Verfahren – verwiesen.

Somit erschließen sich daraus die ersten vier Fragestellungen F1 bis F4 und die zugehörigen Annahmen (A) mit folgenden Schwerpunkten:

F1 Objektivität
F2 Differenzierungsfähigkeit/ Itemanalyse
F3 Reliabilität
F4 Validität

F1 Ist Phonolino ein objektives Messverfahren?

A1.1 Die Durchführungsobjektivität ist sichergestellt.
A1.2 Die Auswertungsobjektivität ist sichergestellt.
A1.3 Die Interpretationsobjektivität ist sichergestellt.

F2 Sind die Messungen von Phonolino brauchbar, also trennscharf und normalverteilt, ist die Schwierigkeit der Aufgaben angemessen und bestehen sie aus den drei Dimensionen „Pseudowörter nachsprechen", „Silben klopfen" und „Lautisolation"? (Itemanalyse)

A2.1 Die Häufigkeitsverteilungen der Items zu den acht Messzeitpunkten sind normalverteilt, sichtbar anhand einer glockenartigen, symmetrischen Kurve.
A2.2 Die Itemschwierigkeiten p_i der Items von Phonolino liegen zwischen p = .20 und p =. 80 (vgl. Bortz und Döring 1995, S. 199).
A2.3 Die Items aller Messungen über die acht Messzeitpunkte hinweg sind trennscharf und es gilt $r \geq .30$ (vgl. Bortz & Döring 2006, S. 220).
A2.4 Die Items laden auf die drei Faktoren „Pseudowörter nachsprechen", „Silben klopfen" und „Lautisolation".

F3 Ist Phonolino ein zuverlässiges Messverfahren, d.h. misst es exakt und genau das, was es messen soll?

A3.1 Die Koeffizienten der Paralleltestreliabilität des Gesamtwerts von Phonolino sowie der Items zu den acht Messzeitpunkten liegen im mindestens mittleren Bereich ($r \geq .30$). (Paralleltestreliabilität)
A3.2 Die Koeffizienten der Testhalbierungsreliabilität des Gesamtwerts von Phonolino sowie der Items zu den acht Messzeitpunkten liegen im mindestens mittleren Bereich ($r \geq .30$). (Split-Half-Reliabilität)
A3.3 Phonolino weist eine interne Konsistenz zu allen acht Messzeitpunkten mit einem Reliabilitätskoeffizienten im mindestens mittleren Bereich (Cronbachs Alpha $\propto \geq .50$) auf. (Interne Konsistenz)

F4 Sind die curriculumbasierten Messungen von Phonolino valide?

A4.1 Phonolino ist inhaltsvalide. (Inhaltsvalidität)
A4.2 Es bestehen mindestens mittlere Zusammenhänge ($r \geq .30$) zwischen den Messungen von Phonolino und der phonologischen Informationsverarbeitung gemessen durch das Bielefelder Screening (BISC), die im gleichen Zeitraum erhoben worden sind. (Übereinstimmungsvalidität)
A4.3 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten korrelieren signifikant im mindestens mittleren Bereich ($r \geq .30$) mit den Leseleistungen der Kinder im ersten Schuljahr, gemessen durch den Knuspel-L von Marx (1998). (Prognostische Validität - Lesen)
A4.4 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten korrelieren signifikant im mindestens mittleren Bereich ($r \geq .30$) mit den Rechtschreibleistungen der Kinder im ersten Schuljahr, gemessen durch den DRT1 von Müller (2003a). (Prognostische Validität - Rechtschreiben)
A4.5 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten tragen signifikant zur Varianzaufklärung im Lesen, gemessen durch den Knuspel-L von Marx (1998), bei. Für den RATZ-Wert gilt RATZ > 66%. (Prognostische Validität - Lesen)
A4.6 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten tragen signifikant zur Varianzaufklärung im Rechtschreiben, gemessen durch den DRT1 von Müller (2003a) bei. Für den RATZ-Wert gilt RATZ > 66%. (Prognostische Validität - Rechtschreiben)

2 Ableitung der Forschungsfragen zur Änderungssensibilität

Wird Phonolino und dessen einzelne Messungen über ein Screening-Verfahren hinaus auch als ein Verfahren zur Messung von Veränderung betrachtet, so reichen die Analysen der klassischen Testtheorie zu Validität, Reliabilität und Objektivität nicht aus. Es stellt sich die Frage, ob nicht nur der aktuelle Status, sondern auch die jeweiligen Veränderungen in den Leistungen durch das Verfahren festgestellt werden können. Weitergehende Analysen im Sinne der zweiten Stufe von L. S. Fuchs (2004, S. 189) werden in der vorlegenden Studie angestrebt, können jedoch nur nicht umfassend erfüllt werden, da die Datenbasis nicht jedwede Art von Berechnungen zulässt. Wie auch Bortz und Döring (2006, S. 248) beschreiben, ist die Analyse von Veränderungen die schwierigste Aufgabe in der Human- und Sozialwissenschaft. So hätte beispielweise der Unterschied zwischen den Werten der Splithalf-Reliabilität sowie der Paralleltestreliabilität nach Klauer (2011, S. 218) betrachtet werden können, wobei die Kennwerte der Testhalbierungsreliabilität aufgrund der

Veränderung deutlich höher sein müssten, als die Kennwerte der Paralleltestreliabilität. Die Splithalf-Reliabiltät hätte dazu jedoch nicht über die Messzeitpunkte hinweg, sondern innerhalb der Messzeitpunkte berechnet werden müssen, was aufgrund nur drei vorliegender Items nicht realisierbar war. Außerdem bezieht Klauer sich in seinem Vergleich auf eindimensionale Tests, wobei hier bei Phonolino eine dreidimensionale Messung vermutet wird.

Für die Messung von Veränderung sollte das Messverfahren die Leistungsentwicklung der Kinder sensibel abbilden. Damit einhergehend ist auch die Schwierigkeit der einzelnen Messungen von Relevanz für ein Verfahren, dass der Lernfortschrittsmessung dienen soll. So konstatiert Klauer (2011, S. 209):

> *„Angenommen, man würde unbeabsichtigt heute einen relativ schweren und morgen einen relativ leichten Test zur gleichen Thematik erheben, so würde vermutlich eine deutliche Leistungsverbesserung zu verzeichnen sein, ohne dass sich die Leistung realiter verbessert haben müsste. Eine Lernverlaufsdiagnostik muss gewährleisten, dass die einzelnen Tests, die da gegeben werden (erstens) dasselbe erfassen und (zweitens) auch stets gleich schwer sind."*

Aus diesem Grund empfiehlt Klauer (2014a, S. 10) für eine Lernverlaufsdiagnostik die Schwierigkeit nebeneinanderliegender Messungen miteinander zu vergleichen, da diese dem geringsten (und dennoch gewollten) Lernzuwachs der Probanden unterliegen. Die Schwierigkeit der einzelnen Messungen wird bereits mit der Annahme A2.2 beschrieben.

Nach Klauer (2014a, S. 10) müsste der Schwierigkeitsgrad der Messungen aufgrund des Lernzuwachses der Probanden kontinuierlich abnehmen. Somit empfiehlt er ausschließlich nebeneinanderliegende Items aufgrund ihrer Schwierigkeit miteinander zu vergleichen. In diesem Fall werden Mittelwerte miteinander verglichen um den Leistungsanstieg zu verzeichnen. Die Itemschwierigkeit selbst sollte dafür parallel betrachtet abnehmen.

Die Ergebnisse der Messungen sollten somit deutliche Unterschiede aufzeigen, da eine Verbesserung von 0,1 Punkten vom einen zum nächsten Messzeitpunkt nicht praktikabel und aussagekräftig genug ist (vgl. Voß & Hartke 2014, S. 93):

> *„Eine CBM erfüllt den Anspruch der Lernverlaufs-dokumentation zur zeitnahen Unterrichtsanpassung nicht, wenn ein durchschnittliches Kind lediglich eine Verbesserung um einen Punkt über einen langen Zeitraum (bspw. ein halbes Schuljahr) erzielt."*

Aufgrund der vorliegenden Datenbasis werden die Unterschiede der übernächsten Messungen betrachtet, da die Messzeitpunkte in den Kitas ggf. zu stark variieren könnten und nur drei bis vier Wochen zwischen den Erhebungen liegen. Hier kann kein großer Leistungszuwachs zu erwarten sein, sodass nebeneinanderliegende Messungen vermutlich eher parallele Ergebnisse abbilden, während weiter auseinanderliegende Messungen einen Leistungszuwachs zu verzeichnen haben.

Die folgenden Hypothesen (H) bilden sich auf der Grundlage der Aussagen von Klauer (2011, S. 218ff., 2014a, S. 10ff.).

F5 Erfasst Phonolino Veränderungen der Leistungen der phonologischen Bewusstheit zu den acht Messzeitpunkten?

H5.1 Die Mittelwerte des Gesamtwerts von Phonolino bzw. der drei Items steigen signifikant über die acht Messzeitpunkte hinweg mit Ausnahme direkt nebeneinanderliegender Messungen.
H5.2 Die Leistung der Probanden steigt durchschnittlich um mindestens einen Punkt zu jeder zweiten Messung.
H5.3 Je größer der Abstand zwischen den Messzeitpunkten ist, umso kleiner sind die Korrelationen zwischen dem Gesamtwert von Phonolino bzw. zwischen den Items zum jeweiligen Messzeitpunkt.

3 Ableitung der Forschungsfragen zur Optimierung des Prognosemodells

Wie in der theoretischen Darstellung bereits unter den endogenen und exogenen Faktoren dargestellt (vgl. Kapitel II.2), könnten noch weitere Faktoren (Kognition, Sprache, soziale Herkunft, ...) für eine bessere Prognose von Leistungen beim Schriftspracherwerb verantwortlich sein. Gleicherweise behilft sich das Bielefelder Screening etwaiger kognitiver Komponenten (z.B. Aufmerksamkeit, Schnelligkeit in Bezug auf den Abruf aus dem Langzeitgedächtnis) zur genaueren Vorhersage von Lese- und Rechtschreibleistungen im ersten Schuljahr (vgl. Kap. I.2.2.1, S. 71). In dieser Untersuchung soll ein mögliches Prognosemodell mit der phonologischen Informationsverarbeitung als Prädiktor durch einen weiteren Prädiktor, der nonverbalen kognitiven Leistung der Probanden, ergänzt und auf eine bessere Prognoseleistung hin überprüft werden. Hierfür werden die Faktoren der nonverbalen kognitiven Voraussetzungen durch den sprachfreien Intelligenztest CFT1

gemessen. Es bleibt zu zeigen, dass sich das Prognosemodell mit der phonologischen Informationsverarbeitung durch diese kognitiven Variablen erweitern lässt und so ggf. eine bessere Vorhersage möglich ist.

Dazu ergeben sich folgende Fragestellung und Hypothesen (H):

F6 Kann die Prognose von Lese- und Rechtschreibschwierigkeiten am Ende des ersten Schuljahrs durch die curriculumbasierten Messungen von Phonolino bei Hinzunahme des Prädiktors der nonverbalen kognitiven Leistung verbessert werden? (Prognosemodell 2)

H6.1 Die Güte des Prognosemodells 2 (mit den Prädiktoren der phonologischen Informationsverarbeitung und der nonverbalen kognitiven Leistung im Kindergarten) zeigt in Bezug auf das Modell-Fit (Chi-Quadrat-Wert) zur Vorhersage der Leseleistung am Ende des ersten Schuljahres eine signifikant ($p < .05$) bessere Passung des Modells. (Prognosemodell 2 - Lesen)

H6.2 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten sowie den kognitiven Leistungen der Kinder tragen mit einem RATZ-Wert > 66% signifikant zur Varianzaufklärung im Lesen bei. (Prognosemodell 2, klassifikatorische Analyse)

H6.3 Die Güte des Prognosemodells 2 (mit den Prädiktoren der phonologischen Informationsverarbeitung und der kognitiven Leistung im Kindergarten) zeigt in Bezug auf das Modell-Fit (Chi-Quadrat-Wert) zur Vorhersage der Rechtschreibleistung am Ende des ersten Schuljahres eine signifikant ($p < .05$) bessere Passung des Modells. (Prognosemodell 2 - Rechtschreiben)

H6.4 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten sowie den kognitiven Leistungen der Kinder tragen mit einem RATZ-Wert > 66% signifikant zur Varianzaufklärung im Rechtschreiben bei. (Prognosemodell 2, klassifikatorische Analyse)

Im Folgenden werden die Methode und der Untersuchungsaufbau der vorliegenden Studie beschrieben. Im Anschluss daran werden die hier verwendeten Untersuchungsinstrumentarien sowie die statistischen Verfahren zur Auswertung der Daten dargestellt.

VI. Methode und Untersuchungsaufbau

Die vorliegende prospektive Längsschnittstudie wurde im Zeitraum von 2008 bis 2010 durchgeführt. Die Studie wurde nicht im Rahmen eines hiesigen Forschungsprojekts der Universität, sondern als einzelner Forschungskomplex durch die Autorin umgesetzt. Bei den Testungen wurde sie zeitweise durch eine explizit geschulte Studentin unterstützt. Im nachfolgenden Abschnitt werden der methodische Aufbau sowie der Zeitplan (Kap. VI.1), die Untersuchungsgruppen und die pädagogischen Einrichtungen Kindertagesstätten und Schulen (Kap. VI.2) sowie die Durchführung der Untersuchung (Kap. VI.3) beschrieben. Anschließend werden die eingesetzten Untersuchungsinstrumentarien (Kap. VI.4) und die statistischen Verfahren, die zur Auswertung der erhobenen Daten genutzt wurden (Kap. VI.5), dargelegt und begründet.

1 Untersuchungsdesign und zeitlich-organisatorischer Ablauf

Es wurde eine prospektive Längsschnittstudie zur Evaluation der curriculumbasierten Messungen von Phonolino (vgl. Kap. IV) vom September 2008 bis Juli 2010 durchgeführt. Zu Beginn der Untersuchung im September 2008 wurden 258 Kinder im letzten Kindergartenjahr für die Evaluation von Phonolino vermerkt. Die Kinder kamen aus 15 Kindergärten aus dem Raum Schleswig-Holstein. Die Kindertagesstätten wurden postalisch über ein Rundschreiben an alle in der Universität Flensburg gelisteten Kindertagesstätten in Schleswig-Holstein im Umkreis von 1,5 bis 2 Autofahrstunden von Flensburg über das Projekt „Phonolino“ informiert. Interessierte Kindergärten meldeten sich per E-Mail oder telefonisch zurück.

Vom Oktober 2008 bis Juli 2009 wurden mit den teilnehmenden Vorschulkindern zu acht Messzeitpunkten monatlich in der ersten Woche des Monats die ca. fünf bis sieben Minuten lange CBM-Messungen durchgeführt. Um die Daten zu pseudonymisieren und damit den Datenschutz zu gewährleisten, wurde jedem Kind eine Codierungsnummer zugewiesen. Alle Unterlagen und Protokollbögen der CBM und der Fragebögen sowie die standardisierten Testverfahren wurden mit dieser Nummer versehen, sodass Testergebnisse nie mit dem Namen des Kindes in Verbindung gebracht werden konnten. Die Zuordnungen der Codierungen zu den Namen der Kinder wurden in den jeweiligen Kindergärten hinterlegt, deren Mitarbeiter wiederum nicht mit den Testergebnissen in Kontakt kommen. Um Verwechslungen zu vermeiden (z. B. Zahlendreher in den Codes) wurden auf den Testbögen zusätzlich Geburtsdaten und Geschlecht der Kinder bei der Testung vermerkt.

Nach Zusage der Kindertagesstätten wurden diese über die Ziele der Studie, den Zeitplan und den Datenschutz informiert. Ebenfalls wurden Elternbriefe versendet, die zudem eine Einverständniserklärung beinhalteten, die die Eltern der teilnehmenden Kinder unterschrieben haben. Die Eltern wurden informiert, dass sie die Ergebnisse ihrer Kinder nach Ablauf der Studie (Ende des ersten Schuljahres 2009/2010) bei der Studienleitung anfragen dürften. Die Rückmeldungen konnten schriftlich angefordert werden, zudem gab es die Möglichkeit ein persönliches Gespräch zu führen. Nur wenige Eltern nahmen diese Möglichkeit der Rückmeldung in Anspruch.

1.1 Zeitplan für die Durchführung von Phonolino

Im September 2008 wurden die Erzieherinnen bzgl. der curriculumbasierten Messung „Phonolino“ fortgebildet und instruiert. Im Oktober wurde das Messverfahren dann zum ersten Mal eingesetzt. Die Erzieherinnen führten die curriculumbasierten Messungen „Phonolino“ einmal im Monat im letzten Kindergartenjahr durch (vgl. Tab. 37).

Tab. 37: Zeitplan für die Phonolino-Messungen im Kindergarten

Zeitraum	Messung
1.10. – 7. 10.2008	Phono1
1.11. – 7.11.2008	Phono2
1.12. – 7.12.2008	Phono3
1.2 – 7.2.2009	Phono4
1.3 – 7.3.2009	Phono5
1.4. – 7.4.2009	Phono6
1.5. – 7.5.2009	Phono7
1.6. – 7.6.2009	Phono8
1.7. – 7.7.2009	Phono9

Die Messungen umfassten ca. zehn Minuten (je nach Begleitumständen), wobei die erste Sitzung im Oktober 2008 etwas länger andauerte, da die Kinder sich zunächst mit Phonolino vertraut machen mussten. An dieser Messung haben aus organisatorischen Gründen nur wenige Kindergärten teilgenommen, sodass die Ergebnisse in die Berechnungen nicht

aufgenommen worden sind. Es war den Erzieherinnen freigestellt, an welchem Tag der Woche sie die Messung durchführen, da viele Erzieherinnen dies aufgrund der großen Anzahl der Kinder nicht an einem Tag leisten konnten. Die Messungen sollten im letzten Kindergartenjahr einmal monatlich in den ersten sieben Tagen des Monats stattfinden. Im Januar fand aufgrund von Schließzeiten der Kindergärten keine Messung statt. An zwei Kindergärten wurden die Messungen von geschulten Studenten übernommen, da die Personalkapazitäten nicht ausreichten.

1.2 Zeitplan für weitere Testungen

Zusätzlich zu den CBM wurden im Zeitraum von Oktober 2008 bis Juli 2010 weitere Erhebungen durchgeführt (vgl. Tab. 38). Die Interviews zur Prädisposition wurden von den Erzieherinnen im Oktober geführt und ebenfalls versendet. Der erste Testzeitraum (T0) orientierte sich am Normierungszeitpunkt des Bielefelder Screenings (BISC) bis zehn Monate vor der Einschulung und erfolgte deshalb vom 13.10. bis zum 28.11.2008 in den Kindertagesstätten. In diesem Zeitraum wurden die Kinder mit dem CFT 1 und dem BISC 1 (bis zehn Monate vor Einschulung) getestet. Die Kinder wurden an unterschiedlichen Tagen überprüft, sodass kein Kind beide Tests hintereinander bearbeiten musste. Der CFT 1 wurde als Gruppentest mit zwei bis sechs Kindern durchgeführt. Nur in wenigen Krankheitsfällen (N = 12) wurde der CFT 1 als Einzeltest nachgeholt. Das Bielefelder Screening wurde, wie im Testmanual vorgesehen, als Einzeltest durchgeführt. Die Räumlichkeiten für die Testdurchführung waren vom restlichen Kindergartenbetrieb abgeteilt, so dass Ruhe und möglichst wenig Ablenkung gewährleistet waren.

Tab. 38: Zeitplan der Erhebung der Kriterien in den Jahren 2008-2010

Testphasen	Messzeitpunkt	Einrichtung	Kriterium
	Oktober 2008	Kindergarten	Interview Prädisposition
T 0	Okt. – Nov. 2008	Kindergarten	CFT 1
	Okt. – Nov. 2008	Kindergarten	BISC 1
T 1	Apr. – Mai 2009	Kindergarten	BISC 2
EINSCHULUNG IM SEPTEMBER 2009			
T 2	Jun. – Jul. 2010	Schule	DRT 1
	Jun. – Jul. 2010	Schule	KNUSPEL-L

Die zweite Testzeitspanne (T1) fand zum zweiten Normierungszeitpunkt des BISC (bis vier Monate vor der Einschulung) vom 30.03. bis 12.05.2009 im Kindergarten statt. Hier wurde der BISC 2 wieder als Einzeltest eingesetzt. Die Einschulung der Kinder war Anfang September 2009 (abhängig von der jeweiligen Schule).

In der Schule wurden dann von Juni bis Juli der Knuspel-L als Gruppentest (ohne Subtest 4, da dieser für die erste Klasse noch nicht geeignet ist) und der DRT 1 ebenfalls als Gruppentest durchgeführt (Testphase T2). Um ein Abschreiben zu verhindern wurden vom DRT 1 und auch vom Knuspel-L die jeweiligen Parallelformen genutzt. Die Gruppengröße beschränkte sich auf maximal 12 Kinder.

2 Beschreibung der Stichprobe

Im Folgenden werden die Stichprobe sowie die Kindertagesstätten und Schulen der vorliegenden Studie beschrieben.

2.1 Beschreibung der Grundgesamtheit

Die Stichprobe umfasste zu Beginn der Studie 258 Kinder im letzten Kindergartenjahr aus 15 Kindertagesstätten des Landes Schleswig-Holstein (vgl. Tab. 39).

Tab. 39: Überblick der Verteilung der Kinder über die 15 Kindertagesstätten in Schleswig-Holstein

Kita	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	Gesamt
Anzahl	16	20	11	33	10	13	11	15	22	14	23	11	27	13	19	258

Es nahmen jeweils alle Vorschulkinder der 15 Kindertagesstätten teil. Nur in zwei Kindertagesstätten willigten die Eltern von insgesamt drei Kindern von vornhinein nicht ein. Diese Kinder sind bei den 258 nicht mit einberechnet worden. Aufgrund eines Umzugs reduzierte sich die Gruppe nochmals um ein Kind auf 257. Der Altersdurchschnitt zu Beginn der Studie am 1.10.2008 betrug 68,66 Monate, also 5;9 Jahre. Die Kinder waren zwischen 4;10 und 6;8 Jahre alt. Die Stichprobe teilte sich in 137 Jungen (54,15 %) und 116 Mädchen (45,85 %). Die Erzieherinnen gaben bei 45 Kindern der Stichprobe an, dass diese in ihrem Alltag mit den Eltern täglich eine andere Sprache sprechen.

Im CFT 1 (Cattell, Weiß & Osterland 1997) betrug die intellektuelle Leistungsfähigkeit im Durchschnitt (vgl. Tab. 40) in der ersten IQ-Summe[24] 101,25 (zwischen 69 und 138) IQ-Punkte, in der zweiten IQ-Summe[25] 97,28 (zwischen 70 und 127) IQ-Punkte und in der dritten IQ-Summe 104,09[26] (zwischen 70 und 145) IQ-Punkte. Die IQ-Punkte konnten jedoch nur für 184 bzw. 185 (ein Kind hat den Test nicht vollständig durchgeführt, es konnte nur eine Summe berechnet werden) der Kinder ermittelt werden, da der CFT 1 für jüngere Kinder keine Altersnorm angeboten hat. In den unten folgenden Berechnungen wurden aus diesem Grund nur die Rohwerte für die Berechnungen genutzt.

Anhand von Elternfragebögen wurden einige sozioökonomische Daten der Kinder erfragt, die hier um die Stichprobe näher zu beschreiben zusammenfassend aufgeführt werden. Dabei ist zu beachten, dass, obwohl bei der Erstellung der Fragebögen dies berücksichtigt wurde, immer die Möglichkeit der sozialen Erwünschtheit (beispielsweise zur Anzahl der Kinderbücher im Haushalt) eine Rolle bei der Beantwortung der Frage spielt (vgl. Mummendey 1995, S. 159ff.). Die Eltern von 170 Kindern beantworteten den Fragebogen. 27 (15,9 %) der 170 Kinder leben bei einer alleinerziehenden Mutter. Die Kinder der Stichprobe haben zwischen keinem und sieben Geschwistern (im Durchschnitt 1,34) (vgl. Abb. 51).

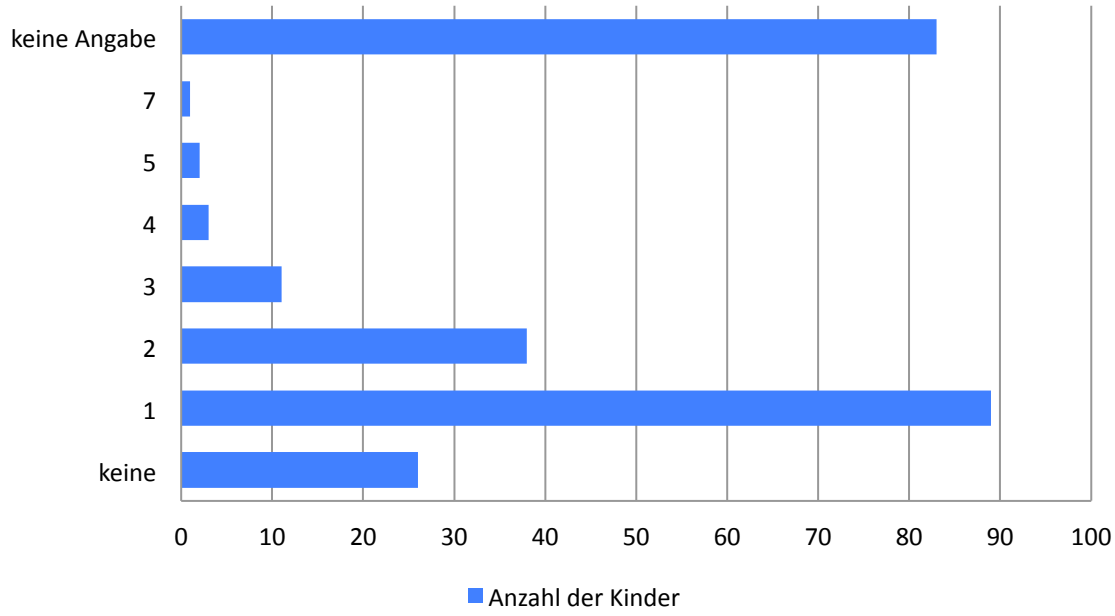


Abb. 51: Anzahl der Geschwister der Probanden

24 IQ-Summe 1 des CFT 1 beschreibt den Gesamtwert bezogen auf alle Subtests.
25 IQ-Summe 2 des CFT 1 beschreibt die Vorbedingungen des Denkens im Bereich der Wahrnehmung.
26 IQ-Summe 3 des CFT 1 beschreibt das Erkennen von Regelmäßigkeiten und Gesetzmäßigkeiten.

Ein Großteil der Eltern gab die Anzahl der Kinderbücher im Haushalt mit 26 bis 100 Büchern an. Nur sieben Kinder haben zu Hause auf weniger als zehn Kinderbücher Zugriff (vgl. Abb. 52).

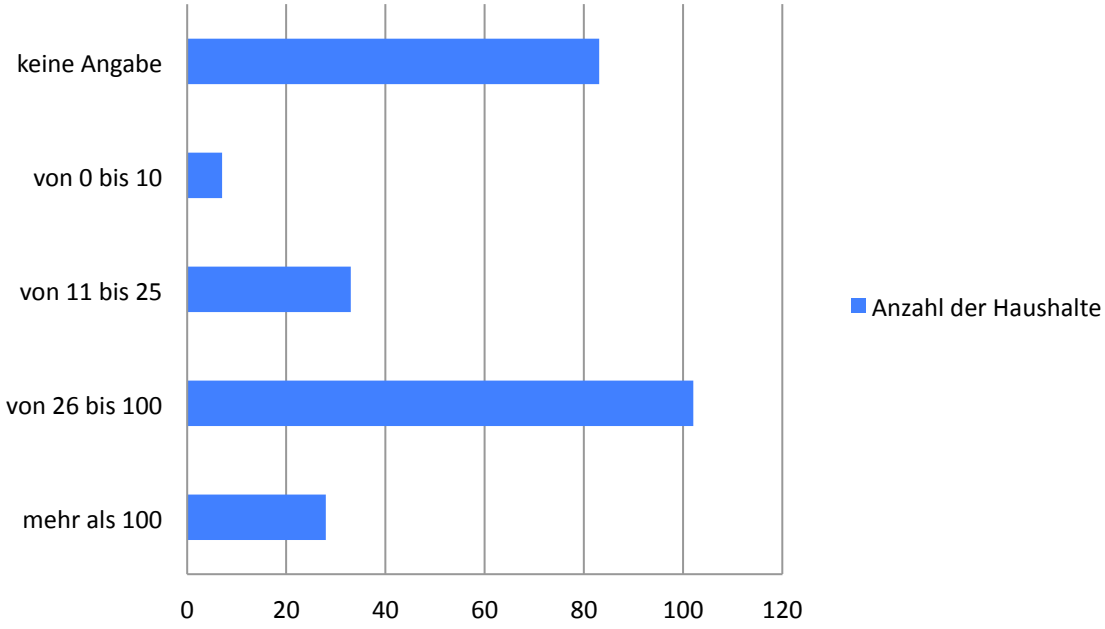


Abb. 52: Anzahl der Kinderbücher im Haushalt der Probanden

Die Erzieher gaben bei 38 der Kinder an, dass sie außerhalb des regulären Besuchs der Kindertagesstätten innerhalb der Einrichtung an einer Sprachförderung teilnehmen. Vier Kinder nehmen an anderweitigen Förderungen (Psychomotorik, Motorik, usw.) teil. 81 der Kinder nahmen an keiner Förderung teil, bei 136 Kindern gab es keine Angaben. Die Eltern von 57 Kindern gaben an, dass diese außerhalb des Kindergartens an einer Sprachförderung teilnehmen oder in den letzten zwei Jahren teilgenommen haben. 14 Kinder erhielten eine anderweitige Therapie und 99 Kinder wurden nicht zusätzlich gefördert. 78 Eltern machten keine Angabe.

29 Kinder der Stichprobe wurden mindestens einmal pro Woche in einer speziellen Vorschulgruppe auf die Schule vorbereitet, 194 Kinder haben an keiner speziellen Vorschulgruppe teilgenommen, zu 35 Kindern wurden durch die Kindertagesstätte keine Angaben gemacht.

Die Abfrage der Prädisposition in einer kurzen Interviewsituation mit den Eltern wurde von den Erziehern in einer in dem Schuljahr stattfindenden Elternsprechstunde oder am Tag des

Elternabends durchgeführt. Es wurde die Form des Interviews gewählt, da es bei einigen Eltern einer Erklärung bedarf, welche Bereiche die Lese-Rechtschreibschwäche umfassen. Außerdem konnte dadurch eine höhere Anzahl an Rückläufern gewährleistet werden, da die Eltern direkt angesprochen wurden. Die Ergebnisse wurden von den Erziehern in einer Tabelle notiert (vgl. Abb. 53). Vor- und Nachname wurde vor der Übermittlung an die Untersuchungsleiterin aus der Tabelle entfernt.

Gibt es Verwandte, die Legasthenie bzw. eine Lese-Rechtschreibschwäche haben?

Code	Vorname	Nachname	Nein	Ja	Mutter/Vater	Geschwister	Großeltern	weiter entfernte Verwandte
N1								
N2								
N3								
N4								
N5								
N6								
N7								
N8								
N9								
N10								
N11								
N12								
N13								

Abb. 53: Protokoll für das Interview zur Prädisposition

85 von 114 Kindern (74,5 %) haben im näheren Verwandtenkreis niemanden, der eine Lese-Rechtschreibschwäche hat oder bei dem diese diagnostiziert wurde. 19 Kinder (16,6 %) haben Eltern oder Geschwister, die von einer Lese-Rechtschreibschwäche betroffen sind, bei sieben Kindern (6,1 %) sind es weiter entfernte Verwandte. Eltern von 142 Kindern konnten keine Angabe machen, da sie sich nicht sicher waren oder sie keine Angabe machen wollten.

2.2 Beschreibung der Bildungsstätten

Acht der teilnehmenden Kindergärten waren in einer Stadt oder Kleinstadt in Schleswig-Holstein angesiedelt, die anderen sieben auf dem Land. Zwei der Kitas zählten insgesamt mehr als 100 Kinder in der Einrichtung, die anderen 13 betreuten zwischen 50 und 100 Kinder. In allen Kindertagesstätten wurde eine Betreuung von mindestens sechs Stunden angeboten, in zehn der fünfzehn Einrichtungen sogar eine ganztägige Betreuung von mindestens acht Stunden täglich. Alle Kindertagesstätten nehmen Kinder im Alter von drei bis sechs Jahren auf. Sieben davon nehmen auch Kinder unter drei Jahren auf. Vier weitere betreuen sowohl Kinder unter drei Jahren als auch im Nachmittagsbereich zusätzlich

Hortkinder aus der angrenzenden Grundschule. Sechs Kindertagesstätten nutzen zudem ein explizites Förderprogramm zur phonologischen Bewusstheit, nur eine Einrichtung nutzt auch ein Diagnostikinstrument zur Erfassung der phonologischen Bewusstheit.

Die Kinder der Untersuchungsgruppe wurden an 41 unterschiedlichen Schulen eingeschult. Beim Übergang in die Schule sind neun Kinder in der Kindertagesstätte verblieben und 18 Kinder in ein anderes Bundesland gezogen oder die zugehörige Schule war nicht mehr auffindbar. Somit wurde die Stichprobengröße beim Übergang in die Schule um insgesamt 27 Kinder verringert. 231 Kinder wurden eingeschult und begleitet. Von den 231 Kindern besuchten 115 eine Schule in der Stadt und 116 Kinder eine Schule auf dem Land. Weitere Verluste von Probanden werden im nächsten Abschnitt erläutert.

2.3 Verluste von Probanden

Die Anzahl der an den einzelnen oben aufgeführten Testungen teilnehmenden Kinder schwankt aufgrund der Länge und der Vielfalt der Testverfahren (vgl. Tab. 41) sowie der neun curriculumbasierten Messungen (vgl. Tab. 42) stark. Eine große Anzahl an Drop-Outs ist für eine längsschnittlich angelegte Studie nicht unüblich.

Tab. 40: Beschreibung der Stichprobe zu den Testzeitpunkten T0, T1, T2 in Kindergarten und Schule

Untersuchungs-zeitpunkte	Testinstrument	N
T0	BISC 1	238
	CFT 1	233
T1	BISC 2	223
T2	DRT 1	216
	Knuspel-L	217

Die erste Messung im Kindergarten (Phono 1) wurde nicht an allen Kindergärten durchgeführt oder protokolliert, da der Termin für einige Erzieher zu kurzfristig war und sie sich zunächst noch in das Material einarbeiten wollten. Dadurch ist die Größe der Stichprobe zum Zeitpunkt Phono 1 im Vergleich zu den anderen Messzeitpunkten stark reduziert (vgl. Tab. 42). Wie in jeder längsschnittlich angelegten Studie wird hier an der Anzahl der Unter-

suchungspersonen zu den unterschiedlichen Testzeitpunkten ersichtlich, dass einige Kinder zeitweise nicht teilnehmen konnten oder gänzlich aus der Studie ausgeschieden sind.
Das vorübergehende Ausscheiden aus einigen Testungen hatte unterschiedliche Gründe. Die Kinder waren zum Beispiel über einen längeren Zeitraum krank, sind umgezogen oder haben sich in einer längeren Therapie aufgehalten. Die geringe Gruppengröße nach dem Wechsel in die Schule lässt sich zum einen damit erklären, dass nicht bei allen Kindern nachverfolgt werden konnte, an welche Schule sie gegangen waren (einige Eltern hatten ihre Kinder bei der zuständigen Grundschule abgemeldet, um sie bei einer anderen Grundschule zu melden, dies wurde jedoch dem Kindergarten nicht mitgeteilt) oder sie waren mit ihren Eltern in einen Ort, der sich außerhalb von Schleswig-Holstein befand, umgezogen. Davon waren 13 Kinder betroffen. Außerdem waren neun Kinder in der Kindertagesstätte verblieben, die entweder zurückgestellt wurden oder zu den sogenannten „Kann-Kindern" gehörten, deren Eltern sich noch entscheiden wollten, ob sie ihre Kinder frühzeitig mit fünf Jahren einschulen wollen und sie bereits in die Schule gehen dürfen.

Tab. 41: Beschreibung der Stichprobe zu den neun Messzeitpunkten der curriculumbasierten Messung im Kindergarten

CBM	Phono1	Phono2	Phono3	Phono4	Phono5	Phono6	Phono7	Phono8	Phono9
N	128	215	211	204	224	197	205	216	212

Der Hauptteil der Untersuchung besteht in der Überprüfung der Phonolino-Messungen. Fehlende Werte können Fehlerquellen in Studien verursachen, die Ergebnisse verzerren oder die Effizienz der statistischen Verfahren verringern. Aus diesem Grund werden diese hier genauer betrachtet. Damit keine Fälle mit eingeschlossen werden, an denen Kinder an nur einer Messung teilgenommen haben, sollen alle Fälle (nach Wirtz 2004, S. 110ff.) ausgeschlossen werden, die an nur 30 % (oder weniger) der Messungen teilgenommen haben.

Diese auszuschließenden Datensätze werden durch die MVA-Analyse mit dem MCAR-Test (missing completly at random) nach Little (vgl. Little & Rubin 1987, S. 39f.) dahingehend überprüft, ob diese Werte zufällig fehlen. Die Eliminierung von Fällen darf nur unter der Annahme, dass es aufgrund von Abhängigkeiten nicht zu Verzerrungen kommen kann, durchgeführt werden. Die Nullhypothese des MCAR-Tests sagt aus, dass die Daten völlig zufällig fehlen. Dies ist der Fall, wenn das Muster der fehlenden Daten nicht abhängig von den anderen Datenwerten ist (vgl. IBM 2011, S. 48). Ist der Test nicht signifikant, so kann davon ausgegangen werden, dass die Daten zufällig fehlen. Nach dem Ausschluss der Fälle

werden die zwei Gruppen in Bezug auf die deskriptiven Daten auf signifikante Unterschiede hin überprüft. Die Gruppenunterschiede zwischen den ausgewählten und den ausgeschlossenen Fällen werden mit dem Kruskal-Wallis-Test („H-Test“) ermittelt. Der H-Test überprüft, ob sich mehr als zwei unabhängige Stichproben in ihrer zentralen Tendenz unterscheiden. Dieser wurde gewählt, da er sowohl für ordinal- als auch für intervallskalierte Daten eingesetzt werden kann und die Daten nicht zwingend normalverteilt sein müssen, wie es für eine einfaktorielle Varianzanalyse als Voraussetzung gelten würde (vgl. Univ. Zürich 2016c). Beim Kruskal-Wallis-Test wird nicht mit den Werten selbst gerechnet, sondern es werden Rangdaten verwendet, mit denen der Test durchgeführt wird. Dabei werden absolute Abstände nicht genauer betrachtet. Da in diesem Fall nur ein Gruppenunterschied für zwei Gruppen (aussortierte Fälle und Fälle, die für die Berechnungen ausgewählt wurden) berechnet wird, wird auf einen exakten (anstatt auf einen asymptotisch-chi-quadrat-verteilten) Test zurückgegriffen (vgl. a. a. O.). Zeigt der Test, dass keine signifikanten Gruppenunterschiede bestehen, so kann davon ausgegangen werden, dass eine zufällige Eliminierung von Fällen stattgefunden hat.

Aufgrund einer zu hohen Anzahl von missing data pro Datensatz ist die completly-case analysis, also der Ausschluss aller Fälle, die einen oder mehr fehlende Daten aufweisen, ausgeschlossen, da die Stichprobe dafür nicht groß genug ist (vgl. Runte o.J., S. 13). Ein Nachteil dabei ist, eine sich je nach Auswertungsverfahren beständig verändernde Stichprobengröße. In dieser Untersuchung sind in den statistischen Daten N = 257 Fälle aufgenommen. Der Fall 118 wurde aufgrund vollständig fehlender Werte zunächst entfernt (das Kind ist schon vor Beginn der Studie umgezogen). Auf eine Imputation der Daten wurde verzichtet, stattdessen wurde auf ein Eliminierungsverfahren zurückgegriffen. Im Anschluss wurden die Fälle entfernt, die an weniger als 30% der Phonolino-Messungen teilgenommen haben (Wirtz 2004, S. 110ff.). Hierbei wurden nur die Phonolino-Messungen Phono 2 bis Phono 9 berücksichtigt, da an der Messung Phono 1, aus organisatorischen Gründen noch nicht alle Kindertagesstätten teilnehmen konnten und diese nicht in die Auswertungen übernommen wird. Somit wurden 43 Fälle entfernt und die Stichprobengröße hat sich auf N* = 214 Fälle um 16,7 % verringert. Somit wird in dieser Studie der fallweise mit dem paarweisen Ausschluss kombiniert. Es werden fallweise alle Probanden mit dem kompletten Datensatz aus der Stichprobe entfernt, die an 30 % der Phonolino-Messungen nicht teilgenommen haben, dennoch sind weiterhin Probanden ohne kompletten Datensatz aller Komponenten in der Stichprobe enthalten, die nur bei einigen Testungen gefehlt haben. Dies verhindert, dass Fälle aufgenommen wurden, die kaum an Messungen teilgenommen haben, verhindert aber gleichzeitig auch, dass der Stichprobenumfang sehr stark sinkt.

Für den fallweisen Ausschluss bestimmter Daten ergibt die MVA-Analyse für die Werte von Phonolino (Phono 2 bis 9 zu allen acht Messzeitpunkten) mit dem MCAR-Test (missing completly at random) nach Little (vgl. Little & Rubin 1987, S. 39f.; vgl. Anhang 5) einen Signifikanzwert von .128 und ist damit nicht signifikant. Es kann davon ausgegangen werden, dass die Daten zufällig fehlen und der oben beschriebene Ausschluss von Fällen zulässig ist.

Das Durchschnittsalter der 43 aussortierten Fälle beträgt 67,82 Monate, also 5,65 Jahre zu Beginn der Studie, das Durchschnittsalter der in der Stichprobe verbleibenden 214 Kinder liegt bei 5,73 Jahren. Die Eigenschaften der aussortierten Fälle wurden durch den Kruskal-Wallis-Test auf Signifikanz der beiden Gruppen überprüft. Der Test ergab für die ausgewählten Variablen „Alter“, „Geschlecht“, „Migrationshintergrund“, „Förderung in der Vergangenheit“, „Risikopunkte im Bielefelder Screening“ (Risiko für Schwierigkeiten in der phonologischen Informationsverarbeitung), „CFT1“ (kognitive Fähigkeiten), „DRT1“ (Rechtschreibleistung im ersten Schuljahr) und „Knuspel-L“ (Leseleistung im ersten Schuljahr) keine Signifikanz (vgl. Tab. 42). Es kann somit davon ausgegangen werden, dass die in der Studie verbleibenden Kinder nicht signifikant von der ursprünglichen Grundgesamt verschieden sind, da die ausgeschlossenen Kinder keine besonderen Eigenschaften aufzeigten. Dennoch ist zu berücksichtigen, dass es sich bei der Grundgesamtheit selbst aufgrund der aktiven Zusage der Kindertagesstätte zur Teilnahme an der Studie und auch aufgrund der örtlichen Begrenzung auf Schleswig-Holstein nicht um eine Zufallsstichprobe handelt.

Tab. 42: Ergebnisse des Krukal-Wallis-Tests zur Überprüfung auf Gruppenunterschiede zwischen den eliminierten Fällen und den in den Berechnungen verbleibenden 214 Fällen

	Alter	Geschlecht	Migrations-hintergrund	Förderung	BISC1	BISC2	CFT1	DRT1	Knuspel-L
Chi-Quadrat	1,722	,094	,775	,205	,074	,028	,428	,265	,214
df	1	1	1	1	1	1	1	1	1
p (asymp)	,189	,759	,379	,651	,786	,867	,513	,607	,644
p (exakt)		,862	,497	,722	,787	,868			

Anmerkung: p (asymp)= asymptotische Signifikanz, p (exakt) = exakte Signifikanzen, konnten aufgrund von mangelnden Arbeitsspeicherkapazitäten nicht für metrische Skalen berechnet werden, Förderung = Hinweis aus dem Erzieherfragebogen zu Logopädie, Ergotherapie usw., BISC 1 = Bielefelder Screening, Risikopunkte in der phonologischen Informationsverarbeitung 10 Monate vor der Einschulung, BISC 2 = Bielefelder Screening, Risikopunkte in der phonologischen Informationsverarbeitung 4 Monate vor der Einschulung, CFT1 = kognitive Leistung (Rohwert) zu Beginn des letzten Kindergartenjahrs, DRT1 = Rohwert (Wortfehler), Rechtschreibleistung am Ende des ersten Schuljahres, Knuspel-L = Klassennorm, Leseleistung am Ende des ersten Schuljahres

Aufgrund einer zu kleinen Anzahl von Probanden wurde hier gegen einen fallweisen Ausschluss entschieden, der voraussetzt, dass nur Kinder, die an allen Messungen von Phonolino teilgenommen haben, Teil der Stichprobe sind. Es wurden Kinder mit einer Teilnahme an weniger als 30 % der Messungen von der Stichprobe ausgeschlossen. Da die statistische Überprüfung des Ausschlusses die Fälle als zufällig fehlende Werte bewertet hat, weicht die ausgewählte Stichprobe nicht signifikant von der Grundgesamtheit ab. Nach Rubin (1976, zit. in Runte o.J., S. 3) liegen die Daten als missing completly at random vor, wobei das Fehlen der Daten weder von der Ausprägung der fehlenden, noch der existierenden Merkmale abhängig ist. Die Ausfälle sind somit zufällig und „verzerren das Untersuchungsergebnis damit in der Regel kaum oder gar nicht“ (vgl. Runte o.J., S. 4). Die Ausfälle sind in größten Teilen anhand von Krankheit oder Abwesenheit der Kinder zu erklären.

Die Größe der Stichprobe (N*=214) ist als mäßig zu bewerten, d.h. eine Wiederholung der Untersuchung mit einer deutlich größeren Stichprobe wäre sinnvoll. Auf eine Imputation oder ein Ersetzen fehlender Werte wurde aufgrund der möglichen Verzerrung der Auswertungen verzichtet. Dennoch werden nicht ausschließlich die Kinder, die an allen Messungen von Phonolino teilgenommen haben, herausgenommen, da diese dann an den Testungen durch die weiteren standardisierten Testverfahren hätten komplett teilnehmen müssen (CFT 1, BISC 1 + 2 usw.), aufgrund dieser Voraussetzungen hätte sich die Stichprobe sonst enorm reduziert. Somit findet bei jeder Auswertung erneut eine Auswahl aus N = 214 Kindern statt, die an den jeweiligen Messungen teilgenommen haben. Dies mindert gleichzeitig die Stabilität der Daten, da Extremwerte in einigen Auswertungen Berücksichtigung finden und in anderen nicht, scheint jedoch eine bessere Aussagekraft zu besitzen, als eine stark reduzierte Stichprobe (mit Kindern, die an allen Phonolino-Messungen, sowie an den Testungen des CFT1 1, BISC 1 und BISC 2 sowie DRT 1 und Knuspel-L, teilgenommen haben).

2.4 Deskriptive Analyse der Untersuchungsgruppe

Als erste Analyse der Daten werden die deskriptiven Statistiken der nun durch einige Fälle bereinigten Stichprobe dargestellt (vgl. Anhang 8). Die 214 Kinder der Stichprobe sind zu Beginn der Untersuchung im Jahr 2008 zwischen 58 und 80 Monate alt. Der Mittelwert beträgt 68,81 Monate, also 5,73 Jahre. Die Stichprobe besteht nunmehr aus N* = 214 Kindern (über den Verbleib der anderen Kinder, s. Kap. VI.2.3 – Verluste von Probanden). Davon sind 115 Kinder männlichen und 99 Kinder weiblichen Geschlechts (vgl. Abb. 54).

40 Kinder haben einen Migrationshintergrund, das ist ein Anteil von 18,7 %. 174 Kinder (81,3 %) sprechen zu Hause ausschließlich deutsch (vgl. Abb. 54).

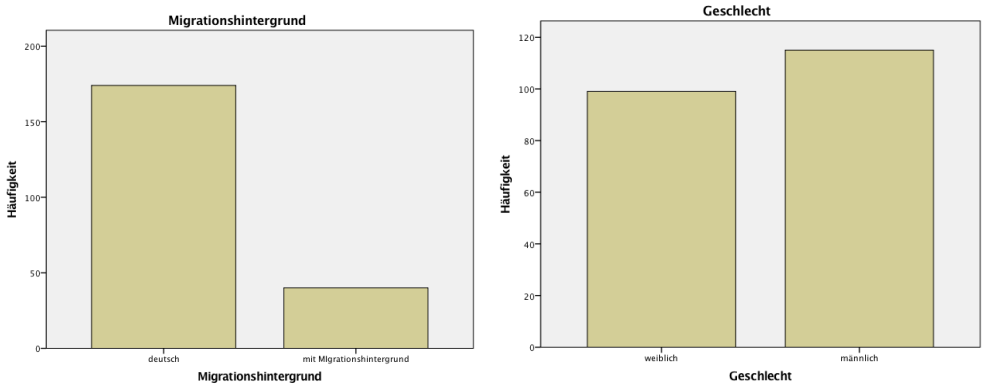


Abb. 54: Balkendiagramme zum Geschlecht und zum Migrationshintergrund der Kinder der vorliegenden Stichprobe von N= 214

Die Eltern von N = 150 Kindern gaben an, dass 86 Kinder bisher keine Förderung oder Therapie erhalten haben, 51 haben bereits an einer Sprachtherapie teilgenommen und 13 an anderweitigen Förderungen. Unter N = 150 Kindern waren 25 Kinder, die mit alleinerziehenden Müttern leben. 125 lebten in einem Haushalt mit Vater und Mutter (oder jeweils neuem Partner) gemeinsam. 50,9 % Prozent der Kinder wurden in der Stadt eingeschult, 44, 9% auf dem Land, bei 4,7 % wurde keine Angabe gemacht.

Die Prävalenzrate bezüglich der Schwierigkeiten im Lesen und Rechtschreiben konnte erst zum Ende der ersten Klasse festgestellt werden (vgl. Tab. 43).

Tab. 43: Prävalenzrate der Lese-Rechtschreibleistungen der Probanden am Ende der ersten Klasse (gemessen durch den DRT 1 und Knuspel-L)

Cut-Score / Kriterium	PR ≤ 15 oder T ≤ 40	PR ≤ 10 oder T ≤ 37
Rechtschreibleistung am Ende der 1. Klasse (DRT 1)	26,8 %	19,2 %
Leseleistung am Ende der 1. Klasse (Knuspel-L)	24,8 %	17,3 %

Anmerkung: PR = Prozentrang der Skala des DRT 1, T = T-Wert der Klassennorm des Knuspel-L

Die gewählte Stichprobe wies je nach gewähltem Cut-Off-Wert eine sehr hohe Prävalenz für Lese- und Rechtschreibschwierigkeiten auf (vgl. Kap. I.1.3.1) und ist damit eine relativ schwache Stichprobe (vgl. Anhang 8).

3 Durchführung der Untersuchung

Die Untersuchung hat im Zeitraum von Sommer 2008 bis Sommer 2010 stattgefunden. Im nächsten Abschnitt wird die Durchführung und Auswertung der CBM Phonolino beschrieben und im Anschluss die Vorgehensweise der Auswertung bei den weiteren Untersuchungsinstrumentarien.

3.1 Durchführung und Auswertung von Phonolino

Jeder Erzieher hat an einem zweistündigen Informationsabend mit einer theoretischen Aufarbeitung des Konstrukts der phonologischen Informationsverarbeitung sowie einer vierstündigen Fortbildung mit Rollenspielen und aktiven Gruppengesprächen teilgenommen. An den Fortbildungen waren insgesamt 33 Erzieher von den 15 teilnehmenden Kindergärten beteiligt. Die Fortbildungen wurden aufgrund der Gruppengröße an zwei unterschiedlichen Terminen durchgeführt. Die Inhalte der beiden Fortbildungen waren identisch. Im Fokus der Betrachtung standen die Art und Weise der Protokollierung und das Einüben der vorgegebenen Instruktionen sowie der Umgang mit der Handpuppe Phonolino. Auch die beiden Studierenden, die an zwei der Kindertagesstätten die Förderung durchgeführt haben, waren bei der Schulung anwesend. Im Oktober und November 2008 fanden während der Phonolino-Messungen Besuche durch die Studienleitung in den teilnehmenden Kindertagesstätten statt, um eine möglichst einheitliche Durchführung an den 15 Kindertagesstätten zu gewährleisten. Zwischen den Messungen wurden regelmäßig Telefongespräche zwischen den durchführenden Personen und der Autorin geführt.

Alle curriculumbasierten Messungen fanden in separaten, möglichst störungsfreien Räumen statt, um den Kindern eine Atmosphäre zu schaffen, bei der sie sich auf die Aufgaben konzentrieren konnten. Es wurde empfohlen, die Messungen am Vormittag zwischen 8 und 12 Uhr durchzuführen. In Ausnahmefällen fand die CBM nachmittags statt. Kinder, die in der jeweiligen Woche krank waren, durften die CBM in der Woche danach nachholen. Wenn der Abstand größer war, wurde die Messung ausgelassen und das Kind nahm erst an der nächsten regulären Messung wieder teil. Zu allen drei Items einer Messung gibt es eine Anfangsaufgabe, bei der die Konzentration der Kinder erneut gebündelt werden soll und überprüft wird, ob sie die Aufgabenstellung ausreichend verstanden haben.

Die Protokollierung wurde wie im Handbuch beschrieben von den Erziehern selbst ausgeführt. Die Auswertung wurde von der Verfasserin anhand der Protokollierung in den Phonolino-Heften durchgeführt. Die Auswertung der Rohpunkte erfolgte bei den Items 1 und

2 („Pseudowörter nachsprechen“ und „Silben klopfen“) anhand des ganzen Wortes. Das heißt, die Rohpunktsummen ergaben sich aus der Anzahl der richtig wiedergegebenen Wörter (ein richtig nachgesprochenes Wort = ein Rohwertpunkt) oder der richtig geklopften Wörter (ein richtig geklopftes Wort = ein Rohwertpunkt). Bei Item 1 galt jedes Wort als falsch, in dem mindestens ein Laut falsch nachgesprochen wurde. Dabei wurden phonetische Schwierigkeiten (z. B. Sigmatismus) jedoch nicht als falsch gewertet. Falls die Kinder in Item 2 nur die Anzahl der Silben verbal nannten, wurden sie einmalig nochmals dazu aufgefordert, die Anzahl der Silben zu klopfen. Bei der Einführungsaufgabe des Items 2 wurden die Kinder noch darauf hingewiesen, dass sie das zu analysierende Wort beim Klopfen mitsprechen sollten. Falls dies nicht stattgefunden hat, sie aber die richtige Anzahl an Silben geklopft hatten, wurde das Wort dennoch als richtig gewertet. Die Verbesserung bei der Einstiegsaufgabe dient ausschließlich der verbalen Hilfestellung und der Konzentration der Kinder, ist aber nicht nötig, wenn die Kinder bereits die Fähigkeit besitzen, ohne zu sprechen die richtige Anzahl an Silben zu klopfen. Bei Item 3 wurde jede richtige Wortstellung (Anlaut oder Auslaut) mit einem Rohwertpunkt bewertet.

Die Person, die die Messung durchführte, hat die Instruktionen wortwörtlich den Instruktionskarten entnommen. Die „Wörter“ wurden von einer CD vorgespielt, somit ist eine Beeinflussung durch die Aussprache oder die Betonung der durchführenden Personen auszuschließen. Die Instruktionen werden wörtlich wiedergegeben. Die Durchführung in Bezug auf die Handpuppe kann je nach durchführender Person ein wenig variieren (Betonung, Tonalität, Bewegung der Puppe), ist aber geprägt von den wörtlich vorgegebenen Instruktionen. Die Auswertung wurde von der Autorin durchgeführt, den durchführenden Personen wurde aber eine klare Notation in dem Phonolino-Protokollheft (s. Anhang 2) vorgegeben.

3.2 Durchführung und Auswertung der standardisierten Tests und Fragebögen

CFT 1 und BISC 1 (zehn Monate vor der Einschulung) und BISC 2 (vier Monate vor der Einschulung) wurden ausschließlich vormittags von der Verfasserin dieser Studie oder von Studierenden in den Kindertagesstätten durchgeführt. Mit Ausnahme von zwei Kindern wurden alle Tests an unterschiedlichen Tagen durchgeführt, damit die Kinder nicht überlastet werden. Bei den zwei Kindern war der Anfahrtsweg sehr weit und sie hatten am Tag der vorherigen Testungen gefehlt. Hier wurde dann aus zeitökonomischen Gründen entschieden, den BISC 1 vormittags und den CFT 1 dann am Nachmittag nach einer Pause durchzuführen.

Bis auf eine Ausnahme, in der ein Kind aus organisatorischen Gründen zu Hause (bzw. im wohnhaften Heim) getestet wurde, wurden alle Kinder in den Klassen- oder Gruppenräumen ihrer Schule getestet. In der Schule fanden die beiden Tests (Knuspel-L und DRT 1) ebenfalls an unterschiedlichen Tagen statt. Die Kinder wurden in Kleingruppen aufgeteilt und parallel zum Unterricht am Vormittag in einem möglichst ruhigen Raum getestet. Die Tests haben oftmals in Einzelmessungen oder zu zweit stattgefunden, da die Kinder einer Kindertagesstätte zum Teil an unterschiedlichen Schulen eingeschult worden sind. Insgesamt wurden die Kinder aus den 15 Kindertagesstätten auf 41 Schulen verteilt.

Die Ergebnisse der standardisierten Verfahren wurden nach Anweisungen der Handbücher der einzelnen Verfahren durchgeführt und von der Autorin ausgewertet. Im Codierungsplan ist ersichtlich, welche Kennwerte der standardisierten Verfahren (CFT 1, BISC, DRT 1, Knuspel-L) übernommen worden sind (vgl. Anhang 4).

4 Untersuchungsinstrumentarien

Zur Evaluation der curriculumbasierten Messung „Phonolino“ wurden für die vorliegende empirische Untersuchung vielfältige Messinstrumente eingesetzt. Zur Überprüfung der Leistungen und Einflussfaktoren der Kinder werden zunächst die eingesetzten Diagnostikinstrumente im Kindergarten (BISC, CFT 1, ein Interview) vorgestellt. Im Anschluss werden die Testverfahren präsentiert, die zur Überprüfung der Lese- und Schreibfähigkeiten im Schulalter genutzt worden sind (DRT 1, Knuspel-L).

4.1 Eingesetzte Testverfahren im Vorschulalter

Die Probanden wurden mit dem Bielefelder Screening (BISC), einem Test zur Erfassung vorschulischer Fähigkeiten in der phonologischen Informationsverarbeitung, getestet (vgl. Kap. I.2.2.1). Die nonverbale Intelligenz wurde außerdem mit dem CFT 1 gemessen. Für die Messung der basalen Voraussetzungen im Vorschulalter und zur Überprüfung der Validität von Phonolino wurden die Fähigkeiten der Kinder zu mehreren Testzeitpunkten im letzten Kindergartenjahr erfasst. Parallel dazu wurden die Phonolino-Messungen durchgeführt.

4.1.1 Bielefelder Screening (BISC)

Das Bielefelder Screening (BISC) (Jansen, Mannhaupt, Marx & Skowronek 2002) wurde bereits in Kapitel I.2.2.1 ausführlich dargestellt. Es wurde in dieser Studie zur Überprüfung der Leistung der Kinder in der phonologischen Informationsverarbeitung zehn Monate (BISC 1) und vier Monate (BISC 2) vor Schulbeginn eingesetzt. Mit dem BISC werden spezifische Schriftsprachvoraussetzungen wie die phonologische Bewusstheit, der schnelle Abruf aus dem Langzeitgedächtnis, das phonetische Rekodieren im Kurzzeitgedächtnis sowie die visuelle Aufmerksamkeitssteuerung überprüft. Der BISC wird als Einzeltest durchgeführt und dauert je nach Fähigkeiten des Kindes ca. 30 bis 40 Minuten. Das Verfahren wurde mit 1 120 Vorschulkindern geeicht (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002, S. 32f.) und durch eine Längsschnittstudie ab dem Jahr 1986/1987 auf seine Vorhersagekraft hin untersucht, die später dann mit der Münchener LOGIC-Studie kreuzvalidiert wurde. 92,2 % der Risikoeinschätzungen durch den BISC ergaben sich als richtig. Im Jahr 2006 erfolgte eine Testrevision, deren Ergebnisse weniger optimistisch ausfielen als die Ergebnisse in der ersten Studie (vgl. Marx & Weber 2006, S. 254ff.). Die hohen RATZ-Indizes der ersten Untersuchung konnten nicht repliziert werden. Nähere Informationen zu den Untersuchungen sind in Kapitel I.2.2.1 aufgeführt.

Das BISC wurde gewählt, da es ein validiertes und anerkanntes Messinstrument zur Überprüfung von vorschulischen Fähigkeiten von Kindern im Bereich der Schriftsprache darstellt und es außerdem zu zwei verschiedenen Messzeitpunkten durchgeführt werden kann (zehn Monate und vier Monate vor der Einschulung), was sich für das Untersuchungsdesign als vorteilhaft gestaltete (vgl. Kap. V).

4.1.2 CFT 1 – Grundintelligenztest Skala 1

Mit dem CFT 1 – Grundintelligenztest Skala 1 (vgl. Weiß & Osterland 1997) kann die Grundintelligenz eines Kindes bestimmt werden. Als Grundintelligenz beschreiben Weiß und Osterland (a. a. O., S. 4) die „Fähigkeit des Kindes, in neuartigen Situationen und anhand von sprachfreiem, figuralem Material, Denkprobleme zu erfassen, Beziehungen herzustellen, Regeln zu erkennen, Merkmale zu identifizieren und rasch wahrzunehmen.“ Der CFT 1 zeigt auf inwieweit die Kinder schon dazu in der Lage sind, nonverbale Probleme zu erfassen und zu lösen und misst damit die nonverbale Intelligenz. Im ersten Teil wird die visuelle Wahrnehmungsfähigkeit und im zweiten Teil das beziehungsstiftende Denken gemessen.

Der CFT 1 ist ein Gruppentest mit zwei Pseudoparalleltests A und B. Die Durchführungsdauer für Kindergartenkinder beträgt ca. 50 Minuten bei einer empfohlenen Gruppengröße von maximal acht Kindern.

Der CFT1 besteht aus fünf Subtests:

- *Substitutionen* (Reproduktiver Aspekt der Wahrnehmung)
  Es wird die Fähigkeit der Kinder erfasst, innerhalb eines vorgegebenen Zeitrahmens schnell passende Symbole zu erkennen.
- *Labyrinthe* (Produktiver Aspekt der Wahrnehmung sowie visuelle Orientierung und Aufmerksamkeit)
  Die Schüler sollen Labyrinthe bei vorgegebener Zeit durchfahren.
- *Klassifikationen*
  Es wird erfasst, wie die Kinder figurales Material klassifizieren können, indem sie die nicht zugehörige Figur unter den angebotenen Figuren finden.
- *Ähnlichkeiten*
  Die Kinder sollen die gleiche Zeichnung unter merkmalsähnlichen Zeichnungen wiederfinden. Es wird erfasst bis zu welchem Komplexitätsgrad die Kinder die figuralen Veränderungen wahrnehmen.

Es liegen sowohl Alters- als auch Klassennormen für Vorschul- und Grundschulkinder in Form von Prozenträngen, T-Werten und Intelligenzquotienten vor. Der CFT 1 wurde in der Hauptuntersuchung im Jahr 1976 mit 6 078 Kindern aus 294 Klassen mit den Parallelformen A und B getestet. Im Jahr 1995 wurden die Untersuchung mit 1 500 Schülern aus Baden-Württemberg und Sachsen wiederholt, wobei für die Normierung nur 1 200 der Testbearbeitungen genutzt wurden. Die Zuverlässigkeit für den CFT 1 ($r = .84$) entspricht den Anforderungen an psychologische Leistungstests (vgl. a. a. O., S. 24). Die Gesamtvarianz wird zu 96,1 % bzw. 94,4 % (Durchführungsart 1 und 2) aufgeklärt. 1995 konnten die Ergebnisse von 1976 bestätigt werden.

In der vorliegenden Studie werden sprachspezifische Eigenheiten der Kinder untersucht. Um keine Abhängigkeit der Werte zwischen den sprachlichen Leistungen im Intelligenztest und den Leistungen in den Tests zu schriftsprachlichen Fähigkeiten zu provozieren, wurde hier ein sprachfreier Intelligenztest gewählt. Gerade bei fremdsprachig aufgewachsenen Kindern oder bei Kindern mit einer Sprachentwicklungsstörung würden die Ergebnisse eines Intelligenztests mit sprachlichen Komponenten von diesen Begebenheiten stark beeinflusst werden.

4.2 Eingesetzte Testverfahren im 1. Schuljahr

Im ersten Schuljahr wurden zur Überprüfung der Rechtschreibleistungen der Deutsche Rechtschreibtest für erste Klassen (DRT 1) und für die Überprüfung der Basisfähigkeiten der Lesefähigkeit und des Leseverständnisses die Knuspels Leseaufgaben (Knuspel-L) verwendet. Beide können zum Ende des ersten Schuljahres eingesetzt werden und sind empirisch validiert und methodisch differenziert erfasst.

4.2.1 Deutscher Rechtschreibtest für 1. Klassen (DRT 1)

Die Reihe der Diagnostischen Rechtschreibtests (DRT 1, 2, 3, 4, 5) wurde für das erste bis fünfte Schuljahr erstellt. Der DRT 1 von Rudolf Müller (2003a) misst die Rechtschreibleistung der Erstklässler zur Früherfassung von Schriftsprachstörungen. Dazu werden den Schülern Wörter zur Notation diktiert. Es liegen zwei Paralleltests mit gleicher Schwierigkeit vor. Der Anwendungszeitraum des DRT 1 bezieht sich auf das Ende des ersten Schuljahrs. Der DRT 1 kann als Einzel- oder Gruppentest durchgeführt werden und dauert 30 bis 45 Minuten. Es existieren Normen für die quantitative Gesamtleistung, sowie für die sechs Fehlerarten, die über die qualitative Auswertung identifiziert werden können. Außerdem gibt es Normtabellen zur Beurteilung der Klassenleistungen, sowohl für die Gesamtleistung als auch für die Fehlerarten.

Die Eichung hat im Jahr 1994 stattgefunden und wurde an 1 488 Schülern aus 67 Klassen in der Bundesrepublik Deutschland durchgeführt. Der Test weist eine hohe Retest-Reliabilität ($r = .89$), eine gute Validität und zusätzlich eine positive Übereinstimmung mit dem Lehrerurteil ($r = .81$) auf. Die Gütekriterien genügen somit den Anforderungen. Es ist sowohl eine qualitative als auch eine quantitative Auswertung bezüglich phonetischer Fehler oder Speicherfehler möglich. Es liegen Individual- und Klassennormen in Prozentrangform vor. Der DRT 1 differenziert insbesondere im unteren Leistungsbereich (vgl. Hemminger, Roth, Schneck, Jans & Warnke 2000, S. 190).

4.2.2 Knuspels Leseaufgaben (Knuspel-L)

Mit Knuspels Leseaufgaben (Knuspel-L) von Marx (1998) liegt ein standardisierter Leseverständnistest vor, der auf dem Simple-View-of-Reading-Ansatz (vgl. Kap. II.1.4) basiert. Der Test dauert ca. 50 Min. und wird als Einzel- oder Gruppentest durchgeführt. Das Testverfahren wurde auf der Grundlage eines theoretischen Leselernmodells entwickelt. Mit

Knuspel-L werden in vier Untertests die folgenden Basisfertigkeiten des Lesens überprüft (vgl. a. a. O., S. 13f.):

- Subtest 1: Hörverstehen
  Verstehen mündlich gestellter Fragen und Aufforderungen
- Subtest 2: Rekodieren
  Erkennen von lautgleichen Wörtern
- Subtest 3: Dekodieren
  Erkennen von Wortbedeutungen
- Subtest 4: Leseverstehen
  Verstehen schriftlich gestellter Fragen und Aufforderungen

Die Auswertung des Tests wird in Knuspel-Score I und Knuspel-Score II eingeteilt. Der Knuspel-Score I beschreibt die Vorläuferfertigkeit für das verstehende Lesen und wird aus den Subtests 1 bis 3, Hörverstehen, Rekodieren und Dekodieren berechnet. Der Knuspel-Score II wird aus dem Rekodieren, Dekodieren und den Ergebnissen zum Leseverstehen der Schüler erstellt und beschreibt die Lesefähigkeit der Schüler (vgl. Marx 1998, S. 10). Der vierte Untertest zur Erfassung des Leseverstehens anhand schriftlicher Aufgaben wurde von den Autoren als für die erste Klassenstufe nur bedingt geeignet bewertet:

> *„Angesichts der großen Unterschiede in der didaktischen Umsetzung der Lehrpläne und Lehrmethoden und den damit verbundenen großen Differenzen in der Schriftsprachentwicklung zum Ende des ersten Schuljahres ist eine Prüfung des Leseverstehens nur in Ausnahmefällen in der Gruppentestung möglich.“ (a. a. O., S. 15).*

Aus diesen Gründen wird der Subtest 4 in dieser Untersuchung nicht verwendet. So konnte nur der Knuspel-Score I ausgewertet werden.

Der Test ist ab Ende der ersten Klasse bis zur vierten Klasse normiert und wurde als Gruppentest eingesetzt. Die Normwerte der Untertests sowie der beiden Knuspel-Scores liegen als Prozentränge und T-Werte vor. Der Test weist eine hohe Retest- und Paralleltest-Reliabilität und hohe bzw. mittlere Validität bzgl. alternativer Lesetests, des Lehrerurteils und der Deutschnote auf. Die Gütekriterien genügen somit den Anforderungen und wurden methodisch sehr differenziert erfasst. Die Eichung des Tests hat in den Jahren 1994-1997 mit 4 746 Grundschulkindern in Nordrhein-Westfalen, Sachsen und Bayern stattgefunden. Bei einem Testintervall von ein bis drei Wochen lag die Retest-Reliabilität der Gesamtstichprobe

zwischen .627 ≤ r ≤ .832, die Paralleltest-Reliabilität zwischen .791 ≤ r ≤ .850 (vgl. Marx 1998, S. 85f.)

5 Statistische Verfahren

Die Zielsetzung der hier vorliegenden Untersuchung ist die Konzeption einer curriculumbasierten Messung für den Vorschulbereich anhand der Erkenntnisse aus der hiesigen deutsch- und US-amerikanischen Literatur, sowie die Evaluation der Messungen durch eine im Raum Schleswig-Holstein längsschnittlich durchgeführte Untersuchung. Die Evaluation einer Lernverlaufsdiagnostik, wie die Messung Phonolino sie beschreibt, basiert grundsätzlich auf der klassischen Testtheorie (vgl. Klauer 2011, S. 207ff.), birgt jedoch Probleme. So beschreibt Klauer (ebd., S. 207):

> *„Ähnlich wie bei der curriculumbasierten Messung (CBM) werden bei der Lernverlaufsdiagnostik über einen längeren Zeitraum hinweg regelmäßig Tests gegeben, die stets ein und dieselbe Kompetenz messen sollen und dann erlauben, den Lernverlauf abzubilden. Solche Tests müssen immer neu sein, jedes Mal aber dasselbe messen und stets gleich schwierig sein, was unerwartet problematisch ist. Darüber hinaus sollen sie in der Lage sein, Änderungen der Kompetenz der Probanden sensibel zu diagnostizieren."*

Die Frage nach der folgerichtigen statistischen Auswertung von Lernverlaufsdiagnostiken ist in der Forschung noch in Diskussion. Wilbert und Linnemann (2011, S. 240) sowie Wilbert (2014, S. 287ff.) beschreiben, dass ebenfalls die Auswertung durch probabilistische Modelle oder durch latente Wachstumsmodelle möglich sei, dies jedoch nur durch ein aufwändigeres Forschungsdesign geleistet werden kann. In Bezug auf die Basiselemente einer Evaluation werden die Reliabilität, Änderungssensitivität, die Überprüfung der gleichen Schwierigkeit der Einzelmessungen, sowie die Validität von Wilbert und Linnemann (2011, S. 226ff.) und auch von Klauer (2011, S. 209ff.) genannt.

Wilbert und Linnemann (2011, S. 226) konstatieren:

> *„Dabei sind diese Anforderungen nicht grundsätzlich verschieden zu denen, die an Tests zur Statusdiagnostik gestellt werden, erfahren aber eine abweichende Akzentuierung, und es entstehen besondere Probleme bei deren Umsetzung."*

Aus pragmatischen Gründen wird nicht auf die Berechnung von probabilistischen oder latenten Wachstumsmodellen zurückgegriffen. Es wird sich auf die klassische Testtheorie beschränkt, um die Lernverlaufsdiagnostik Phonolino für den ersten Schritt evaluieren zu können. In weiterführenden Untersuchungen kann die Messung weitergehend statistisch (z.B. über die von Wilbert 2014, S. 287 benannte Rasch-Modellierung für Veränderungsmessungen) wie auch auf seine Praktikabilität (Stufe 3 nach Fuchs) hin untersucht werden.

Nach Deno, Fuchs, Marston und Shin (2001, S. 510ff.) sind die klassischen Gütekriterien zwar eine wichtige Grundlage eines Messinstruments, um jedoch Wachstum messen zu können, müssen noch weitere Kriterien erfüllt werden. In der vorliegenden Studie wird somit die klassische Testtheorie angewendet, die als Grundlage zur Evaluation der *Stufe 1 nach Fuchs* (2004, S. 189, vgl. auch Kap. III.6, S. 249f.) dient. Für die Überprüfung der Hauptgütekriterien Objektivität, Reliabilität und Validität sowie einer Itemanalyse der einzelnen Messungen werden zunächst die statistischen Methoden erläutert, die bei der Auswertung der Datensätze genutzt wurden.

Für eine genauere Analyse werden die von Klauer (2011, S. 209) aufgestellten Postulate für die Änderungssensibilität einer curriculumbasierten Messung verwendet. Dies ist ein Ansatz der Überprüfung der *Stufe 2 nach Fuchs* (2004, S. 189, vgl. auch Kap. III.6, S. 249f.). Das Hauptaugenmerk dieser Studie liegt jedoch in der Erstellung und Evaluation der Messungen. Die *Stufe 3 nach Fuchs* (ebd.) muss Teil einer Folgestudie sein, sobald die Voraussetzungen der Stufe 1 und 2 positiv bestätigt werden können.

Im Anschluss wird mittels der logistischen Regression geprüft, inwieweit eine *Verbesserung des Prognosemodells* zur Vorhersage von Schwierigkeiten im Lesen und Rechtschreiben im ersten Schuljahr mittels Hinzunahme der nonverbalen kognitiven Leistung der Kinder möglich ist.

Ein Überblick über die verwendeten statistischen Verfahren bietet Tab. 44.

Im Folgenden werden in einem Exkurs zunächst die Grundbegriffe zur Korrelationsanalyse und den Signifikanztests erläutert, im Anschluss wird die Analyse der klassischen Testgütekriterien nach der klassischen Testtheorie vorgestellt und die einzelnen in Tab. 44 beschriebenen statistischen Verfahren mit ihren Kennwerten erläutert.

Tab. 44: Statistische Verfahren der vorliegenden Untersuchung

Untersuchungen in Bezug auf ...			Statistische Verfahren	Annahme (A), Hypothese (H)
Stufe 1 nach L.S. Fuchs	Objektivität	Durchführungsobjektivität Auswertungsobjektivität Interpretationsobjektivität	-	A1.1 A1.2 A1.3
	Differenzierungs-fähigkeit/ Itemanalyse	Rohwertverteilung	Häufigkeitsverteilungen	A2.1
		Itemschwierigkeit	Deskriptive Analysen Mittelwertvergleiche	A2.2
		Trennschärfe	Korrelative Analysen	A2.3
		Dimensionalitätsprüfung	Faktorenanalyse	A2.4
	Reliabilität	Paralleltestreliabilität	Korrelative Analysen	A3.1
		Split-Half-Reliabilität	Korrelative Analysen	A3.2
		Interne Konsistenz	Reliabilitätsberechnung (Cronbachs Alpha)	A3.3
	Validität	Inhaltsvalidität	-	A4.1
		Übereinstimmungsvalidität (Konvergenzvalidität)	Korrelative Analysen	A4.2
		Prognostische Validität	Korrelative Analysen Klassifikatorische Analysen	A4.3 A4.4 A4.5 A4.6
Stufe 2 nach L.S. Fuchs	Änderungssensibilität	Anstieg der Mittelwerte	Mittelwertvergleiche	H5.1
		Anstieg der Leistung in Punktwerten (alle zwei Messungen)	Deskriptive Analysen	H5.2
		Leistungssteigerung über die acht Messungen	Korrelativen Analysen	H5.3
Zusätzliche Analysen	Erweiterung des Prognosemodells	Prognostische Validität bei Hinzunahme der kognitiven Leistung	Logistische Regressionsanalyse Klassifikatorische Analysen	H6.1 H6.2 H6.3 H6.4

5.1 EXKURS: Grundbegriffe verwendeter statistischer Methoden

Wichtige Grundlagen der hier vorgestellten statistischen Untersuchung sind die Prüfung der Daten bzgl. etwaiger Korrelationen und auf Signifikanz. Die Begriffe werden an dieser Stelle in Bezugnahme auf die Annahmen und Hypothesen geklärt, sowie die Kennwerte für die Interpretation als Orientierung festgehalten.

5.1.1 Korrelationen

Zur Ermittlung von signifikanten Zusammenhängen wird (häufig zur Reliabilitäts- und Validitätsanalyse) die Korrelationsanalyse verwendet. Die Korrelationsanalyse entspringt der deskriptiven Statistik und dient der Untersuchung und Beschreibung von Abhängigkeiten verschiedener Variablen. Durch dieses Verfahren werden statistisch signifikante Korrelationen identifiziert, die zu einem Zeitpunkt einen Zusammenhang zwischen zwei oder mehr Variablen beschreiben (vgl. Hartke & Koch 2008, S. 46). Mit der Korrelations-analyse wird die Ausprägung eines Merkmals X auf ein Merkmal Y gemessen. Die metrischen Variablen X und Y werden dabei als eine Punktwolke analysiert (vgl. Clauß, Finze & Partzsch 2004, S. 271f.).

Durch die Bestimmung eines Korrelationskoeffizienten („r“) ist es möglich, das Maß des Zusammenhangs zu identifizieren. Der Korrelationskoeffizient bewegt sich zwischen den Werten 1 und -1. Je näher sich der Betrag an 1 oder -1 befindet, umso höher ist der Zusammenhang einzustufen, analog dazu ist der Zusammenhang umso schwächer zu interpretieren, je näher sich der Wert an 0 bewegt (vgl. Bühl 2008, S. 346 und Tab. 45).

Tab. 45: Interpretation des Korrelationskoeffizienten (nach Cohen 1988)

Korrelations-koeffizient (r)	**Interpretation**
0,1	geringe Korrelation
0,3	mittlere Korrelation
0,5	hohe Korrelation

Zur Interpretation der numerischen Werte der hier berechneten Korrelationskoeffizienten gibt es laut Clauß und Ebner (vgl. 1982, S. 122f.) keine verbindlichen Richtlinien. In der vorliegenden Studie wird somit zunächst (vor der inhaltlichen Interpretation) mit den Bezeichnungen geringe, mittlere und hohe Korrelation nach Cohen gearbeitet (vgl. Tab. 43).

Die Berechnung des Korrelationskoeffizienten ist abhängig vom Skalenniveau der jeweiligen Variablen. In der vorgelegten Studie werden Produkt-Moment-Korrelationen nach Pearson (vgl. Kleiter 1988, S. 228) genutzt, die auf Intervallskalen bzw. normalverteilten Daten basieren. Die zu vergleichenden Merkmale (BISC, DRT 1, Knuspel-L, CBM) sind intervallskaliert. Im Folgenden ist der Korrelationskoeffizient r als Korrelationskoeffizient von Pearson und Bravais[27] zu verstehen.

Die Irrtumswahrscheinlichkeit reduziert sich bei steigendem Stichprobenumfang. Je größer die Stichprobe, desto eher können Messfehler kompensiert werden. Der Korrelations-koeffizient r muss also immer in Zusammenhang mit dem Stichprobenumfang N betrachtet werden. Nach Weber (1972, S. 652) kann der Korrelationskoeffizient mit den Irrtumswahrscheinlichkeiten auf 0,1 %, 0,27 %, 1 % und 5 %-Niveau gekennzeichnet werden. Das bedeutet, dass ein korrelativer Zusammenhang mit einer Wahrscheinlichkeit von 99,9 %, 99,7 %, 99 % oder 95 % bestimmt werden kann. Die berechneten Korrelationskoeffizienten werden in dieser Untersuchung auf ihre Verlässlichkeit hin durch die Überprüfung der Irrtumswahrscheinlichkeit auf dem Signifikanzniveau von 5 % und 1 % anhand der Tabelle der Freiheitsgrade für Maßkorrelationen (bzw. Produkt-Moment-Korrelationen) nach Weber überprüft (vgl. Weber in Clauß und Ebner, 1968 bzw. die Maßkorrelationen aus Clauß, Finze & Partzsch 2004, S. 64f.).

Notwendig für eine kausale Interpretation sind jedoch immer theoretisch begründete Annahmen, die sich aus wissenschaftlichen Modellen entwickeln lassen. Eine statistisch relevante Korrelation erweist sich somit erst durch die Berücksichtigung theoretischer Hypothesen als wissenschaftlich relevant. Der Korrelationskoeffizient lässt keine Aussage bezüglich der Richtung des Zusammenhangs zu.

Bei der Auswertung der Analysen sollte die unterschiedliche Ausrichtung des Merkmals Berücksichtigung finden. Die Merkmale CBM und die Rohwerte des DRT 1 sind unter-schiedlich gepolt: Die CBM misst die Anzahl der Treffer, während der DRT 1 bei den Rohwerten die Fehleranzahl zusammenzählt: Je weniger Rohpunkte ein Kind in der CBM erreicht, so ist zu vermuten, dass es umso mehr Fehlerpunkte im DRT 1 erreicht. Wenn ein korrelativer Zusammenhang erwartet wird, dann wird in diesem Fall ein negativer Korrelationskoeffizient ermittelt. Die Korrelationen zwischen dem ermittelten Prozentrang (als normierte Größe) und Phonolino ergibt dagegen bei einem erwarteten Zusammenhang

27 In SPSS wird die Kurzform „Pearson-Korrelation“ verwendet.

einen positiven Korrelationskoeffizienten. Je höher der Prozentrang ist, umso höher sollte dann auch die Anzahl der Treffer in der CBM ausfallen.

5.1.2 Signifikanz

Für die statistische Hypothesenprüfung wird häufig ein Signifikanztest durchgeführt, der die Entscheidung über die Beantwortung einer Hypothese fällen kann. Üblicherweise wird zu der Forschungsfrage eine Nullhypothese aufgestellt, die dann mittels des Signifikanztests geprüft wird. Hierfür wird ein Signifikanzniveau festgelegt z.B. 0,05 unter dessen Annahme der Gültigkeit ein Wertebereich für die Nullhypothese hergestellt wird. Dieser umfasst dann „ungewöhnliche" Ergebnisse (vgl. Bortz & Schuster 2010, S. 10f.). Je kleiner das Signifikanzniveau gewählt wird, desto kleiner ist auch der Wertebereich der „ungewöhnlichen Ergebnisse". Das Signifikanzniveau (oder auch Irrtumswahrscheinlichkeit) bezeichnet die Wahrscheinlichkeit, wenn die Nullhypothese fälschlicherweise verworfen wird (Fehler erster Art). Die im Folgenden durch SPSS ermittelten Werte geben einen sogenannten p-Wert an, der auch als „Signifikanzwert" bezeichnet wird (vgl. Interpretation des p-Werts in Tab. 46). Die Höhe der Signifikanz ist dem Wert des Signifikanzniveaus entgegengesetzt (vgl. Bortz & Döring 2006, S. 740).

Tab. 46: Interpretation des Signifikanzniveaus
(vgl. Bortz & Döring 2006, S. 740)

Wert (Betrag)	**Interpretation**
$> 5\%$	nicht signifikant
$\leq 5\%$	signifikant
$\leq 1\%$	sehr signifikant
$\leq 0{,}1\%$	hoch signifikant

Die Nennung von Signifikanzen als einzige Überprüfung von Hypothesen wird von Bortz und Schuster (2010, S. 112f.) aufgrund der fehlenden praktischen Bedeutung des Wertes kritisiert. Dagegen unterstützen sie die praktische Bedeutsamkeit, da bspw. bei einer hohen Mittelwertdifferenz geprüft werden muss, ob diese nicht zufällig zustande gekommen ist. Jedoch sollte der Wert einer statistischen Arbeit nicht ausschließlich über Signifikanzen beurteilt werden, sondern bedarf stets einer kritischen Überprüfung des Versuchsdesigns, der Durchführung und Auswertung.

Im nächsten Abschnitt werden die statistische Vorgehensweise und die Methodenwahl in Bezug auf die klassischen Testgütekriterien der Reliabilität, der Validität und der Itemanalyse vorgestellt.

5.2 Analyse der klassischen Testgütekriterien

Zur Beschreibung der klassischen Testgütekriterien sei auf den Exkurs in Kapitel I.2.1 verwiesen.

In dieser Untersuchung wird die Objektivität in Bezug auf die Konzeption des Untersuchungsmaterials (vgl. Kap. IV.2, S. 256) geprüft. Dabei wird zwischen Ausführungs–, Durchführungs- und Interpretationsobjektivität unterschieden und die Konzeption der Messung daraufhin analysiert. Im Weiteren werden die Auswahl der statistischen Methoden für die Item-, die Reliabilitäts- und die Validitätsanalyse vorgestellt.

5.2.1 Itemanalyse

Die Itemanalyse ist ein zentrales Element der klassischen Testkonstruktion (vgl. Bortz und Döring 2006, S. 217f.). Sie bewertet anhand von Qualitätsstandards die Zusammensetzung der Messung anhand der Items. Der Begriff selbst ist in der Literatur nicht eindeutig, sodass die Komponenten einer Itemanalyse in Untersuchungen unterschiedlich sein können. In dieser Untersuchung werden, wie auch in Bortz und Döring (a.a.O.) beschrieben, zunächst die Rohwerte, dann die Itemschwierigkeit und die Trennschärfe und zuletzt die Dimensionalität analysiert.

5.2.1.1 Analyse der Rohwertverteilung

Zur Überprüfung der Verteilung der Rohwerte werden Häufigkeitsverteilungen erstellt. Es ist zu prüfen, ob die Verteilung der Rohwerte von Phonolino zu den acht Messzeitpunkten einer Normalverteilung entspricht, da die Nutzung vieler inferenzstatistischer Verfahren normalverteilte Rohwerte voraussetzt (vgl. Bortz & Döring 2006, S. 218). Beim Konstrukt der phonologischen Informationsverarbeitung ist zudem aufgrund der unterschiedlichen Entwicklung von Individuen zu erwarten, dass bei Phonolino eine normalverteilte Kurve vorliegen wird.

5.2.1.2 Itemschwierigkeit

Die Schwierigkeit eines Items ist abhängig von der Anzahl der Probanden sowie von der Anzahl der Antwortalternativen. Schwierige Items werden nur von wenigen Probanden richtig beantwortet. Somit wird die Verteilung der Rohwerte stark von der Itemschwierigkeit beeinflusst (vgl. Bortz & Döring 2006, S. 218). Die Aufgabenschwierigkeit der Phonolino-Messungen wird einerseits über deskriptive Analysen beschrieben und andererseits durch den Vergleich von Mittelwerten geprüft. Die Itemschwierigkeiten der Items von Phonolino lassen sich durch folgende Formel berechnen (vgl. Kelava 2010, S. 8):

Formel 9: Berechnung der Itemschwierigkeiten

$$p_i = \frac{x_i}{max(x_i)} \cdot 100$$

Anmerkung: x_i = erreichte Punkte auf Item i, $max(x_i)$ = maximale Punktzahl auf Item i

Aufgrund der Speedvariante wird die maximale Punktzahl des Items (z.B. Pseudowörter nachsprechen) über den gesamten Zeitraum (Phono 2 bis Phono 8) hinweg festgelegt. Jede monatliche Messung von Phonolino besteht aus drei Items (Pseudowörter nachsprechen, Silben klopfen, Lautisolation) mit Speedvariante. Jedem Item sind verschiedene (Pseudo-) Wörter zugeordnet. Somit handelt es sich nach Bortz und Döring (2006, S. 219) um ein mehrstufiges Item. Zur Berechnung der allgemeinen Schwierigkeit eines mehrstufigen Items nutzen Bortz und Döring (a.a.O.) die Formel 10.

Formel 10: Berechnung der Itemschwierigkeit bei mehrstufigen Items

$$p_i = \frac{\sum_{m=1}^{n} x_{im}}{k_i \cdot n}$$

Anmerkung: x_i = erreichte Punkte auf Item i, k = maximale Punktzahl auf Item i, n= Anzahl der antwortenden Personen, blau markiert: Mittelwertsberechnung

Der Wertebereich der Itemschwierigkeit beschrieben durch die Formel 10 liegt zwischen 0 (schwerstes Item) und 1 (leichtestes Item). Der blau markierte Teil beschreibt den Mittelwert, somit kann die Formel verkürzt werden (s. Formel 11). Zur besseren Interpretation wird die Itemschwierigkeit mit 100 multipliziert, um Prozentwerte vergleichen zu können:

Formel 11: Verkürzte Formel zur Berechnung der Itemschwierigkeit bei mehrstufigen Items

$$p_i = \frac{m}{k_i} \cdot 100$$

Anmerkung: m = Mittelwert der Items i, k = maximale Punktzahl der Items i

Die mittlere Itemschwierigkeit sollte 50% betragen. Damit die Messungen auch im oberen und vor allem im unteren Bereich gut differenzieren, sind Items mit unterschiedlicher Schwierigkeit notwendig. Zur Interpretation der Itemschwierigkeit empfehlen Bortz und Döring (a.a.O.):

> *„Damit ein Test Untersuchungsteilnehmer mit unterschiedlichen Fähigkeiten annähernd gleich gut differenziert, ist darauf zu achten, dass die Items eine möglichst breite Schwierigkeitsstreuung aufweisen. Im Allgemeinen werden Itemschwierigkeiten im mittleren Bereich (zwischen 0,2 und 0,8) bevorzugt."*

5.2.1.3 Trennschärfe

Die Trennschärfe der Items ist ein Indikator für die Messung eines ähnlichen Konstrukts wie die anderen Items. Sie zeigt, wie ein einzelnes Item das Gesamtergebnis des Tests repräsentiert. Sie wird für jedes Item einzeln berechnet (vgl. Bortz & Döring 2006, S. 219). Bei Phonolino wird sie somit für jedes Item „Pseudowörter nachsprechen", „Silben klatschen" und „Lautisolation" zu allen acht Messzeitpunkten berechnet. So kann gezeigt werden, wie sehr ein einzelnes Item im Vergleich zu den zusammengefassten übrigen Items differenziert. Die Trennschärfe eines Items ist damit die Korrelation des Wertes des Items mit den Werten aus sämtlichen Werten des Tests (vgl. a.a.O.). Bei intervallskalierten Tests wird die Produkt-Moment-Korrelation mit folgender Formel berechnet (vgl. a.a.O.):

Formel 12: Produkt-Moment-Korrelation (nach Bortz & Döring 2006, S. 219)

$$r_{it} = \frac{cov(i,t)}{s_i \cdot s_t}$$

Anmerkung: Trennschärfe: r_{it}, einzelnes Item: i, Gesamttestwert: t

Als Maß des Zusammenhangs wird die Trennschärfe als Korrelation berechnet und liegt im Wertebereich von -1 bis 1 (vgl. auch Kap. VI.5.1.1, S. 316 zu Korrelationen). Eine hohe Trennschärfe ist bei Items zu vermuten, deren Itemschwierigkeit im mittleren Bereich einzuordnen ist. Bei einem Test mit sehr leichten und sehr schwierigen Items ist ein geringerer Wert der Trennschärfe zu erwarten (vgl. Bortz & Döring 2006, S. 220).
Nach Fisseni sollte die Trennschärfe mindestens einen Wert von 0,3 betragen (vgl. Fisseni 2004, S. 39f.).
Insgesamt sind hohe Korrelationen wünschenswert. Dabei sind im Fall von positiven Werten in Bezug auf die Trennschärfe Korrelationen von mehr als 0,5 als hoch einzuschätzen, während mittelmäßige Korrelationen zwischen 0,3 und 0,5 liegen (vgl. Weise 1975, S. 219, zit. in Bortz und Döring 2006, S. 220).

Die Berechnung der Trennschärfe wird in dieser Untersuchung über die gesamte Skala hinweg durchgeführt, da die je drei Items pro Messung alle auf den gleichen Zusammenhang der phonologischen Informationsverarbeitung hinweisen. Somit müssen alle Teilmessungen bei den Berechnungen mit einbezogen werden.

5.2.1.4 Dimensionalität

Das Ziel der Überprüfung der Dimensionalität ist es, die aus der Theorie begründeten Items durch die Faktoriellität des Messverfahrens zu beweisen. Hierfür wird die Methode der konfirmatorischen Faktorenanalyse genutzt. Nach einer faktorenanalytischen Bestimmung der Dimensionalität werden üblicherweise homogene Items zu einer Skala und heterogene Items zu mehrdimensionalen Skalen zusammengefasst. Die bestimmten Faktoren stehen dabei inhaltlich für das zu messende Konstrukt. Die Anzahl der zu extrahierenden Faktoren ist nicht immer eindeutig. Ein mehrdimensionaler Test sollte somit stets theoretisch begründet und durch eine inhaltlich festgelegte Anzahl von Faktoren überprüft werden (vgl. Bortz & Döring 2006, S. 221). Für die hier vorgelegten Messungen lassen sich inhaltlich drei mögliche Faktoren begründen. Über alle acht Messungen im letzten Kindergartenjahr hinweg sollten sich die drei Faktoren aufgrund ihrer Aufgabenstellung (vgl. Kap. IV.2.5, S. 265) „Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“ als drei übergeordnete Merkmale in drei Skalen zusammenfassen. Hierfür wird bei der Berechnung über SPSS für die Faktorenzahl der Wert 3 festgelegt. Es wird durch eine Faktorenanalyse geprüft, ob repräsentative Faktoren für die Teilbereiche der phonologischen Informationsverarbeitung „der Fähigkeit des phonetischen Rekodierens im Arbeitsgedächtnis“ (gemessen über: Pseudowörter nachsprechen), „Silbensegmentierung“ (gemessen über: Silben klopfen) und „Laute erkennen“ (gemessen über: Lautisolation) in den Messungen vorliegen und ob es als gerechtfertigt erscheint, die Items der verschiedenen Skalen zu einer möglichen Gesamt- oder Teilskala zusammenzufassen. Durch die Faktorenanalyse, die hier durch die Hauptkomponentenanalyse mit der Varimax-Rotation durchgeführt wird, werden Faktorladungen für jedes Item produziert. Durch die Hauptkomponentenanalyse wird die Gesamtvarianz eines Items berücksichtigt. Die Varimax-Rotation bietet eine einfache Form der Interpretation, da die Faktoren so rotiert werden, dass hohe Faktorladungen sich erhöhen und kleine Faktorladungen sich verringern, sodass die größtmögliche Varianz der Ladungen erreicht werden kann (vgl. Bortz 2010, S. 419). Items mit geringen Faktorladungen sollten aus der Messung entfernt werden. Stevens (2002, S. 221) postuliert, dass Items bei einer Faktorladung über 0,4 eine interpretierbare Repräsentation des jeweiligen Faktors darstellen.

Bevor eine Faktorenanalyse durchgeführt wird, muss festgestellt werden, inwiefern die Varianz einer Variablen durch die anderen erklärt wird. Dies kann durch das Kaiser-Meyer-Olkin-Kriterium geprüft werden, das anzeigt, wie stark die Variablen voneinander abhängen. Der Kennwert des KMO-Kriteriums zeigt an, ob eine Faktorenanalyse als sinnvoll erscheint. In Tab. 47 wird das Kriterium nach Bühner (2006, S. 206) bewertet. Somit sind bei einem Wert oberhalb von .50 die gewählten Items für eine Faktorenanalyse geeignet (vgl. Bühner 2006, S. 206ff.).

Tab. 47: Interpretation der Kennwerts des KMO-Kriteriums (nach Bühner 2006, S. 206)

Wert (Betrag)	**Interpretation**
unter 0.5	ungeeignet
0,5 bis 0,59	schlecht
0,6 bis 0,69	mäßig
0,7 bis 0,79	mittel
0,8 bis 0,89	gut
über 0,9	sehr gut

Eine weitere Voraussetzung zur Durchführung einer Faktorenanalyse ist, dass die Items der Korrelationsmatrix nicht miteinander korrelieren. Hierfür kann der Bartlett-Test auf Sphärizität verwendet werden. Der Test zeigt, ob die Abweichung von der Nullhypothese, dass die erhobenen Kennwerte aus einer Stichprobe mit lauter Nullkorrelationen stammen, signifikant ist (vgl. a.a.O.).

Die Ergebnisse der Faktorenanalyse werden von der Kommunalität (h2) der Items im Zusammenhang mit der Stichprobengröße beeinflusst. Die Kommunalität gibt an, in welchem Maß eine Variable durch die Faktoren aufgeklärt werden kann. Gleichzeitig erklärt sie den gemeinsamen Anteil an Varianz und kann kleiner 1 sein. Sollte die Kommunalität stark von 1 abweichen, ist zu vermuten, dass die Faktoren ungünstig gewählt worden sind (vgl. Bortz & Schuster 2010, S. 393). Den Zusammenhang von Stichprobengröße und Kommunalität beschreibt Klopp (2013, S. 4) angelehnt an Bühner (2006, S. 193) folgendermaßen:

Tab. 48. Zusammenhang zwischen Stichprobengröße, Kommunalität und Durchführbarkeit einer Faktorenanalyse (nach Klopp 2013, S.4 angelehnt an Bühner 2006, S. 193)

Stichprobengröße	Kommunalität	Beurteilung
n < 60	$h^2 < .60$	keine Faktorenanalyse durchführbar
n = 60	$h^2 < .60$	gerade ausreichend
n = 100	$h^2 < .50$	ausreichend
n = 200	$h^2 < .50$	fair
n = 300	$h^2 < .50$	gut
n = 500	$h^2 < .50$	sehr gut
n = 1000	$h^2 < .50$	exzellent

Die Überprüfung der Rohwerte, der Itemschwierigkeit, der Trennschärfe und der Dimensionalität dienen dazu, „schlechte“ Items zu identifizieren. Im Vergleich zu einem hypothesengeleiteten Vorgehen gibt es keine eindeutigen Vorgaben, ab welchem Wert ein Item aus dem Gesamtverfahren ausgeschlossen werden soll. Daher wird sich die Interpretation an den oben angegebenen Kennwerten orientieren. Es wird im Allgemeinen darauf geachtet, dass kein einzelnes Item durch deutliche Unterschiede zu den anderen Items heraussticht.

5.2.2 Reliabilitätsanalyse

Die Reliabilität wird hier als „Grad der Messgenauigkeit (Präzision) eines Instruments“ (Bortz & Döring 2006, S. 196) bezeichnet. Sie wird in dieser Untersuchung durch die Paralleltestreliabilität, die Split-Half-Reliabilität (Testhalbierungsreliabilität) und die Interne Konsistenz geprüft.

5.2.2.1 Paralleltest-Reliabilität

Zur Überprüfung der Paralleltest-Reliabilität der vorliegenden CBM wird eine Korrelations-analyse verwendet. Die curriculumbasierte Messung ist ein nicht im klassischen Sinne konstruiertes Messverfahren. Hier werden – im Gegensatz zu klassischen Verfahren - die einzelnen Messungen in Zeitabständen von beispielsweise einem Monat durchgeführt. Die Items variieren dabei bei jeder Messung, sodass die Retest-Reliabilität nicht durch Testwiederholung gemessen werden kann. Curriculumbasierte Messungen können somit nicht im herkömmlichen Sinne mit der klassischen Testtheorie überprüft werden. Im Allgemeinen werden die Messergebnisse der Einzelmessungen in der maßgeblichen Literatur

miteinander korreliert. Es handelt sich damit um eine Art der Paralleltest-Reliabilität, die keine Paralleltestreliabilität im engeren Sinne ist, da die Paralleltests nicht zu gleichen Testzeitpunkten, sondern im Abstand von einem Monat durchgeführt werden (vgl. Strathmann et al. 2010, S. 65). Die curriculumbasierten Messungen von Phonolino können als Parallelversionen verstanden werden, die dasselbe Konstrukt operationalisieren. Es ist zu erwarten, dass je ähnlicher die Ergebnisse der einzelnen Messungen in Bezug auf die Vorschulkinder ausfallen, desto weniger Fehlereffekte darauf einzuwirken scheinen (vgl. Bortz & Döring 2006, S. 197).

Es ist zu vermuten, dass nur nebeneinanderliegende Kennwerte eine hohe Korrelation aufweisen, während weiterauseinander liegende Messungen der Erweiterung des phonologischen Wissens der Kinder unterliegen und ggf. eine Weiterentwicklung der Fähigkeit zu verzeichnen ist. Es ist anzunehmen, dass sich die Vorschulkinder ggf. in ihren Leistungen in Bezug auf die phonologische Informationsverarbeitung geringfügig verbessert haben. Für eine Lernverlaufsanalyse durch die CBM Phonolino ist dieser Effekt sogar gewollt, damit sich die Lernentwicklung der Kinder durch das Verfahren abbilden lässt. Somit ist zu erwarten, dass mindestens der erste und der letzte Messwert von Phonolino (Phono 2 und Phono 9) einer Veränderung aufgrund von Entwicklung unterliegen, die auch statistisch signifikant ist. Zur Prüfung der Paralleltestreliabilität wird eine Korrelationsanalyse (vgl. Kap. VI.5.1.1) mit den nebeneinanderliegenden Rohwerten der Phonolino-Messungen durchgeführt, da der Abstand von einem Monat den geringstmöglichen Entwicklungszeitraum für diese Untersuchung darstellt. Die fachlichen Begründungen zur Konzeption der Paralleltests mit der daraus resultierenden Annahme, dass es sich bei den acht Phonolino-Messungen um Items handelt, die alle das Konstrukt der phonologischen Informationsverarbeitung abbilden, sind in Kap. IV.2.5 (vgl. Konzeption der Items, S. 265) zu finden. Die Korrelation der Werte der jeweils nebeneinanderliegenden Messzeitpunkte ergibt dann den Paralleltest-Reliabilitätskoeffizienten, der zwischen 0 und 1 liegt. 0 bedeutet, dass der Messwert nur aus Messfehlern besteht, 1 bedeutet, dass der Messwert identisch mit dem wahren Wert ist (vgl. Bortz & Döring 2006, S. 197). Weist der Koeffizient (vgl. Kap. VI.5.1.1) eine hohe Korrelation aus, liegt somit eine geringere Wahrscheinlichkeit für Messfehler vor und Messungen können als reliabel interpretiert werden.

5.2.2.2 Split-Half-Reliabilität

Als weitere Bestimmung der Reliabilität von Phonolino wurde die Split-Half-Reliabilität oder auch Testhalbierungsreliabilität als Berechnungsmethode gewählt. Dazu werden pro

Proband ein Wert für die erste Hälfte des Tests und ein Wert für die zweite Hälfte des Tests berechnet. Die Testhalbierungsreliabilität entspricht damit der Korrelation der beiden Testhälften miteinander (vgl. Bortz & Döring 2006, S. 198, sowie Kap. VI.5.1.1 zu „Korrelationen"). Die Splithalf-Methode kann auch als Sonderform der Paralleltestmethode angesehen werden, da die Testhälften auch als Paralleltests mit halber Länge ausgelegt werden können. Daher beschreiben Bortz und Döring (2006, S. 198), dass die Reliabilität des Tests unterschätzt wird. Je größer die Anzahl der Items ist, umso mehr nimmt die Reliabilität eines Tests zu. Deshalb wird die Spearman-Brown-Korrektur angewandt, die die Reliabilität zur eigentlichen Gesamtlänge des Tests schätzt (vgl. Spearman 1910, zit. nach Bortz & Döring 2006, S. 198).
Zur Art und Weise der Halbierung des Tests gibt es unterschiedliche Möglichkeiten: Zufallsauswahl aus allen Testitems, erste und letzte Testhälfte, Items mit gerade und ungerader Nummer usw. (vgl. auch Bortz & Döring 2006, S. 198). Für eine Messung mit Speedvariante wie sie in Phonolino vorliegt, wird von Amelang und Zielinski (1994, S. 110) empfohlen Testhälften mit zeitlich gleich langen Testabschnitten zu bilden:

> *„Die gesamte Bearbeitungszeit wird in 2 Teile getrennt. Nach Ablauf der ersten Halbzeit müssen die Probanden auf Anweisung des Testleiters die Stelle im Antwortblatt markieren, an der sie momentan arbeiten, so daß für die erste und zweite Halbzeit erhebungstechnisch unabhängige Rohwertteilleistun-gen bestimmbar sind."*

Diese Methode der Berechnung der Split-Half-Reliabilität innerhalb einer Messung zu jedem Zeitpunkt konnte in dieser Studie nicht umgesetzt werden, da die Durchführung der Phonolino-Messungen von Erziehern geleistet wurde und damit das Aufgabenspektrum für die durchführende Person angemessen sein musste. Die Prüfung der Split-Half-Reliabilität innerhalb einer Messung hätte vorausgesetzt, dass die Erzieher nach der Hälfte der Zeit die Messung stoppen und im Auswertungsheft vermerken, wie viele Items das Kind in der Zeit bearbeiten konnte. Danach hätte das Kind die restliche Zeit zur Beantwortung weiterer Items nutzen können. Auf diese Art der Auswertung wurde verzichtet, da die Erzieher bereits unterschiedlichste Medien (CD-Player, Protokollheft, Handpuppe, Zeitmesser) bedienen mussten und weitere Vorgaben und Regeln die ordentliche Durchführung der Messungen gefährdet hätten. Eine Aufteilung der Items in zwei Teile der Messung innerhalb eines Messzeitpunkts war ebenfalls nicht möglich, da die Anzahl der Items (drei) ungerade ist und sich nicht splitten lässt. Wenn alle Messzeitpunkte berücksichtigt werden, könnte der Testhalbierungskoeffizient ermittelt werden, indem die ersten vier und die letzten vier Messzeitpunkte in je eine Testhälfte übergehen. Da Phonolino jedoch als

Verändungsmessung konzipiert ist, würde die Entwicklung der Kinder im letzten Kindergartenjahr den Reliabilitätskoeffizienten verfälschen. Somit werden die Messzeitpunkte aller Phonolino-Messungen zur Feststellung der Split-Half-Reliabilität in veränderter Reihenfolge in zwei Hälften aufgeteilt (vgl. Kap. VII.1.3.2 zur „Splithalf-Reliabilität“, Tab. 63, S. 360), wobei dann jeweils ein Item der ersten Messhälfte („Pseudowörter nachsprechen“ oder „Silben klopfen“) mit einem Item der zweiten Hälfte („Silben klopfen“ oder „Lautisolation“) verglichen wird. Klauer (2011, S. 218) hat für eindimensionale CBMs vorausgesagt:

> *„Wenn ein Testverfahren Änderungen erfasst, so bringt das Split-Half-Verfahren höhere Reliabilitäten als Reliabilitäts-schätzungen mittels Retest oder Paralleltests.“*

Dies ist aufgrund der ggf. drei vorhandenen Dimensionen (vgl. Ergebnisse zur Faktorenanalyse in Kap. VII.1.2.4, S. 353) hier in der Form nicht zwingend zu erwarten. Aufgrund der vermuteten Dreidimensionalität der Messungen, beschrieben durch die drei Items „Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“, kann die Split-Half-Reliabilität nicht viel höher sein, als die Paralleltestreliabilität.

Eine Reliabilität von $r < .50$ wird als niedrig, von $.50 \leq r \leq .70$ als mittel und von $r > .70$ als hoch bewertet (vgl. Bühl 2008, S. 507ff.).

5.2.2.3 Interne Konsistenz

Die Verlässlichkeit und Genauigkeit eines Messinstruments kann durch die interne Konsistenz als eine Form der Reliabilität überprüft werden. Die Überprüfung dieser Kriterien über die Split-Half-Reliabilität soll durch die Berechnung der internen Konsistenz gestützt werden, da die Ergebnisse von der Art und Weise der Erstellung der zwei Testhälften abhängig sind. Dafür wird die Messung nicht nur in zwei Teile zerlegt, sondern in die Anzahl der Testitems, die miteinander korreliert werden. Dies sind in Bezug auf Phonolino die drei Items pro Messung: „Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“. Hierfür wird dann das Maß *Cronbachs Alpha* (vgl. Cronbach 1951) berechnet (vgl. Bortz & Döring 2006, S. 198f.). Cronbachs Alpha zeigt an, wie gut die Items gemeinsam die latenten Variablen beschreiben.

Der Alpha-Koeffizient lässt sich nach Formel 13 berechnen (vgl. Bortz 2005, S. 559 in Bortz und Döring 2006, S. 199). Cronbachs Alpha kann einen Wert zwischen 0 und 1 annehmen. Der Wert 0 bedeutet, dass keine Konsistenz besteht, der Wert 1 beschreibt eine vollständige

Zuverlässigkeit. Eine Messung ist umso zuverlässiger, je näher die Cronbachs Alpha-Werte gegen 1 tendieren (vgl. Bortz & Döring 2006, S. 725).

Formel 13: Berechnung des Alpha-Koeffizenten (Bortz 2005, S. 559, nach Bortz & Döring 2006, S. 199)

$$\alpha = \frac{p}{\mathrm{p}-1} \cdot \left(1 - \frac{\sum_{i=1}^{p} s^2_{item(i)}}{s^2_{Testwert}}\right)$$

Anmerkung: Anzahl der Items = p, Varianz des Items bzw. des Testwerts = s

Die Interpretation der Reliabilität der Skala wird in der Literatur unterschiedlich eingeschätzt. Bortz und Döring schätzen α-Werte > 0.8 als gut ein, Cortina (1993, S. 102) beschreibt Werte > 0.6 als befriedigend. Brosius (2013, S. 826) beschreibt:

> *„In der Literatur wird häufig ein Wert von mindestens 0.8 gefordert, damit eine zusammengesetzte Skala als hinreichend zuverlässig angesehen werden kann, in einigen Fällen begnügt man sich jedoch auch mit einem Wert von 0.7."*

Cronbachs Alpha unterschätzt bei mehrdimensionalen Tests oftmals die Reliabilität und ist dagegen umso höher, je mehr Items ein Test aufweist und je mehr diese Items untereinander interkorrelieren (vgl. Cortina 1993, S. 102). Daher werden in der Literatur bei Messverfahren mit nur zwei oder drei Items auch Werte von mindestens 0.4 akzeptiert (vgl. Cratzius 2003, S. 165).
Die Cronbachs Alpha werden für diese Studie zu jedem einzelnen Messzeitpunkt für die je drei Items berechnet, um nachzuweisen, dass es sich bei allen acht Messzeitpunkten um reliable Messungen handelt. Auf eine Überprüfung der einzelnen Wörter der drei Items wird verzichtet, da dies aufgrund der Speed-Variante in den Messungen nicht aussagekräftig ist.

5.2.3 Validitätsanalyse

Die Validität als wichtigstes Testgütekriterium soll angeben, ob ein Test das misst, was er messen soll (vgl. Bortz & Döring 2006, S. 200). Zur Bewertung der Validität eines Verfahrens sollen verschiedene Aspekte der Validität geprüft werden (für eine allgemeine Beschreibung vgl. auch Kap. I.2.1.3, S. 57 „Validität"). Im Folgenden werden die Aspekte der Inhaltsvalidität, Konstruktvalidität und die damit einhergehende Übereinstimmungsvalidität sowie prognostische Validität sowie die klassifikatorischen Analysen vorgestellt.

5.2.3.1 Inhaltsvalidität

Die Inhaltsvalidität oder auch Augenscheinvalidität eines Testes liegt vor, wenn das Konstrukt in seinen Teilaspekten erschöpft ist (vgl. Kap. I.2.1.3, S. 57). Diese Inhalte werden nach dem „Augenschein" beurteilt, sodass es keiner besonderen Validierung bedarf. Dies ist nicht in Bezug auf eine numerische Darstellung möglich, sondern beruht auf subjektiven Einschätzungen, die in diesem Fall auf der wissenschaftlichen Grundlage anderer Studien getätigt wurden. Die Inhaltsvalidität ist mehr eine „Zielvorgabe, die bei der Testkonstruktion bedacht werden sollte." (vgl. Bortz & Döring 2006, S. 200). Dies erfordert eine genaue Definition der Inhalte der hier vorliegenden CBM. Somit wird diesbezüglich auf die theoretische Herleitung der CBM „Phonolino" im Kapitel IV (vgl. S. 255) hingewiesen.
Die Kontrolle der Inhaltsvalidität ist bei einem curriculumbasierten Messverfahren wie der hier vorgelegten CBM Phonolino nicht ausreichend, um die Validität des Verfahrens zu bestimmen, da die interessierenden Merkmale des Konstrukts der Lese- und Schreibfertigkeiten bzw. der phonologischen Informationsverarbeitung nicht zwingend direkt repräsentiert werden (können) (vgl. a.a.O., S. 200). Somit kann auf die Validierung durch weitere Analysen keinesfalls verzichtet werden.

5.2.3.2 Konstruktvalidität

Die Konstruktvalidität ist gegeben, wenn die Messung alle Facetten des theoretischen Konstrukts erfasst, die durch den Test erfasst werden sollen.
Die bereits oben (vgl. Kapitel VI.5.2.1.4 „Dimensionalität", S. 322) erwähnte Faktorenanalyse stellt fest, wie viele Dimensionen die vorliegende CBM Phonolino hat und inwieweit diese mit den Dimensionen des zu untersuchenden Konstrukts der phonologischen Informationsverarbeitung übereinstimmen (*faktorielle Validität*). Dies wird hier nicht weiter beleuchtet, da diese Berechnungen bereits Teil der Itemanalyse sind.

Die *Kriteriumsvalidität* oder auch kriteriumsbezogene Validität ist ein spezieller Aspekt der Konstruktvalidität. Die Kriteriumsvalidität beschreibt, ob die Ergebnisse des Tests auf ein Außenkriterium schließen lassen. In dieser Arbeit wird speziell die Übereinstimmungsvalidität und die prognostische Validität der curriculumbasierten Messung Phonolino beleuchtet.

Die *Übereinstimmungsvalidität* ermittelt den Zusammenhang zwischen einem parallel existenten Kriterium, das ebenfalls die phonologische Informationsverarbeitung misst. Dies wird in dieser Arbeit anhand von Korrelationen zwischen der CBM und dem Verfahren

BISC (Bielefelder Screening) festgestellt. Da in der CBM nicht alle Komponenten des BISC verwendet wurden, werden auch speziell die Untertests, die mit der CBM inhaltlich ähnlich sind, verglichen. Hierfür werden die Messungen verwendet, die ungefähr zu den zwei Testzeitpunkten des Bielefelder Screenings stattgefunden haben, um Effekte der Verbesserung einzelner Leistungen als Fehler zu vermeiden. Die Zusammenhänge werden über eine Korrelationsanalyse zwischen den BISC-Werten sowie den Werten der Phonolino-Messungen, die zum gleichen Zeitpunkt erhoben wurden, berechnet.

Die *prognostische Validität* zeigt dagegen die Vorhersageleistung der curriculumbasierten Messung auf. Marx und Lenhard (2010, S. 9) beschreiben die prognostische Validität als ein sehr wichtiges Kriterium für Screening-Verfahren. Um die CBM als eine Art Screening nutzbar zu machen, steht außer Frage, dass diese Form der Validität ebenfalls überprüft werden soll. Zur Überprüfung werden korrelative Zusammenhänge zu den Außenkriterien im Lesen (mittels Knuspels Lesetest) und Rechtschreiben (mittels DRT1) im ersten Schuljahr zur Interpretation herangezogen. Eine hohe Korrelation zwischen den einzelnen Messungen der CBM und den Ergebnissen der Kriterien zum Lesen und Schreiben weist auf eine gute prognostische Validität des Verfahrens hin. Es ist zu vermuten, dass je weiter die Messung der CBM von der Messung des Kriteriums zeitlich entfernt ist, desto geringer ist die Korrelation, da die Kinder im letzten Kindergartenjahr im Bereich der phonologischen Informationsverarbeitung lernen und sich weiterentwickeln. Zusätzlich werden klassifikatorische Analysen über das Vierfelderschema herangezogen, um differenziertere Ergebnisse zur prognostischen Validität zu erhalten, die in den folgenden Abschnitten näher beschrieben werden. Hierbei werden zunächst das Vierfelderschema und dann die Kennwerte der klassifikatorischen Analyse dargestellt.

5.2.3.3 Das Vierfelderschema

Zur Überprüfung der Klassifikationsvalidität reicht eine einfache Produkt-Moment-Korrelation nicht aus (vgl. Jansen, Mannhaupt, Marx & Skowronek 2002, S. 45). Zur näheren Analyse wird das Vierfelderschema genutzt, das bereits bei den Güteindizes für Screening-Verfahren im Theorieteil beschrieben wurde (vgl. Kap. I.2.1.3.2 - Prognostische Validität).

Der Zielgedanke der Überprüfung der Vorläuferfertigkeiten des Schriftspracherwerbs ist eine Verbesserung und eine frühzeitige Förderung von Schülern, die gefährdet sind, Schwierigkeiten beim Lesen und Rechtschreiben zu entwickeln. Dies bedarf einer Klassifikation der Kinder aufgrund ihrer Leistung. Hohe Signifikanzen zwischen Prädiktor

(phonologische Bewusstheit) und Kriterium (Lese- und Schreibkompetenz) sind für eine Vorhersage von Schwierigkeiten in der Entwicklung von Lese- und Rechtschreibleistungen im Schulalter nicht ausreichend und sind zusätzlich auf der Grundlage ihrer klassifikatorischen Gütekriterien zu beurteilen. Es soll herausgefunden werden, ob die 25 % schwächsten Kinder, die durch die hier vorgestellte CBM Phonolino erkannt werden, auch durch die angegebenen Kriterien Knuspel-L und DRT 1 identifiziert werden können.

Die binären Daten (Risiko- und Nichtrisikokinder), die durch eine Dichotomisierung entstehen, lassen sich in dieser Vierfeldertabelle abbilden. Als Basis für die Dichotomisierung werden die Prozentränge (PR), T-Werte (T) oder Risikopunkte (RiPu) des jeweiligen Kriteriums (BISC, DRT 1, Knuspel-L) verwendet. Probanden werden anhand eines Cut-Scores, z.B. einem Prozentrang kleiner als 15, als Risikokinder klassifiziert, Kinder mit einem größeren Prozentrang gehören der Gruppe der Nicht-Risikokinder an. Die Daten der CBM dienen wiederum als Prädiktor und werden gleichermaßen unterteilt.

> *„Die Ermittlung der klassifikatorischen Vorhersagegüte insgesamt ergibt sich aus der Übereinstimmung der dichotomisierten Prädiktor- mit der dichotomisierten Kriteriumsverteilung“ (Lichtenstein & Ireton 1984; Marx 1989, zit. in Jansen, Mannhaupt, Marx & Skowronek 2002, S. 46).*

Nach dem Eintrag in einer Vierfeldertabelle werden die Beziehungen zwischen Prädiktor und Kriterium in einer Matrix angezeigt, die vier unterschiedliche Gruppen beschreibt (vgl. Tab. 49).

Zur Vorhersage der Lese- und Rechtschreibleistung im ersten Schuljahr können damit folgende Gruppen unterschieden werden (vgl. Barth & Gomm 2008b, S. 35f.; Tab. 49):

(a) Valid positiv: Die Kinder werden als Risikokinder klassifiziert und sie haben in der ersten Klasse tatsächlich massive Schwierigkeiten beim Schriftspracherwerb.
(b) Falsch positiv: Die Kinder werden als Risikokinder klassifiziert, es stellt sich jedoch heraus, dass sie in der ersten Klasse keine Schwierigkeiten beim Schriftspracherwerb haben.
(c) Falsch negativ: Diese Kinder werden als „Nicht-Risikokinder“ klassifiziert, haben jedoch in der ersten Klasse massive Schwierigkeiten beim Schriftspracherwerb.
(d) Valid negativ: Die Kinder werden als „Nicht-Risikokinder“ klassifiziert und haben auch keine Schwierigkeiten beim Schriftspracherwerb.

Tab. 49: Einteilung der Schüler in Gruppen im Vierfelderschema
(abgeändert nach Marx 1992a, S. 255)

	Kriterium	
Prädiktor	(a) valid positiv	(b) falsch positiv
	(c) falsch negativ	(d) valide negativ

Anmerkung: blau hinterlegt: „Treffer2; grau hinterlegt: „Fehler“

Die von der CBM klassifizierten Risikokinder ergeben sich aus der Summe von (a) und (b), die Nicht-Risikokinder aus der Summe von (c) und (d). Die Felder a und d können auch als Treffer, die Felder b und c wiederum als Fehler bezeichnet werden. Insgesamt bedeutet dies, dass je mehr Treffer im Vergleich zu Fehlern klassifiziert wurden, desto besser ist die Güte der Vorhersage zu bewerten (vgl. Jansen, Mannhaupt, Marx, Skowronek 2002, S. 46ff.).

Für die Dichotomisierung der Daten muss jeweils ein Cut-Score gewählt werden, der die Kinder in Risiko- und Nicht-Risikokinder unterteilt. Ein veränderter Cut-Score verändert dabei auch die Genauigkeit der Messung eines Screening (vgl. National Research Center 2006, S. 1.3). Der Cut-Score beschreibt die Größe anhand derer entschieden wird, welches Kind ein Risiko- bzw. ein Nicht-Risikokind ist. Es ist sozusagen die Linie, bei dessen Schwellenwert die Trennung von Risiko und Nicht-Risiko stattfindet und damit der Grenzwert, anhand dessen Kinder in einem Frühdiagnostikverfahren identifiziert werden, die eine gewisse Risikogefährdung haben im Schulalter eine Lese-Rechtschreibschwäche zu entwickeln.

Bei der Zuordnung der Kinder anhand der dargestellten Gruppen in ein Vierfelderschema wird in der vorliegenden Studie die CBM Phonolino als Prädiktor und die beiden standardisierten Testverfahren DRT 1 und Knuspel-L als Kriterien herangezogen (vgl. Tab. 50). Durch die Dichotomisierung der Daten können die Kennwerte der klassifikatorischen Validität (vgl. Kap. i. F. VI.5.2.3.4 - Kennwerte der klassifikatorischen Validität) des Prädiktors bestimmt werden. In der vorliegenden Studie werden verschiedene Cut-Off-Werte genutzt (Cut-Score PR = 10 bzw. T-Wert = 40 und T-Wert = 37). Dabei wurde für den DRT 1 der in der Literatur übliche Wert der Unterteilung anhand des Prozentrangs PR ≤ 25 bzw. PR > 25 unterschritten, da die Kinder dieser Studie eine relativ schwache Stichprobe waren.

Tab. 50: Einteilung der Schüler in Gruppen im Vierfelderschema mit der CBM Phonolino als Prädiktor und den standardisierten Testinstrumenten Knuspel-L (Leseverständnis) und DRT 1 (Rechtschreibung)

		Kriterium	
		DRT 1 (PR ≤ 10) Knuspel-L (T ≤ 43)	DRT 1 (PR > 10) Knuspel-L (T > 43)
Prädiktor	$Phono_k$ (< 25% Punkte des Gesamtrohwerts)	(a) valid positiv	(b) falsch positiv
	$Phono_k$ > 25% Punkte des Gesamtrohwerts)	(c) falsch negativ	(d) valide negativ

Die Prozentrangskala zeigt die relative Position des Schülers in einer Rangordnung an. Es wird beschrieben, wie viele Schüler genauso gut oder schlechter abgeschnitten haben als der getestete Schüler. Prozentränge werden zur Interpretation meistens in vier Quartile geteilt, wodurch hier der Wert PR ≤ 25 entsteht (vgl. Ingenkamp & Lissmann 2008, S. 64f.). Analog wird für den Knuspel-L der T-Wert T ≤ 37 und T > 37 bzw. T ≤ 40 bzw. T > 40 für die Dichotomisierung gewählt.

Der Cut-Score sollte näher betrachtet werden. In Abbildung 55 wird ersichtlich, wie sich die Verteilung von Risiko- und Nichtrisikokindern ergeben würde, wenn der Prädiktor exakt die Kinder prognostiziert, die später gemessen durch das Kriterium auch Schwierigkeiten aufweisen.

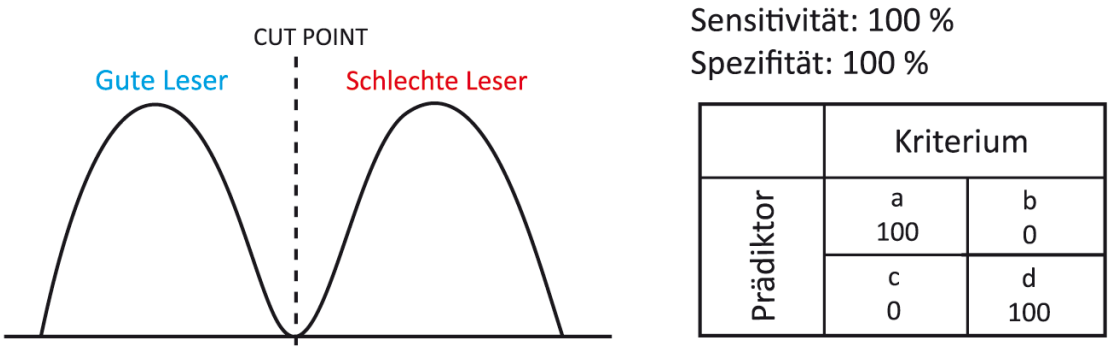
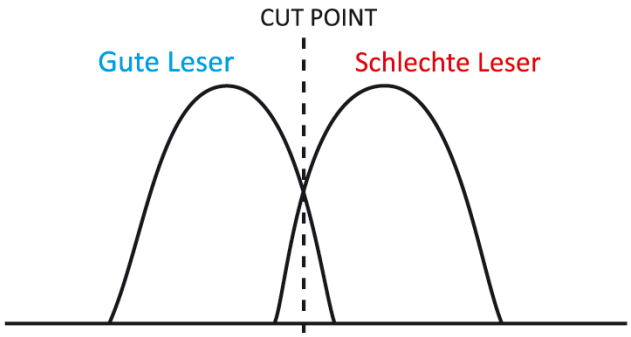


Abb. 55: Modell des „optimalen Screenings“ (angelehnt an Catts 2006, S. 11)
Anmerkung: a = Anteil richtig-positiver Screening-Befunde, b = Anteil der richtig-negativen Befunde, c = Anteil der falsch-negativen Befunde, d = Anteil der falsch-positiver Befunde, in Prozent

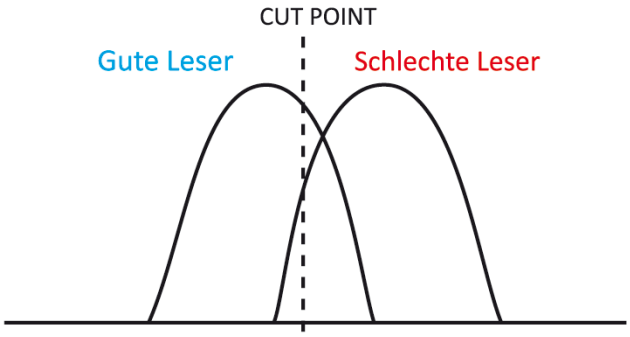
Der Cut-Score bildet in dem Fall die Symmetrieachse zwischen den Risiko- und Nicht-Risikokindern. Bei diesem optimalen Modell würden sowohl Sensitivität als auch Spezifität bei 100 % liegen. Diese Werte sind in der pädagogisch-psychologischen Forschung nur als

Gedankenkonstrukt möglich und würden in der Realität mit hoher Wahrscheinlichkeit nicht auftreten. Abb. 56 zeigt ein Beispiel, wie sich die Kennwerte der Messung verändern, wenn der Cut-Score verschoben wird. Hier wird anhand der Kurven sichtbar, dass der Cut-Score die Anzahl der richtig identifizierten Risiko- und Nicht-Risikokinder stark beeinflusst.



Sensitivität: 80 %
Spezifität: 80 %

		Kriterium	
Prädiktor	a	80	b 20
	c	20	d 80



Sensitivität: 95 %
Spezifität: 65 %

		Kriterium	
Prädiktor	a	95	b 35
	c	5	d 65

Abb. 56: Modell zweier typischer Screenings mit unterschiedlichem Cut-Off-Wert (angelehnt an Catts 2006, S. 12)
Anmerkung: a = Anteil richtig-positiver Screening-Befunde, b = Anteil der richtig-negativen Befunde, c = Anteil der falsch-negativen Befunde, d = Anteil der falsch-positiver Befunde, in Prozent

Der prozentuale Anteil der Kinder lässt sich durch den Cut-Off-Wert verschieben. Durch Veränderung des Cut-Offs werden automatisch auch die klassifikatorischen Gütekriterien (vgl. Kap. VI.5.2.3.4 - Kennwerte der klassifikatorischen Validität) wie Sensitivität, Prädiktortrefferquote oder Spezifität verändert.

Je strenger der Cut-Off-Wert definiert ist, umso wahrscheinlicher ist es, dass die identifizierten Kinder auch wirklich eine Lese-Rechtschreibschwäche entwickeln, jedoch ist es damit auch wahrscheinlicher, dass einige der Kinder, die später eine Lese-Rechtschreibschwäche entwickeln, nicht entdeckt worden sind. Diese Kinder hätten nicht präventiv gefördert werden können. Je weiter der Cut-Off-Wert definiert ist, umso mehr steigt die Selektionsrate der Risikokinder. Dagegen besteht die Gefahr, dass Risikokinder diagnostiziert wurden, die in der späteren Entwicklung keinerlei Schwierigkeiten im Lesen

und Schreiben haben. Dies könnte sich ebenfalls nachteilig auf das Kind auswirken im Sinne einer Stigmatisierung und der „self-fulfilling-prophecy" (vgl. Barth und Gomm 2008a, S. 23).

Die Früherkennungsdiagnostik versucht im Sinne des weiteren Cut-Off-Wertes Risikokinder nicht zu übersehen, da die Forscher davon ausgehen, dass eine nicht erforderliche Förderung besser ist als keine Förderung, wenn es eigentlich notwendig ist (vgl. Barth und Gomm 2008a, S. 23). Barth und Gomm (ebd.) haben im PB-LRS Kinder mit einer Standardabweichung (unter PR 15) als Risikokind identifiziert. Die Beispiele zeigen, dass durch den Cut-Score die Ergebnisse der Vierfeldertafel verändert werden können und damit auch die Aussagekraft des Screenings. Deshalb ist es wichtig, die Güteindizes der Vierfeldertafel zu berechnen, um anhand derer eine Interpretation zur Klassifikation des Verfahrens treffen zu können. Die Berechnung dieser Kennwerte der klassifikatorischen Validität wurden bereits in Kapitel I.2.1.3.2 beschrieben und im folgenden Kapitel ergänzt.

5.2.3.4 Kennwerte der klassifikatorischen Validität

Die Güteindizes für Screening-Verfahren (vgl. auch Kap. I.2.1.3.2) können durch die Vierfeldertafel berechnet werden. Diese Kennwerte sind für den Vergleich von zwei klassifikatorischen Vorhersagen erforderlich. Die Kennwerte werden im Bielefelder Screening (vgl. Jansen et al. 2000, S. 46ff.) sowie bei Tröster (Tröster 2009, S.105ff.) wie zusammenfassend in Tabelle 2 (in Kap. I.2.1.3.2 Prognostische Validität, S. 61) beschrieben. Für die Interpretation der Kennwerte der klassifikatorischen Validität stellen Marx und Lenhard (2010, S. 9) besonders die Kennwerte Sensitivität (Anteil der richtig identifizierten Kinder mit späteren Schwierigkeiten), Spezifität (Anteil der richtig klassifizierten Kinder, die später keine Schwierigkeiten entwickeln) sowie die Prädiktortrefferquote (auch Positive Korrektheit) heraus, also die Wahrscheinlichkeit, dass ein positiver Screening-Befund auftritt. Die Gesamttrefferquote sollte dagegen nicht überinterpretiert werden. Anhand eines plastischen Beispiels zeigen sie, dass es möglich wäre, bei einer sehr niedrigen Selektionsrate dennoch eine Gesamttrefferquote von 82 % erhalten zu können. Die Spezifizität weist von Natur aus ebenfalls oft hohe Werte auf, da der Anteil der Kinder, die als Nicht-Risikokinder identifiziert werden, größer ist als der Anteil der Risikokinder und dementsprechend viele Kinder auch im Schulalter dann keine Schwierigkeiten entwickeln. Durch Veränderung des Cut-Off-Kriteriums ist es möglich, die Werte der Spezifität

nochmals nach oben anzupassen, auch ohne dass der Test von diagnostischer Relevanz sei[28]. Beispielsweise wäre dies möglich, wenn alle Kinder durch den Cut-Score als Nicht-Risikokinder eingestuft werden. Die Sensitivität hat eine hohe Bedeutung, da sie aufzeigt, welche Kinder mit späteren Schwierigkeiten durch das Screening richtig klassifiziert worden sind. Zur Interpretation der Kennwerte dient für diese Untersuchung als grobe Orientierung die Einteilung in Tab. 51.

Tab. 51: Interpretation der Werte der Sensitivität, Angaben in Prozent

Wert	Interpretation
0-30	schwach
31-60	moderat
61-100	stark

Um jedoch die Wahrscheinlichkeit der Prognose zu betrachten, heben Marx und Lenhard (2010, S. 11f.) die Prädiktortrefferquote (oder auch positiver prädiktiver Wert) hervor. Dieser Wert gibt die Wahrscheinlichkeit an, dass ein identifiziertes Risikokind später tatsächlich Schwierigkeiten bekommt. Der negativ prädikative Wert dagegen kann ebenso wie die Spezifität per Zufall schon hohe Werte annehmen. Zur Bewertung des Likelihood Ratio empfiehlt Tröster (2009, S. 96) die Einteilung von Bender (2001, vgl. Tab. 52).

Tab. 52: Interpretation der Testeffizenz des Likelihood Ratio nach Bender 2001 (aus Tröster 2009, S. 96)

LR+	**LR-**	**Testeffizienz**
> 10	< 0.10	sehr gut
5 – 10	0.1 – 0.2	gut
2 – 5	0.2 – 0.5	mäßig
< 2	> 0.5	schlecht

28 Beispiel (nach Marx & Lenhard 2010, S. 9f.): 10 % der leistungsschwächsten Kinder werden bereits im Kindergarten erkannt. Bei einer vollkommen zufälligen Klassifikation der Kinder könnte man bei einer Selektionsrate von 10 % also theoretisch erwarten, dass 10 % der betroffenen und 90 % der nicht-betroffenen Kinder richtig vorhergesagt worden sind. Das wäre eine Zufallstrefferquote von 82 %, da 81 % per Zufall als nicht-betroffenen Kinder identifiziert und 1 % als richtigerweise betroffene Kinder identifiziert wurden.

Der RATZ-Index dagegen beschreibt den relativen Anstieg der Trefferquote gegenüber der Zufallstrefferquote. Der RATZ-Index wird von mehreren Autoren als der entscheidende Wert für die Überprüfung der praktischen Relevanz der klassifikatorischen Validität gesehen (vgl. Marx & Lenhard 2010, S. 11f.; Walter 2010a, S. 33f.). Der RATZ-Index beschreibt, wie zuverlässig das Screening gegenüber dem Zufall ist.

Die RATZ-Werte nach Jansen (1992, in Jansen, Mannhaupt, Marx & Skowronek 2002, S. 47) können wie in Tab. 53 dargestellt interpretiert werden.

Tab. 53: Interpretation der RATZ-Werte nach Jansen (1992, in Jansen, Mannhaupt, Marx & Skowronek 2002, S. 47), Angaben in Prozent

Wert	**Interpretation**
> 66	sehr gute Klassifikation
34 - 66	gute, aber noch unspezifische Klassifikation
< 34	ungenügende Klassifikation

In der vorliegenden Arbeit sollen Angaben zur Sensitivität, Spezifität und zum RATZ-Index genauer ausgewertet werden. Hierfür wird für die curriculumbasierte Messung Phonolino ein Cut-Off-Wert von 25% festgelegt. Vorschulkinder, die zu dem jeweiligen Messzeitpunkt zu den unteren 25% zählen, werden damit der Risikogruppe zugeordnet, die gefährdet sind, Schwierigkeiten im Lesen und Rechtschreiben in der Schule zu entwickeln. Es wird dann anhand des Kriteriums Lese- (Knuspel-L) bzw. Rechtschreibleistung (DRT1) in der ersten Klasse geprüft, wie viele Kinder tatsächlich Minderleistungen aufweisen. Hierfür wird beim Knuspel-L ein Cut-Off-Wert bei $T \leq 37$ und beim DRT1 von $PR \leq 10$ gewählt.

5.3 Methoden zur Bestimmung der Änderungssensibilität

Zur Überprüfung der Hypothese H5.1 (*Die Mittelwerte des Gesamtwerts von Phonolino bzw. der drei Items steigen signifikant über die acht Messzeitpunkte hinweg mit Ausnahme direkt nebeneinanderliegender Messungen.*) muss zunächst überprüft werden, ob die Mittelwerte der acht Gesamtmessungen sowie auch der einzelnen Items von Phonolino über die Zeit hinweg ansteigen (deskriptive Analyse). Im Anschluss wird anhand einer einfaktoriellen

Varianzanalyse mit Messwiederholung geprüft, ob sich diese Mittelwerte auch signifikant unterscheiden.

Zunächst wird der Unterschied der Itemschwierigkeiten der drei Items der jeweils acht Messzeitpunkte miteinander verglichen, um auszuschließen, dass der Lernzuwachs aufgrund sinkender Itemschwierigkeiten zustande kommt und der vermerkte Lernzuwachs der einzelnen Probanden nur aufgrund leichterer Aufgaben zu verzeichnen ist. Dies wird sozusagen als Grundvoraussetzung für die Überprüfung der Hypothese H5.1 angenommen. Dabei werden nur nebeneinanderliegende Messungen betrachtet, da davon auszugehen ist, dass bei den Messungen der Leistungszuwachs der Probanden am geringsten ist (vgl. Klauer 2014a, S. 10). Der Vergleich der Itemschwierigkeit wird auf die Differenzen zurückgeführt und an dieser Stelle nur deskriptiv beschrieben. In einer nachfolgenden Untersuchung wäre es möglich, wie bspw. Strathmann bei der Entwicklung der LVD-M beschreibt, die nebeneinanderliegenden Itemschwierigkeiten zufallskritisch zu prüfen und die jeweiligen Effektstärken miteinander zu vergleichen. Dabei greift er auf eine Bonferroni-Holm-Prozedur zurück (vgl. Strathmann 2014, S. 214). Beweist sich die Voraussetzung von möglichst gleichbleibenden Itemschwierigkeiten von nebeneinanderliegenden Messungen als richtig, können im nachfolgenden Schritt die Mittelwerte miteinander verglichen werden. Eine Differenz der Itemschwierigkeit von unter 10 % wird an dieser Stelle als Lernzuwachs interpretiert.

Zur Überprüfung, ob sich die Mittelwerte der einzelnen Phonolino-Messungen signifikant unterscheiden, wird eine einfaktorielle Varianzanalyse mit Messwiederholung (unter SPSS: „Allgemein Lineares Modell“) berechnet. Es handelt sich dabei um eine Verallgemeinerung des t-Tests für abhängige Stichproben für mehr als zwei Gruppen (vgl. Univ. Zürich, 2016a). Bei Phonolino handelt es sich um abhängige Stichproben, da die jeweiligen Messwerte zu den Messzeitpunkten von den gleichen Personen stammen. In diesem Fall soll ein Verlauf über die Zeit untersucht werden, der überprüft, ob sich die Leistung der phonologischen Informationsverarbeitung (gemessen durch Phonolino) über die acht Messzeitpunkte im letzten Kindergartenjahr hinweg signifikant verändert. Da Phonolino als curriculumbasierte Messung und damit als Veränderungsmessung konzipiert wurde, ist zu vermuten, dass sich die Mittelwerte der Messungen Phono 2 bis Phono 9 signifikant voneinander unterscheiden, wobei der Abstand von einem Monat (also die nebeneinanderliegenden Messungen wie bspw. Phono 2 und Phono 3) aufgrund des geringen Abstands keinen signifikanten Unterschied ergeben sollten, sie können vielmehr als quasi „Parallelmessungen“ interpretiert werden.

Durch die Varianzanalyse können die Veränderungen innerhalb der Personen geprüft und die Unterschiede zwischen den Mittelwerten erkannt werden. Es wurde die einfaktorielle Varianzanalyse gewählt, da der Phonolino-Gesamtwert betrachtet wird, dessen Messung über die Zeit gesehen wiederholt wird.
Als Voraussetzung zur Durchführung der einfaktoriellen Varianzanalyse sollte die abhängige Variable (also die Phonolino-Werte zu den jeweiligen Messzeitpunkten) intervallskaliert und möglichst normalverteilt sein, wobei bei einer Stichprobengröße von mehr als 25 Probanden eine Verletzung dieser Voraussetzung unproblematisch ist (vgl. Univ. Zürich 2016a). Außerdem sollte die Sphärizität, berechnet über den Mauchly-Test, gegeben sein, andernfalls müssen Korrekturen berechnet werden. Sphärizität (oder auch Zirkularität) ist gegeben, wenn die Varianzen der Differenzen aller Messpaare der Messungen gleich sind. Der Mauchly-Test zeigt gegebene Sphärizität über den p-Wert an. Ist dieser größer oder gleich .05, kann man davon ausgehen, dass Sphärizität gegeben ist. Ist der Wert signifikant, also kleiner als .05, dann müssen die Freiheitgrade noch unten korrigiert werden. Hierfür kann entweder die Greenhouse-Geisser oder die Huynth-Feldt-Korrektur verwendet werden. Girden (1992, zit. nach Holling & Gediga 2016, S. 293) empfiehlt bei einem Epsilon (Greenhouse-Geisser) größer als .75, dass die Huynh-Feldt-Korrektur genutzt werden sollen, bei einem Epsilon kleiner oder gleich .75 wird die etwas konservativere Greenhouse-Geisser-Korrektur verwendet.

Die Varianzanalyse versucht die Abweichungen der individuellen Werte vom Gesamtmittelwert zu erklären. Dafür werden alle individuellen Abweichungen vom Gesamtmittelwert quadriert. Daraus wird eine Gesamtquadratsumme (SS_{ges}, vgl. Formel 14) gebildet (vgl. a.a.O.).

Formel 14: Gesamtquadratsumme einer einfaktoriellen Varianzanalyse mit Messwiederholung

$$SS_{ges} = \sum_{k=1}^{K} \sum_{g=1}^{G} (y_{kg} - \bar{y})^2$$

Anmerkung: k= Laufindex der Personen, g= Laufindex der Messwerte einer Person, G= Anzahl der Messwerte, $\bar{y}$ = Gesamtmittelwert über alle Personen und Messzeitpunkte hinweg

Diese Quadratsumme wird durch die Varianzanalyse geteilt: in die Quadratsumme zwischen den Personen ($SS_{zwischen}$) und die Quadratsumme, die die Veränderung innnerhalb der Personen ($SS_{innerhalb}$) beschreibt. Die Quadratsummen basieren auf dem Unterschied zwischen dem Mittelwert einer Person und dem Gesamtmittelwert ($SS_{zwischen}$) bzw. den individuellen Abweichungen der unterschiedlichen Messzeitpunkte und dem persönlichen

Mittelwert ($SS_{innerhalb}$). Die Quadratsumme innerhalb der Personen wird nochmals aufgeteilt in einen Anteil, der der Messwiederholung ($SS_{Messwiederholung}$) zugeschrieben wird, sowie einem nicht erklärbaren Residualwert ($SS_{Residual}$, vgl. a.a.O., vgl. Formel 15).

Formel 15: Berechnung innerhalb der einfaktoriellen Varianzanalyse

$$SS_{ges} = SS_{zwischen} + SS_{innerhalb}$$
$$SS_{innerhalb} = SS_{Messwiederholung} + SS_{Residual}$$
$$SS_{ges} = SS_{zwischen} + SS_{Messwiederholung} + SS_{Residual}$$

Anmerkung: SS_{ges} = Gesamtquadratsumme einer einfaktoriellen Varianzanalyse mit Messwiederholung, $SS_{zwischen}$ = Quadratsumme, die die Veränderung zwischen Personen beschreibt, $SS_{innerhalb}$ = Quadratsumme, die die Veränderung innerhalb von Personen beschreibt, $SS_{Messwiederholung}$ = Anteil der Messwiederholung, $SS_{Residual}$ = unerklärbarer Residualwert

Für die Berechnungen werden dann die mittleren Quadratsummen benötigt, um F zu berechnen. Der F-Wert fällt umso höher aus, je mehr Variation sich durch die Messwiederholung erklären lässt. Dieser F-Wert wird dann auf Signifikanz geprüft.

Es wird außerdem ein Signifikanztest für den „Haupteffekt“ der Messwiederholung durchgeführt. Dieser Effekt beschreibt den direkten Effekt des Faktors (Zeit: Messwiederholung im letzten Kindergartenjahr) auf die abhängige Variable (Leistungen in der phonologischen Informationsverarbeitung). Es werden die Effekte innerhalb von Personen betrachtet. Sind die Werte signifikant, dann kann von einem signifikanten Einfluss zwischen der Zeit und der Leistung ausgegangen werden. Hierbei wird auch das partielle Eta-Quadrat berechnet, das in dieser Studie in eine Effektstärke nach Cohen (1992, zit. nach Univ. Zürich 2016a) umgerechnet werden soll (vgl. Formel 16).

Formel 16: Berechnung der Effektstärke nach Cohen über das partielle Eta-Quadrat

$$f = \sqrt{\frac{\eta_p^2}{1 - \eta_p^2}}$$

Anmerkung: f = Effektstärke nach Cohen, η_p^2= Partielles Eta-Quadrat

Nach Cohen (1988, zit. nach Univ. Zürich 2016a) ist eine Effektstärke von f = .10 als schwach, f = .25 als moderat und f = .40 als stark einzuschätzen.

Wenn der F-Test zeigen konnte, dass ein Haupteffekt der Zeit (Abstände zwischen den Messungen) vorlag, wird danach ein Post-Hoc-Test durchgeführt, um genauer feststellen zu

können, bei welchen Messzeitpunkten signifikante Unterschiede in Bezug auf die Leistung in der phonologischen Informationsverarbeitung festzustellen sind. Der Post-hoc-Test berechnet für je zwei Mittelwerte einen t-Test. Aufgrund der hohen Anzahl der Tests steigt die Wahrscheinlichkeit, dass die Nullhypothese fälschlicherweise abgelehnt wird (Alpha-Fehler). Daher wird zusätzlich eine Bonferroni-Korrektur angewendet (vgl. Univ. Zürich 2016a). Alpha wird dabei durch die Anzahl der Paarvergleiche (hier 9, da es sich um 8 Gruppen handelt) dividiert und jeder Test wird durch dieses errechnete Niveau geprüft. Anhand der Signifikanzen kann nun festgestellt werden, ob die Zeit zwischen den Messungen signifikant zur Veränderung der Leistung beiträgt.

Zur Überprüfung der Hypothese H5.2 (*Die Leistung der Probanden steigt durchschnittlich um mindestens 1 Punkt zu jeder zweiten Messung.*) werden erneut die Mittelwertdifferenzen berechnet. Da die obige Berechnung über eine einfaktorielle Varianzanalyse mit Messwiederholung eine große Anzahl von Probanden aufgrund des fallweisen Ausschlusses aus der Berechnung ausschließt, werden für diese Hypothese die Mittelwertdifferenzen von je zwei Phonolino-Messungen, die zwei Messzeitpunkte auseinander liegen (z.B. Phono 2 und Phono 4), erneut über einen T-Test für abhängige Stichproben geprüft. Da es sich bei den zwei Phonolino-Messungen um die gleiche Stichprobe mit einer Messwiederholung handelt, liegen abhängige Stichproben vor, da die Messwerte von der gleichen Person stammen. Zur Überprüfung der statistischen Signifikanz wird eine zugehörige Teststatistik berechnet. Diese folgt einer theoretischen t-Verteilung. Der berechnete t-Wert wird dann auf Signifikanz geprüft. Dafür wird ein Vergleich zwischen der Teststatistik t und dem kritischen Wert, der durch die Freiheitsgrade bestimmten t-Verteilung, errechnet. Wenn der Betrag der Teststatistik höher ist als dieser kritische Wert, dann handelt es sich um einen signifikanten Unterschied.

Zur Überprüfung der Hypothese H5.3 (*Je größer der Abstand zwischen den Messzeitpunkten ist, umso kleiner sind die Korrelationen zwischen dem Gesamtwert von Phonolino bzw. zwischen den Items zum jeweiligen Messzeitpunkt.*) werden Korrelationen (vgl. Kap. VI.5.1.1, S. 316) zwischen den Phonolino-Messungen berechnet und grafisch anhand einer Trendlinie ausgewertet. Die Trendlinie zeigt an, ob die Korrelationen über die acht Messzeitpunkte hinweg fällt oder ansteigt, und kann damit aussagen, ob die Korrelationen über die Zeit abnehmen, zunehmen oder gleich bleiben.

5.4 Methoden zur Erweiterung des Prognosemodells

Auf theoretischer Basis gilt es zu überprüfen, ob das oben benannte möglich Prognosemodell, durch Phonolino die Lese- und Rechtschreibleistungen im ersten Schuljahr vorherzusagen, durch einen weiteren Prädiktor, den nonverbalen kognitiven Fähigkeiten, gemessen durch den CFT-Rohwert der Summe 1 (Gesamtsumme aller Subtests), verbessert werden kann. Um diese mögliche Verbesserung der Prognose statistisch nachweisen zu können, wird auf die logistische Regressionsanalyse sowie auf die Güteindizes für Klassifikationsanalysen (vgl. Kap. VI.5.2.3.4, S. 335) zurückgegriffen. Die Wahl dieser Methode beruht zum einen darauf, dass durch die Klassifikationsanalysen genauer dargestellt werden kann, wie viele Kinder mit unterdurchschnittlichen Leistungen im Lesen und Rechtschreiben durch das dargestellte Prognosemodell richtig identifiziert werden können (Sensitivität). Außerdem ist es für die Praxis sinnvoll auch über einzelne Kinder eine persönliche Aussage zu treffen, mit welcher Wahrscheinlichkeit diese Schwierigkeiten aufweisen werden (vgl. Walter 2014b, S. 23). Dies ist wiederum mit einer logistischen Regressionsanalyse machbar. Diese bietet zudem den Vorteil, dass die Risikopunkte bzw. der sogenannte Cut-Off-Wert der beteiligten Variablen zur Vorhersage der Lese- und Rechtschreibleistungen nicht festgelegt werden müssen, sondern durch die anhand der logistischen Regression berechneten Regressionskoeffizienten für die einzelnen Variablen innerhalb der Berechnung festgelegt werden.

Für diese Untersuchung wird die logistische Regressionsanalyse gerechnet. Es werden die acht einzelnen Phonolino-Messungen (Gesamtsumme) in Bezug auf ihre Vorhersageleistung einmal auf die Lese- und einmal auf die Rechtschreibleistung hin überprüft (Modell 1). Im Anschluss wird zu jeder der acht Phonolino-Messungen zusätzlich die Intelligenz der Probanden als Prädiktor zur Vorhersage des jeweiligen Kriteriums (Lesen oder Rechtschreiben) hinzugezogen (Modell 2). Hierfür wird bei der logistischen Regression bei der Auswahl der Kovariaten die Methode „Einschluss“ gewählt, was bedeutet, dass die ausgewählten Kovariaten in der Rechnung alle mit einbezogen werden. Die Ergebnisse (nur Phonolino-Messung vs. Phonolino-Messung + IQ) werden im Anschluss miteinander verglichen und anschließend durch klassifikatorische Analysen überprüft, die wie bereits in Kapitel VI.5.2.3.4 (vgl. S. 335) beschrieben, durch die Auswertungstabelle von Lenhard und Marx (2010) berechnet werden.

Die binäre logistische Regressionsanalyse kann einen Zusammenhang zwischen mehreren unabhängigen Variablen und einer binären abhängigen Variablen testen (vgl. Backhaus, Ericson, Plinke & Weiber 2006, S. 284). Sie ist ein strukturen-prüfendes Verfahren und

bildet eine Variante der Regressionsanalyse (vgl. a. a. O.). Im Vergleich zu einer einfachen linearen Regressionsanalyse, die einen empirischen Beobachtungswert als Ergebnis liefert, kann die logistische Regression die Eintretenswahrscheinlichkeit des empirischen Beobachtungswerts berechnen. Die logistische Regression ist ein multivariates Analyseverfahren, das dichotom abhängige Variablen analysiert (vgl. Best & Wolf 2010, S. 827). Dies sind Variablen, die zwei Ausprägungen haben. Dafür werden die Lese- und Rechtschreibleistungen im ersten Schuljahr, wie bereits für die Klassifikationsanalyse getätigt, in Kinder mit und Kinder ohne Lese- bzw. Rechtschreibschwierigkeiten in der ersten Klasse mit den Kennwerten 0 (Kind ohne Lese- oder Rechtschreibschwierigkeiten) und 1 (Kind mit Lese- oder Rechtschreibschwierigkeiten) dichotomisiert.

Die logistische Funktion (vgl. Abb. 57) bildet die Grundlage der logistischen Regression und lässt sich als Verteilungsfunktion interpretieren, die der Verteilungsfunktion der Normalverteilung ähnelt (vgl. a.a.O.). Die logistische Funktion kann als „s-förmig“ beschrieben werden und ist damit im Vergleich zur linearen Regressionsanalyse keine Gerade. Die Werte der logistischen Funktion liegen zwischen 0 und 1 (vgl. Univ. Zürich 2016b). Die Werte der logistischen Funktion sind als Wahrscheinlichkeit, dass die abhängige Variable den Wert 1 annimmt, interpretiert. Somit ist bei einem Wert in der Nähe von 0 die Eintretenswahrscheinlichkeit sehr unwahrscheinlich und bei einem Wert in der Nähe von 1 dagegen sehr wahrscheinlich (vgl. a. a. O.).

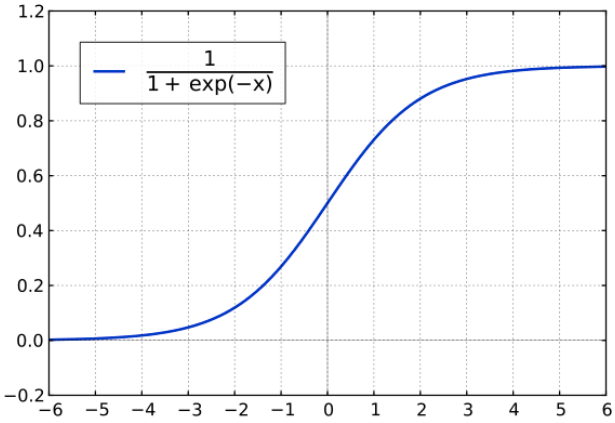


Abb. 57: Grafische Darstellung einer logistischen Funktion[29]

Die abhängige Variable wird bei der logistischen Regressionsanalyse meistens auch Y-Variable genannt, die unabhängigen Variablen sind die X-Variablen, also die Prädiktoren oder auch Kovariaten. Die logistische Regression nutzt die Maximum-Likelihood-Schätzung, während lineare Regressionsanalysen auf der Methode der kleinsten Quadrate

29 Bildquelle: Geek3, Wikimedia Commons, lizenziert unter CreativeCommons-Lizenz by-sa-2.0-de, URL: http://commons.wikimedia.org/wiki/File:mplwp_logistic

basieren (vgl. Univ. Zürich 2016b). Walter (2014b, S. 24) formuliert dazu den Vorteil dieser Schätzung:

> *„Damit optimiert das Verfahren auch das jeweilige Gewicht einer Prädiktorvariablen bezüglich der Prognose, so dass auch die à priori schwierig zu beantwortende Frage z.B. nach der Anzahl der Risikopunkte für eine Risiko-Klassifikation bzw. deren Bestimmung mittels eines Cutoff-Werts durch das Verfahren selbst geleistet wird.“*

Die logistische Regressionsfunktion wird in Formel 17 beschrieben (vgl. Univ. Zürich 2016b). Sie entsteht aus der Logit-Transformation und damit dem natürlichen Logarithmus der Odds (Chancen auf Erfolg), die durch $\frac{p}{1-p}$ beschrieben werden, wobei p die Wahrscheinlichkeit des Erfolgs beschreibt (vgl. Tarazona 2015, S. 6).

Formel 17: Logistische Regressionsfunktion
(vgl. Univ. Zürich 2016b)

$$P(y = 1) = \frac{1}{1 + e^{-z}}$$

Anmerkung: P (y=1)= Wahrscheinlichkeit, dass y=1, e = Eulersche Zahl, z = Logit (lineares Regressionsmodell der unabhängigen Variablen)

Bei der Vorhersage einer dichotomen Variable Y durch einen oder mehrere Prädiktoren x_1, x_2, ..., wird angenommen, dass zwischen den Prädiktoren und dem Logarithmus der Odds eine lineare Beziehung besteht, dass Y den Wert 1 annimmt (vgl. Formel 18). Dies beschreibt Z. Ist Z größer als 0, so nimmt Y den Wert 1 an, ist Z kleiner oder gleich 0, dann nimmt Y den Wert 0 an.

Formel 18: Logit (z) – lineares Regressionsmodell

$$Z = logit(p) = LN\left(\frac{p}{(1-p)}\right) = \beta_0 + \beta_1 x_1 + \cdots \beta_n x_n + u$$

Anmerkung: Es gilt p = P (Y=1)., Xn = unabhängige Variablen, β_n= Regressionskoeffizienten, u=Fehlerwert

Der Logit wird nun in die logistische Funktion eingesetzt (vgl. Formel 17).

Die Regressionskoeffizienten entstehen durch die Maximum-Likelihood-Schätzung. Diese werden für y = 1 so bestimmt, dass möglichst hohe und für y = 0 möglichst geringe Wahrscheinlichkeiten vorhergesagt werden (vgl. Univ. Zürich 2016). Dabei wird eine Likelihood-Funktion maximiert, die eine Aussage über die Wahrscheinlichkeit trifft, dass die unabhängigen Werte den abhängigen Wert voraussagen können. Positive

Regressionskoeffizienten bedeuten einen Anstieg der Wahrscheinlichkeit, dass y = 1 ist und negative Vorzeichen weisen auf eine Abnahme der Wahrscheinlichkeit hin (vgl. a. a. O.).
Die Regressionskoeffizienten werden nicht direkt, sondern über die Odds Ratios (Exp(B)) interpretiert, die aus dem Verhältnis zweier Odds bestehen. Die Berechnung der Odds Ratios in SPSS sind in Formel 19 dargestellt. Sie geben damit bei Anstieg einer unabhängigen Variablen um 1 die Veränderung der relativen Wahrscheinlichkeit von y = 1 an, während die anderen Variablen konstant gehalten werden. Ein Odds Ratio von 1 bedeutet damit keine Veränderung, ist Odds Ratio > 1 ist es eine Zunahme und bei < 1 eine Abnahme der Odds.

Formel 19: Berechnung der Odds Ratio in SPSS (vgl. Univ. Zürich 2016)

$$Odds\ Ratio = Exp(B) = \frac{Odds\ nach\ dem\ Anstieg\ von\ x\ um\ eine\ Einheit}{Odds\ vor\ dem\ Anstieg\ von\ x\ um\ eine\ Einheit}$$

Für die vorliegende Untersuchung wird nun zunächst das Regressionsmodell in Formel 20 für alle acht Phonolino-Messungen (Phono 2 bis Phono 9) einzeln berechnet. Das Modell wird in Bezug auf die unabhängige Variable „Leseleistung im ersten Schuljahr“ (Erfolg/Misserfolg) und auf die unabhängige Variable „Rechtschreibleistung im ersten Schuljahr“ (Erfolg/Misserfolg) jeweils separat überprüft. Hierbei wird „Erfolg im Lesen/Rechtschreiben“ mit y = 0 und „Misserfolg im Lesen/Rechtschreiben“ mit y = 1 codiert. Die abhängigen Variablen werden dabei in acht Berechnungen durch die Phonolino-Messungen (Gesamtwert) präsentiert und dienen der Vorhersage des Erfolgs/Misserfolgs im Lesen bzw. Rechtschreiben im ersten Schuljahr. Diese Kodierung wird genutzt, um die Wahrscheinlichkeit von erwarteten Schwierigkeiten im Lesen und Rechtschreiben berechnen zu können. Dies kann Vorteile haben, wenn die mit Regressionskoeffizienten ausgefüllte Formel später dafür verwendet wird, die Wahrscheinlichkeit für ein Risiko im Lesen und Rechtschreiben zu identifizieren, um den Kindern die notwendige Intervention zukommen lassen zu können.

Formel 20: Regressionsmodell 1 für die acht Phonolino-Messungen als Prädiktoren

$$P\ (Misserfolg\ im\ Lesen = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot Phono_k)}}$$

$$P\ (Misserfolg\ im\ Rechtschreiben = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot Phono_k)}}$$

Anmerkung: $Phono_k$= Wert der Phonolino-Messung 2 bis 9, β_n= Regressionskoeffizienten

Dieses Modell soll dann um den Prädiktor der nonverbalen kognitiven Leistungsfähigkeit, gemessen durch den CFT1, ergänzt werden (vgl. Formel 21). Im Anschluss werden Modell 1

und Modell 2 in Bezug auf ihre Prognoseleistung zu den jeweiligen Messzeitpunkten verglichen.

Formel 21: Regressionsmodell 2 für die jeweils acht Phonolino-Messungen und den kognitiven Leistungen der Probanden im letzten Kindergartenjahr als Prädiktoren

$$P\left(Misserfolg\ im\ Lesen = 1\right) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot Phono_k + \beta_2 \cdot CFT)}}$$

$$P\left(Misserfolg\ im\ Rechtschreiben = 1\right) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot Phono_k + \beta_2 \cdot CFT)}}$$

Anmerkung: $Phono_k$= Wert der Phonolino-Messung 2 bis 9, β_n= Regressionskoeffizienten, CFT = Messung für kognitive Leistung

Die Methodenauswahl für die Aufnahme der Variablen ist dabei in den Sozialwissenschaften relevant, da die Variablen oftmals nicht vollständig unkorreliert sind, was auch in dieser Untersuchung der Fall ist und als Kritik an der Wahl dieser Berechnungsmethode zu sehen ist. Bei den Modellen 1 und 2 (vgl. Formel 20 und Formel 21) soll jeweils mit der Einschlussmethode gerechnet werden. Dabei werden alle unabhängigen Variablen in einem einzigen Schritt aufgenommen. Die Auswahl der unabhängigen Variablen sollte selbstverständlich theoretisch fundiert sein (vgl. Univ. Zürich 2016b).

Die Voraussetzungen für die Durchführung einer binären logistischen Regression ist, dass die abhängige Variable binär codiert ist, die unabhängigen Variablen metrisch oder bei kategorialen Variablen als Dummy-Variablen codiert sind und dass die unabhängigen Variablen untereinander nicht hoch korrelieren (Multikollinearität, vgl. Univ. Zürich 2016b). Die Multikollinearität der Variablen führt dazu, dass die Treffsicherheit ungenauer wird und damit sich die Standardabweichung der geschätzten Werte vergrößert. Die Signifikanz der Koeffizienten sinkt (vgl. Wooldridge 2003, S. 97f.). Jedoch bleibt die Schätzung unverzerrt und varianzminimal (vgl. Stocker 2015, S. 2), damit bleiben die Konfidenzintervalle und Teststatistiken dennoch gültig. Wenn bei vorliegender Multikollinearität ein Koeffizient signifikant ist, kann dieser ganz normal interpretiert werden (vgl. a. a. O., S. 10). Besteht bei den Koeffizienten jedoch keine Signifikanz, kann der Grund die hohe Multikollinearität der Variablen sein, da die abhängige Variable aufgrund der hohen Korrelation der unabhängigen Variablen nicht genügend Informationen enthält, dies schlägt sich in einem großen Standardfehler nieder (vgl. a. a. O., S. 11). Somit kann die logistische Regression bei Multikollinearität durchgeführt werden., zu hohe Korrelationen zwischen den unabhängigen Variablen sollten dennoch vermieden werden. Aus diesem Grund werden die parallelen Messungen von Phonolino nicht innerhalb eines Modells abgebildet. Stocker (a. a. O. , S. 12) beschreibt, dass „die Frage, ab wann ein Multikollinearitätsproblem ‚zu groß' ist, ebenso

schwer zu beantworten *sei (Anm. des Verf.)* wie die Frage, ab wann eine Stichprobe ‚zu klein' ist." Außerdem sollte die Fallzahl mindestens N > 25 Fälle umfassen (vgl. Diaz-Bone 2013, S. 236).

Ob das Regressionsmodell einen Erklärungsbeitrag gegenüber der Vorhersage eines modalen Wertes leistet, wird durch die Signifikanz über einen Chi-Quadrat-Test geprüft. Dazu wird der logarithmierte Wert aus der Likelihood-Funktion genutzt (Log-Likelihood – LL) und mit -2 multipliziert. Diese -2LL-Werte des Nullmodells (berücksichtigt nur die Konstante) und des Regressionsmodelles werden dann auf Signifikanz hin überprüft. Dies wird über einen Chi-Quadrat-Test durchgeführt. Je höher der Chi-Quadrat-Wert ist, umso besser ist das „Modell-Fit" also die Passung des Modells. Wenn das Modell als Ganzes signifikant ist, dann kann die Analyse über die logistische Regression fortgesetzt werden (vgl. Univ. Zürich 2016b).

Zum Vergleich der Modelle 1 und 2 wird ein Chi-Quadrat-Differenz-Test berechnet. Dies ist möglich, wenn sich die Modelle nur in Bezug auf bestimmte Parameter unterscheiden. Dazu werden die Differenzen für Chi-Quadrat (entnommen aus dem Omnibus-Test der logistischen Regression) und die Differenzen der Freiheitsgrade (df) berechnet. Im Chi-Quadrat-Differenz-Test wird die Differenz der beiden Chi-Quadrat-Tests anhand der Differenz der Freiheitsgrade auf Signifikanz hin bewertet. Ein signifikanter Wert bedeutet, dass das Modell mit dem höheren Chi-Quadrat-Wert eine signifikant bessere Passung vorweist. Die Berechnung ist möglich, da die Differenz zweier Chi-Quadrat-Werte weiterhin eine Chi-Quadrat-Verteilung aufweist.

Die Regressionskoeffizienten werden ebenfalls auf Signifikanz hin überprüft. Dies geschieht über den Wald-Test, der den Regressionskoeffizienten (β) einer Variable durch den Standardfehler von β dividiert. Wird eine Signifikanz nachgewiesen, bedeutet dies, dass die Regressionskoeffizienten nicht Null sind und die jeweilige Variable einen signifikanten Einfluss auf die abhängige Variable hat. Der Einfluss der Variablen wird über Odds Ratio geprüft. Diese sind signifikant, wenn das Konfidenzintervall von Exp(B) den Wert 1 nicht mit einschließt (vgl. a.a.O.).

Um die Modellgüte beurteilen zu können, wird auf das Cox und Snell R-Quadrat sowie das Nagelkerke R-Quadrat zurückgegriffen. Diese messen die Stärke der Beziehung zwischen der abhängigen und den unabhängigen Variablen (vgl. Tarazona 2015, S. 13). Das Nagelkerke R-Quadrat kann nur Werte zwischen 0 und 1 annehmen (es standardisiert das Cox und Snell R-Quadrat). Je höher dieses ist, umso besser ist die Passung zwischen Modell

und Daten zu beurteilen. Um die Modellgüte genauer bewerten zu können, wird das Nagelkerke R-Quadrat in eine Effektstärke f nach Cohen (vgl. Cohen 1992, zit. in Univ. Zürich 2016b) umgerechnet (vgl. Formel 22, Univ. Zürich 2016b).

Formel 22: Berechnung der Effektstärke nach Cohen (1992)

$$f = \sqrt{\frac{R^2}{1 - R^2}}$$

Anmerkung: f = Effektstärke nach Cohen, R^2 = Nagelkerkes R-Quadrat

Die Effektstärke kann wie folgt beurteilt werden (vgl. Cohen 1988, zit. in Univ. Zürich 2016b):

- f = .10 entspricht einem schwachen Effekt
- f = .25 entspricht einem mittleren Effekt
- f = .40 entspricht einem starken Effekt

Über die Klassifizierungstabelle können die durch das jeweilige Modell richtig vorhergesagten Fälle festgestellt werden und zwischen verschiedenen Modellen verglichen werden. Die errechneten Klassifikationen werden dann im Anschluss über die Güteindizes für klassifikatorische Analysen (vgl. Kap. VI.5.2.3.4, S. 335) auf ihre Güte hin überprüft und verglichen.

5.5 Datenauswertung

Für die empirische Analyse der Daten der vorgelegten Studie wurde das statistische Programm SPSS Version 21 der Firma IBM verwendet. Außerdem wurde zur Ermittlung der klassifikatorischen Eckdaten eine Hilfe zur „Berechnung der Klassifikationsvalidität“ (vgl. Walter, o.J.) sowie die „Auswertungshilfe zur Berechnung von Testkennwerten von Screeningverfahren“ (vgl. Lenhard & Marx 2010) in Form einer Excel-Tabelle benutzt. Für die Berechnung der Effektstärken und des Chi-Quadrat-Differenz-Tests wird Excel als Berechnungsgrundlage verwendet.

VII. Darstellung der Ergebnisse

Im folgenden Abschnitt werden die berechneten Ergebnisse der Längsschnittstudie beschrieben. Zunächst werden die Ergebnisse zu den klassischen Testgütekriterien zur Objektivität, zur Reliabilität und zur Validität vorgestellt. Danach werden die Befunde zur Änderungssensibilität des curriculumbasierten Messverfahrens Phonolino präsentiert. Die Änderungssensibilität leistet ebenfalls einen Beitrag zur Konstruktvalidität des Verfahrens (vgl. Walter 2011a, S. 115f.) und wird aufgrund der eher nicht klassischen Vorgehensweise bei den oben dargestellten Früherkennungsverfahren hier explizit (im Sinne der Stufe 2 nach Fuchs, vgl. Kap. III.6, S. 249) betrachtet. Außerdem wird als Ausblick im letzten Teil der Ergebnisdarstellung das hier dargestellte Prognosemodell durch den Prädiktor der kognitiven Leistung der Probanden erweitert, um herauszufinden, ob eine Vorhersage der Lese- und Rechtschreibleistung dadurch optimiert werden kann.

1 Klassische Testgütekriterien

Im Folgenden werden die Ergebnisse zu den Kriterien zur Testgüte der untersuchten curriculumbasierten Messung Phonolino dargelegt. Zunächst wird die Objektivität des Verfahrens näher betrachtet. Danach werden die Datensätze für die Ergebnisse der Reliabilität ausgewertet. Im Anschluss wird die Validität der Messungen überprüft.

1.1 Objektivität

Die Objektivität der Phonolino-Messungen konnte durch standardisierte Instruktions- und Durchführungshinweise gesichert werden (vgl. Kap. VI.3.1). Die Erzieherinnen und Erzieher haben wortwörtlich vorformulierte Instruktionen erhalten, die Items wurden den Kindern auf CD präsentiert und die Durchführung wurde in gleicher Reihenfolge ausgeführt. Das genutzte Material wurde für alle Kindergärten einheitlich produziert. Selbst die begleitende Handpuppe, das Schaf Phonolino, wurde allen Kindergärten zur Verfügung gestellt. Ebenfalls die Vergabe der Punkte und eine feste vorgegebene Zeitspanne sind durch die standardisierte Auswertung vorgegeben. Die Untersuchungsleitung hat nach Abgabe der Phonolino-Hefte lediglich die Punkte der drei Items pro Messzeitpunkt einzeln aufsummiert.

1.2 Itemanalyse

Zunächst werden die Häufigkeitsverteilungen, Mittelwerte und Standardabweichungen für alle Items der acht Messungen (Phono 2 bis Phono 9) dargelegt. Im Anschluss werden dann die Itemschwierigkeiten (vgl. Tab. 55) und Trennschärfen sowie die Dimensionalität des Messverfahrens analysiert.

1.2.1 Analyse der Rohwertverteilung

Anhand der Histogramme zur Häufigkeitsverteilung über die acht Messzeitpunkte hinweg wird sichtbar, dass die Rohwerte der einzelnen Items „Pseudowörter", „Silben klopfen" und „Lautisolation" (vgl. Anhang 8) optisch weitestgehend normalverteilt sind. Zudem scheinen keine starken Deckeneffekte für die Gesamtgruppe aufzutreten, da alle drei Items bis zum letzten Messzeitpunkt keine erhöhte Anzahl an Kindern aufweisen, die im höheren Bereich viele Punkte erzielt haben. Dennoch ist insbesondere bei der ersten Messung (Phono 2) zu sehen, dass diese für die Kinder noch deutlich schwerer war (Linksverschiebung der Normalverteilung) als die folgenden Messungen.

Für die acht Messungen liegen für 184 bis 204 Vorschulkinder gültige Werte vor (bereinigt um die Kinder, die an nur unter 30 % der Messungen teilgenommen haben, vgl. Tab. 54).

Tab. 54: Deskriptive Daten

	Pseudowörter					Silben klopfen					Lautisolation				
MZP	N	MW	SD	Min	Max	N	MW	SD	Min	Max	N	MW	SD	Min	Max
Phono2	197	2,62	1,95	0	9	198	5,96	2,91	0	13	197	3,36	1,85	0	9
Phono3	197	3,15	1,93	0	9	196	6,77	2,75	0	14	197	4,03	1,91	0	8
Phono4	193	3,26	1,85	0	8	193	7,31	2,83	0	14	193	3,90	2,12	0	9
Phono5	204	3,65	1,87	0	9	205	7,33	2,95	0	13	205	4,32	2,01	0	13
Phono6	185	3,76	2,06	0	8	184	8,07	2,65	2	14	185	4,67	2,34	0	9
Phono7	184	3,93	2,10	0	9	184	8,27	2,76	1	14	184	4,97	2,21	0	9
Phono8	197	3,82	2,25	0	9	197	9,00	2,55	3	14	197	5,19	2,29	0	9
Phono9	193	4,24	2,13	0	9	193	8,92	2,74	2	14	193	5,28	2,10	0	9

Anmerkung: MZP = Messzeitpunkt, Anzahl der Vorschulkinder (N), Mittelwerte (MW), Standardabweichungen (SD), minimaler Rohwert (Min) und maximaler Rohwert (Max) für die Items "Pseudowörter", "Silben klopfen" und "Lautisolation" für die Messzeitpunkte Phono 2 bis Phono 9

Tab. 54 beschreibt die Mittelwerte, Standardabweichungen sowie die minimale und maximale erreichte Punktzahl zu den acht Messzeitpunkten der drei Items „Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“. Bis auf einen „Ausreißer“ (es handelt sich hier um ein Kind mit dem Rohwert 13, als einziges Kind mit einem Rohwert > 9) in Phono 5 beim Item „Lautisolation“ differieren die maximalen Rohwerte nur um maximal einen Rohwertpunkt (vgl. Anhang 8).

1.2.2 Itemschwierigkeit

Die Itemschwierigkeit lässt sich anhand der Mittelwerte der einzelnen Items darstellen (vgl. Anhang 8). Item 1 („Pseudowörter“) weist über die acht Messzeitpunkte hinweg Mittelwerte zwischen 2,62 und 4,24 bei einer maximal zu erreichenden Anzahl von 9 Rohwertpunkten auf. Die Mittelwerte sind liegen nah beieinander und es ist zu beobachten, dass sich die Werte über die Zeit hinweg leicht vergrößern. Es handelt sich um ein eher schwieriges Item. Die Erhöhung der Mittelwerte ist auch bei Item 2 („Silben klopfen“) und bei Item 3 („Lautisolation“) zu beobachten. Das Item „Silben klopfen“ zeigt bei einem Maximalwert von 14 Rohwertpunkten Mittelwerte von 5,96 bis 9,00 auf und das Item „Lautisolation“ hat mit einem Maximalwert von 13, Mittelwerte zwischen 3,36 und 5,28. Dieser Maximalwert von 13 kann auf einem Messfehler beruhen, da kein anderes Kind mehr als neun Punkte erreicht hat und das betroffene Kind diese Leistung auch in keiner der anderen Messungen replizieren konnte. Daher wird der Wert an dieser Stelle in die Berechnungen nicht miteinbezogen.

Nach der im Kapitel VI.5.2.1.2 dargelegten Formel 11 (vgl. S. 320) wurden die Itemschwierigkeiten der drei Items zu allen drei Messzeitpunkten berechnet. In Tab. 55 findet sich eine Übersicht der berechneten Itemschwierigkeiten. Aufgrund der Speedvariante der Messungen wurde als maximale Punktzahl des Items die von allen teilnehmenden Kindern maximal erreichte Punktzahl des jeweiligen Items über alle Messzeitpunkte (Phono 2 bis Phono 9) gewählt. Die Itemschwierigkeit gibt eine prozentuale Auskunft darüber, wie viele Kinder ein Item lösen konnten. 100 % bedeutet dabei, dass alle Kinder dieses Item mit höchster Punktzahl gelöst haben, 0 % bedeutet, dass kein Kind dieses Item mit höchster Punktzahl lösen konnte (was in diesem Fall nicht möglich ist, da der Maximalwert durch die höchste erreichte Punktzahl eines Probanden gewählt wurde). Für das Item „Pseudowörter“ konnte eine maximale Punktzahl von 9 Rohwerten über alle Messzeitpunkte hinweg festgestellt werden, beim Item „Silben klopfen“ war 14 die maximale Punktzahl, und beim Item „Lautisolation“ wurde k = 9 festgelegt, da der „Ausreißer“ k = 13 ausgeschlossen wurde.

Das Item „Pseudowörter nachsprechen“ hat eine Itemschwierigkeit von 29,11 % bis 47,11 %, wobei die Itemschwierigkeit vom ersten Messzeitpunkt bis zum achten Messzeitpunkt kontinuierlich ansteigt (vgl. Tab. 55). Dies ist ebenfalls beim Item „Silben klopfen“ sichtbar. Die Itemschwierigkeiten liegen zwischen 42,57 % und 64,29 %, wobei nur Phono 9 leicht niedriger ist, als Phono 8, alle anderen Werte jedoch ebenfalls kontinuierlich ansteigen.
Das Item „Lautisolation“ verzeichnet Itemschwierigkeiten zwischen 37,33 % und 58,67 %. Diese steigen ebenfalls über die Messzeitpunkte hinweg kontinuierlich an.

Tab. 55: Itemschwierigkeit für Phono 2 bis Phono 9, k = maximale Punktzahl des jeweiligen Items über alle acht Messezeitpunkte hinweg

	Itemschwierigkeit in %		
Messzeitpunkt	Pseudowörter (k = 9)	Silben klopfen (k = 14)	Lautisolation (k = 9)
Phono 2	29,11 %	42,57 %	37,33 %
Phono 3	35,00 %	48,36 %	44,78 %
Phono 4	36,22 %	52,21 %	43,33 %
Phono 5	40,56 %	52,36 %	48,00 %
Phono 6	41,78 %	57,64 %	51,89 %
Phono 7	43,67 %	59,07 %	55,22 %
Phono 8	42,44 %	64,29 %	57,67 %
Phono 9	47,11 %	63,71 %	58,67 %

Das Item „Silben klopfen“ liegt damit in Bezug auf die Schwierigkeit im mittleren Bereich. Das Item „Pseudowörter“ kann als eher schweres bis mittleres Item bezeichnet werden, während die „Lautisolation“ etwas leichter ist als das „Pseudowörter nachsprechen“, wobei die Kennwerte nur teilweise unter 50 % liegen und ab der fünften Messung bereits die 50 % und damit die mittlere Itemschwierigkeit überschreiten.

1.2.3 Trennschärfe

Die Messungen von Phonolino zu den acht Messzeitpunkten weisen bis auf drei Ausnahmen hohe Trennschärfekoeffizienten auf (vgl. Tab. 56). Die Ausnahmen lassen sich auf der Grundlage von Bortz und Döring (2006, S. 220) ebenfalls noch als mittelmäßig beschreiben.

Ein genauerer Überblick über die Item-Skala-Statistiken ist im Anhang zu finden (vgl. Anhang 9).

Tab. 56: Trennschärfekoeffizienten von Phono 2 bis Phono 9 für alle drei Items

	Korrigierte Item-Skala-Korrelation		
Messzeitpunkt	Pseudowörter	Silben klopfen	Lautisolation
Phono2	,447	,533*	,554*
Phono3	,554*	,702*	,645*
Phono4	,354	,636*	,596*
Phono5	,552*	,646*	,486
Phono6	,588*	,750*	,600*
Phono7	,519*	,625*	,575*
Phono8	,660*	,605*	,680*
Phono9	,637*	,604*	,616*

Anmerkung: * hohe Korrelation > 0,5

1.2.4 Dimensionalität

Die Dimensionalität wurde anhand einer Faktorenanalyse überprüft (vgl. Kap. VI.5.2.1.4, S. 322 Dimensionalität, vgl. Anhang 10). Phonolino wird in Bezug auf drei Faktoren analysiert, die sich aus der Theorie heraus begründen („Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“). Die Faktoren gehören alle zu dem Konstrukt der phonologischen Informationsverarbeitung, lassen sich jedoch theoriegeleitet begründet durch unterschiedlich geforderte Fähigkeiten begründen.

Das Kaiser-Meyer-Olkin-Kriterium (KMO) prüft zunächst, ob die Durchführung einer Faktorenanalyse sinnvoll erscheint. Das KMO hat einen Wert von .894 und ist damit größer als .50 und nach Bühner (2006, S. 206ff.) für eine Faktorenanalyse gut geeignet (vgl. Tab. 47, S. 323). Der Bartlett-Test auf Sphärizität konnte zeigen, dass die Abweichung von der Nullhypothese (die erhobenen Kennwerte stammen alle aus einer Stichprobe mit lauter Nullkorrelationen) hochsignifikant ist ($p = .000$).

Die Kommunalitäten werden in Tab. 57 beschrieben. Bis auf einen Wert (Phono 2 – Lautisolation) liegen alle Werte über .50 und sind damit für eine Faktorenanalyse als fair einzustufen (vgl. Tab. 48). Alle Werte von Phono 2 sind etwas niedriger als die anderen Kommunalitäten, dies kann an den Anfangsschwierigkeiten der Messungen liegen.

Tab. 57: Kommunalitäten anhand der Extraktionsmethode (Analyse der Hauptkomponente)

Item 1		Item 2		Item 3	
Messzeitpunkt	Extraktion	Messzeitpunkt	Extraktion	Messzeitpunkt	Extraktion
Phono 2 - PSW	,517	Phono 2 - SK	,511	Phono 2 - LI	,460
Phono 3 - PSW	,566	Phono 3 - SK	,722	Phono 3 - LI	,641
Phono 4 - PSW	,638	Phono 4 - SK	,670	Phono 4 - LI	,662
Phono 5 - PSW	,619	Phono 5 - SK	,597	Phono 5 - LI	,512
Phono 6 - PSW	,612	Phono 6 - SK	,676	Phono 6 - LI	,722
Phono 7 - PSW	,656	Phono 7 - SK	,733	Phono 97 -	,684
Phono 8 - PSW	,764	Phono 8 - SK	,668	Phono 8 - LI	,722
Phono 9 - PSW	,716	Phono 9 - SK	,621	Phono 9 - LI	,640

Anmerkung: PSW=Pseudowörter nachsprechen, SK= Silben klopfen, LI = Lautisolation

Die Faktorladungen, die durch die Hauptkomponentenanalyse durch Varimax-Rotation produziert wurden, sind für jedes einzelne Item der Tab. 58 zu entnehmen.

Tab. 58: Rotierte Komponentenmatrix (Rotationsmethode: Varimax)

Messzeitpunkt	1	2	3
Phono 2 - PSW	,115	,085	,705
Phono 3 - PSW	,150	,225	,702
Phono 4 - PSW	,029	-,058	,796
Phono 5 - PSW	,146	,192	,749
Phono 6 - PSW	,287	,169	,708
Phono 7 - PSW	,072	,173	,788
Phono 8 - PSW	,234	,226	,811
Phono 9 - PSW	,231	,211	,786
Phono 2 - SK	,033	,660	,272
Phono 3 - SK	,237	,792	,199
Phono 4 - SK	,175	,783	,160
Phono 5 - SK	,319	,690	,139
Phono 6 - SK	,394	,677	,251
Phono 7 - SK	,159	,836	,097
Phono 8 - SK	,201	,789	,066
Phono 9 - SK	,222	,749	,102
Phono 2 - LI	,575	,346	,102
Phono 3 - LI	,724	,300	,163
Phono 4 - LI	,778	,204	,120
Phono 5 - LI	,696	,092	,137
Phono 6 - LI	,823	,176	,112
Phono 7 - LI	,804	,172	,089
Phono 8 - LI	,775	,188	,294
Phono 9 - LI	,746	,178	,227

Anmerkung: blau hervorgehoben = zusammengehörige Faktoren

Wie in Tab. 58 anhand der farbigen Markierungen sichtbar, entsprechen die Items zum „Pseudowörter nachsprechen“, zum „Silben klopfen“ und zur „Lautisolation“ einem Faktor. Sie weisen alle eine Faktorladung über .40 auf und stellen damit eine interpretierbare Repräsentation der Faktoren „Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“ dar.

1.3 Reliabilität

Im Folgenden werden die Ergebnisse zur Paralleltestreliabilität, Split-Half-Reliabilität sowie zur Internen Konsistenz berichtet.

1.3.1 Paralleltestreliabilität

In diesem Abschnitt werden die Ergebnisse zur Paralleltestreliabilität dargestellt und anhand der berechneten Korrelationskoeffizienten interpretiert. Dazu wurden zunächst die Korrelationen der nebeneinanderliegenden Messungen des Gesamtrohwerts einer Messung und dann die je drei Items einzeln („Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“) zu den jeweiligen Messzeitpunkten verglichen.

In den unten präsentierten vier Tabellen (vgl. Tab. 59, Tab. 60, Tab. 61, Tab. 62) sind die Korrelationen zwischen allen Messzeitpunkten abgebildet. Alle Korrelationen werden als signifikant angezeigt. Die jeweiligen nebeneinanderliegenden Messungen und deren Werte sind blau hinterlegt. Nimmt man alle Korrelationen in den Blick (nicht nur die nebeneinanderliegenden Messungen) liegen die Korrelationskoeffizienten zwischen r = .335 (Phono 2 und Phono 9 – Lautisolation) und r = .823 (Phono 8 und Phono 9 - Gesamtmessung) und sind damit als moderat bis hoch einzustufen (vgl. Tab. 45, S. 316).

Zur Untersuchung der Paralleltestreliabilität der acht curriculumbasierten Messungen werden die nebeneinanderliegenden Werte verglichen, da diese den geringsten Lernentwicklungen unterliegen, die aufgrund der Zeit (ca. 1 Monat zwischen den Messungen) zu vermuten ist.

Zunächst werden die Korrelationen für die Gesamtmessung (aufsummierte Rohwerte der Items PSW, SK, LI zum jeweiligen Messzeitpunkt) der nebeneinanderliegenden Messungen verglichen (vgl. Tab. 59 – blau hinterlegte Felder). Für diese aufsummierten Rohwerte ergaben sich Korrelationskoeffizienten zwischen r = .644 (Phono 2 und Phono 3) und r = .823 (Phono 8 und Phono 9) mit einer Stichprobengröße zwischen N = 157 und N = 184.

Die Korrelationen sind auf dem 0,01 %-Signifikanzniveau als hoch signifikant einzustufen. Alle Zusammenhänge können als hohe Korrelationen interpretiert werden (vgl. Tab. 45, S. 316).

Tab. 59: Korrelationen zwischen der Gesamtsumme zu den Messzeitpunkten Phono 2 bis Phono 9

		Phono 2	Phono 3	Phono 4	Phono 5	Phono 6	Phono 7	Phono 8	Phono 9
Phono 2	r	1	,644**	,686**	,610**	,564**	,562**	,612**	,469**
	Sig.		,000	,000	,000	,000	,000	,000	,000
	N	198	181	177	189	168	168	181	179
Phono 3	r		1	,731**	,716**	,729**	,691**	,736**	,701**
	Sig.			,000	,000	,000	,000	,000	,000
	N		197	183	187	169	167	181	176
Phono 4	r			1	,731**	,725**	,692**	,700**	,637**
	Sig.				,000	,000	,000	,000	,000
	N			194	184	166	169	179	174
Phono 5	r				1	,777**	,742**	,721**	,631**
	Sig.					,000	,000	,000	,000
	N				204	174	175	188	184
Phono 6	r					1	,757**	,790**	,728**
	Sig.						,000	,000	,000
	N					185	157	169	168
Phono 7	r						1	,806**	,785**
	Sig.							,000	,000
	N						185	171	165
Phono 8	r							1	,823**
	Sig.								,000
	N							198	179
Phono 9	r								1
	Sig.								
	N								194

Anmerkungen: Sig.= Signifikanzniveau (2-seitig), N= Anzahl der Probanden, r = Korrelationskoeffizient
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant, grau hervorgehoben sind alle r < 0,5, blau hervorgehoben sind die direkt nebeneinander liegenden Messungen

Die Items zum *„Pseudowörter nachsprechen" (PSW)* (vgl. Tab. 60) ergaben bei Betrachtung der nebeneinanderliegenden Messungen Korrelationskoeffizienten zwischen r = .454 (Phono 2 und Phono 3) und r = .765 (Phono 8 und Phono 9) mit einer Stichprobengröße zwischen N = 158 und N = 185. Alle Korrelationen sind auf dem 0,1 %-Niveau statistisch hoch signifikant. Alle Korrelationskoeffizienten sind als mittlere Korrelation (Phono 2 und Phono

3) und als hohe Korrelation (Phono 3 und Phono 4, Phono 4 und Phono 5, Phono 5 und Phono 6, Phono 6 und Phono 7, Phono 7 und Phono 8, Phono 8 und Phono 9) zu bewerten.

Tab. 60: Korrelationen zwischen den Items „Pseudowörter nachsprechen“ zu den Messzeitpunkten Phono 2 bis Phono 9

		Phono 2	Phono 3	Phono 4	Phono 5	Phono 6	Phono 7	Phono 8	Phono 9
Phono 2	r	1	,454**	,568**	,636**	,433**	,453**	,503**	,533**
	Sig.		,000	,000	,000	,000	,000	,000	,000
	N	198	182	177	190	169	168	181	179
Phono 3	r		1	,521**	,531**	,515**	,505**	,619**	,556**
	Sig.			,000	,000	,000	,000	,000	,000
	N		198	184	189	171	168	182	177
Phono 4	r			1	,545**	,484**	,531**	,555**	,577**
	Sig.				,000	,000	,000	,000	,000
	N			194	185	167	169	179	174
Phono 5	r				1	,517**	,596**	,564**	,594**
	Sig.					,000	,000	,000	,000
	N				205	176	176	189	185
Phono 6	r					1	,567**	,652**	,642**
	Sig.						,000	,000	,000
	N					186	158	170	169
Phono 7	r						1	,660**	,616**
	Sig.							,000	,000
	N						185	171	165
Phono 8	r							1	,765**
	Sig.								,000
	N							198	179
Phono 9	r								1
	Sig.								
	N								194

Anmerkungen: Sig.= Signifikanzniveau (2-seitig), N= Anzahl der Probanden, r = Korrelationskoeffizient
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant, grau hervorgehoben sind alle r < 0,5, blau hervorgehoben sind die direkt nebeneinander liegenden Messungen

Die Items zum *„Silben klopfen“ (SK)* (vgl. Tab. 61) ergaben bei Betrachtung der nebeneinanderliegenden Messungen Korrelationskoeffizienten zwischen r = .568 (Phono 2 und Phono 3) und r = .716 (Phono 5 und Phono 6) mit einer Stichprobengröße zwischen N = 157 und N = 185. Alle Korrelationen sind auf dem 0,1 %-Niveau statistisch hoch signifikant. Alle Korrelationskoeffizienten sind als hohe Korrelation zu bewerten.

Tab. 61: Korrelationen zwischen den Items „Silben klopfen“ zu den Messzeitpunkten Phono 2 bis Phono 9

		Phono 2	Phono 3	Phono 4	Phono 5	Phono 6	Phono 7	Phono 8	Phono 9
Phono 2	r	1	,568**	,570**	,461**	,414**	,472**	,493**	,337**
	Sig.		,000	,000	,000	,000	,000	,000	,000
	N	199	182	178	192	169	169	182	180
Phono 3	r		1	,650**	,600**	,564**	,570**	,571**	,550**
	Sig.			,000	,000	,000	,000	,000	,000
	N		197	183	188	169	167	181	176
Phono 4	r			1	,609**	,598**	,571**	,559**	,504**
	Sig.				,000	,000	,000	,000	,000
	N			194	185	166	169	179	174
Phono 5	r				1	,716**	,568**	,616**	,531**
	Sig.					,000	,000	,000	,000
	N				206	176	177	190	186
Phono 6	r					1	,636**	,638**	,560**
	Sig.						,000	,000	,000
	N					185	157	169	168
Phono 7	r						1	,684**	,680**
	Sig.							,000	,000
	N						185	171	165
Phono 8	r							1	,660**
	Sig.								,000
	N							198	179
Phono 9	r								1
	Sig.								
	N								194

Anmerkungen: Sig.= Signifikanzniveau (2-seitig), N= Anzahl der Probanden, r = Korrelationskoeffizient
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant, grau hervorgehoben sind alle r < 0,5, blau hervorgehoben sind die direkt nebeneinander liegenden Messungen

Die Items zur *„Lautisolation“ (LI)* (vgl. Tab. 62) ergaben bei Betrachtung der nebeneinanderliegenden Messungen Korrelationskoeffizienten zwischen r = .520 (Phono 2 und Phono 3) und r = .705 (Phono 8 und Phono 9) mit einer Stichprobengröße zwischen N = 158 und N = 185. Alle Korrelationen sind auf dem 0,1 %-Niveau statistisch hoch signifikant. Alle Korrelationskoeffizienten sind als hohe Korrelationen zu bewerten.

Tab. 62: Korrelationen zwischen den Items „Lautisolation“ zu den Messzeitpunkten Phono 2 bis Phono 9

		Phono 2	Phono 3	Phono 4	Phono 5	Phono 6	Phono 7	Phono 8	Phono 9
Phono 2	r	1	,520**	,585**	,393**	,403**	,404**	,456**	,335**
	Sig.		,000	,000	,000	,000	,000	,000	,000
	N	198	182	177	191	169	168	181	179
Phono 3	r		1	,640**	,460**	,541**	,628**	,517**	,538**
	Sig.			,000	,000	,000	,000	,000	,000
	N		198	184	189	171	168	182	177
Phono 4	r			1	,529**	,638**	,612**	,587**	,528**
	Sig.				,000	,000	,000	,000	,000
	N			194	185	167	169	179	174
Phono 5	r				1	,558**	,597**	,538**	,414**
	Sig.					,000	,000	,000	,000
	N				206	177	177	190	186
Phono 6	r					1	,699**	,659**	,554**
	Sig.						,000	,000	,000
	N					186	158	170	169
Phono 7	r						1	,634**	,615**
	Sig.							,000	,000
	N						185	171	165
Phono 8	r							1	,705**
	Sig.								,000
	N							198	179
Phono 9	r								1
	Sig.								
	N								194

Anmerkungen: Sig.= Signifikanzniveau (2-seitig), N= Anzahl der Probanden, r = Korrelationskoeffizient
**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant, grau hervorgehoben sind alle r < 0,5, blau hervorgehoben sind die direkt nebeneinander liegenden Messungen

1.3.2 Split-Half-Reliabilität

Zur Berechnung der Split-Half-Reliabilität wurden die Messungen von Phonolino gleichmäßig auf zwei Hälften verteilt, wobei darauf geachtet wurde, dass die Messungen immer abwechselnd einer Hälfte zugeteilt wurden. Da Phonolino als Veränderungsmessung konzipiert ist, würde eine Aufteilung zur Hälfte des Kindergartenjahres eine Verfälschung des Reliabilitätskoeffizienten darstellen. Somit wurde die unten stehende Aufteilung gewählt (vgl. Tab. 63), um die zeitlichen Verbesserung von Phono 4 auf Phono 3 bspw. mit der angenommenen ähnlich großen Verschlechterung von Phono 7 auf Phono 6 auszugleichen. Aufgrund der je drei Items und damit einer ungeraden Anzahl an Items pro Messung wird

die Split-Half-Reliabilität dreimal berechnet: 1) für die Items Pseudowörter und Silben klopfen, 2) für die Items Silben klopfen und Lautisolation und 3) für die Items Pseudowörter und Lautisolation (vgl. Tab. 63). Hierfür wird in der ersten Hälfte immer ein Item auch aus der ersten Hälfte der Messung verwendet (Pseudowörter nachsprechen oder Silben klopfen) und in der zweiten Hälfte auch immer ein Item aus der zweiten Hälfte einer Messung (Silben klopfen oder Lautisolation). Somit werden pro Testhälfte je vier Items also insgesamt 8 Items miteinander verglichen.

Tab. 63: Aufteilung der Messungen für die Berechnung der Split-Half-Reliabilität

1. Hälfte	2. Hälfte
Phono 2, Phono 4, Phono 7, Phono 9 • Pseudowörter • Silben klopfen • Pseudowörter	Phono 3, Phono 5, Phono 6, Phono 8 • Silben klopfen • Lautisolation • Lautisolation

Die Split-Half-Reliabilität entspricht nun den Korrelationen der beiden Hälften der Messungen (vgl. Bortz & Döring 2006, S. 198). Die Ergebnisse für die drei Berechnungen der Split-Half-Reliabilitäten sind in Tab. 64 zu sehen. Für die Berechnung zu den Items „Pseudowörter“ und „Lautisolation“ sind bei beiden Hälften der Messungen bereits zufriedenstellende Reliabilitäten (Cronbachs Alpha) zu finden. Die zweite Hälfte weist eine leicht höhere interne Konsistenz (α = .838) als die erste Hälfte (α = .820) auf. Der Spearman-Brown-Koeffizient (r = .538) weist eine mittlere Reliabilität auf. Für die Berechnung zu den Items „Silben klopfen“ und „Lautisolation“ sind ebenfalls bei beiden Hälften der Messungen bereits zufriedenstellende Reliabilitäten (Cronbachs Alpha) zu finden.

Die zweite Hälfte weist eine leicht höhere interne Konsistenz (α = .840) als die erste Hälfte (α = .822) auf. Der Spearman-Brown-Koeffizient (r =. 609) weist eine mittlere Reliabilität auf. Für die Berechnung zu den Items „Pseudowörter nachsprechen“ und „Silben klopfen“ sind ebenfalls bei beiden Hälften der Messungen bereits zufriedenstellende Reliabilitäten (Cronbachs Alpha) zu finden. Die zweite Hälfte weist eine leicht höhere interne Konsistenz (α = .864) als die erste Hälfte (α = .828) auf. Der Spearman-Brown-Koeffizient (r =. 512) weist eine mittlere Reliabilität auf. Für alle drei Berechnungen konnten N = 92 bis N = 95 Fälle berücksichtigt werden, dies entspricht einem Anteil der Probanden der Gesamtstichprobe von 43 bis 44,4 %.

Tab. 64: Ergebnisse zur Split-Half-Reliabilität für drei Modelle

	PSW und LI	SK und LI	PSW und SK
N	94	95	92
Cronbachs Alpha Teil 1	.820	.822	.828
Cronbachs Alpha Teil 2	.838	.840	.864
Korrelation Teil 1 und Teil 2	.368	.438	.344
Spearman-Brown-Koeffizient	.538	.609	.512

Anmerkungen: N= Anzahl der Probanden, PSW = Pseudowörter nachsprechen, SK = Silben klopfen, LI = Lautisolation

1.3.3 Interne Konsistenz

Die Einschätzung der internen Konsistenz wird über Cronbachs Alpha bestimmt. Es ergeben sich zu den einzelnen CBM Reliabilitätskoeffizienten Werte zwischen $\alpha = .485$ und $\alpha = .625$ (vgl. Anhang 11). Die Anzahl der Probanden liegt in dieser Berechnung zwischen N = 184 und N = 204.

1.4 Validität

Die Validität beinhaltet verschiedene zu betrachtende Teilbereiche (vgl. Kap. I.2.1.3 - Validität).

Die *Inhaltsvalidität* wird an dieser Stelle nicht weiter geprüft, da die Items und deren Inhalte bereits in der theoretischen Darlegung (vgl. Kap. I - III) und in der Beschreibung der Konstruktion der CBM Phonolino (vgl. Kap. IV.2) mit dem aktuellen Stand der Forschung verglichen worden sind.

Die *kriteriumsorientierte Validität* unterteilt sich in verschiedene Aspekte (vgl. Kap. I.2.1.3) und wird in dieser Untersuchung empirisch und kriterienbezogen anhand der Übereinstimmungsvalidität und der prognostischen Validität überprüft.

1.4.1 Übereinstimmungsvalidität

Die *Übereinstimmungsvalidität oder auch Konvergenzvalidität* wird anhand von Produkt-Moment-Korrelationen nach Pearson und die Signifikanz durch Weber (vgl. Weber in Clauß und Ebner, 1968 aus Clauß, Finze & Partzsch 2004, S. 64f.) zum Kriterium des Bielefelder Screenings (BISC) überprüft. Dazu wurde eine Korrelationsanalyse zu den curriculumbasierten Messungen Phono 2 (Zeitpunkt 1) bzw. Phono 6 und Phono 7 (Zeitpunkt 2) und der darin erreichten Anzahl der korrekt bearbeiteten Items als Rohwert sowie den Risikopunkten von BISC 1 (Zeitpunkt 1) bzw. BISC 2 (Zeitpunkt 2) als Gesamtwert, sowie den Rohwerten dreier weiterer Untertests des Bielefelder Screenings durchgeführt. Die Untertests des Bielefelder Screenings werden mit den inhaltlich ähnlichen Items der Phonolino-Messungen korreliert, d.h. „Pseudowörter“ aus dem BISC mit „Pseudowörter nachsprechen“, „Silben segmentieren“ aus dem BISC mit „Silben klopfen“ und „Laut-zu-Wort Vergleich“ aus dem BISC mit „Lautisolation“. Der Messzeitpunkt „Phono 2“ wurde gewählt, da die Messung im gleichen Monat wie die erste Testung des Bielefelder Screening durchgeführt wurde. Phono 6 und Phono 7 wurden beide ausgewählt, da die Testung des BISC 2 zwischen diesen beiden Messungen stattgefunden hat.

Der Untertest „Pseudowörter“ aus dem BISC ähnelt dem Item „Pseudowörter nachsprechen“ von Phonolino in Bezug auf die Aufgabenstellung und die Art und Weise der Durchführung sehr. Es ist ebenfalls zur Erfassung des phonetischen Rekodierens im Kurzzeitgedächtnis geeignet. Bei beiden Aufgaben werden die Aufgabenstellungen dem Kind von einer CD Pseudowörter vorgesprochen, die es dann nachsprechen soll. Alle richtig produzierten Wörter werden notiert. Falsch artikulierte oder in der Reihenfolge vertauschte oder auch vergessene Silben werden als Fehler gewertet. Der Unterschied von BISC und Phonolino besteht in der Wortauswahl. Das BISC nutzt bspw. das Wort „Fan-go-fän-ger“, das zweimal den Nasalkonsonanten [ŋ] umfasst, der in der deutschen Sprache weniger häufig Verwendung findet und üblicherweise ohne hörbare Verschlussöffnung ([g]) gebildet wird. Auf diesen Laut wurde in Phonolino verzichtet. Dagegen enthalten die Pseudowörter von Phonolino Mehrfachkonsonanzen innerhalb der Silben. Sowohl im BISC als auch in Phonolino sind Wörter zwischen vier und sechs Silben enthalten.

Der Untertest „Silben segmentieren“ aus dem BISC ähnelt inhaltlich dem Item „Silben klopfen“ aus Phonolino. Beim Silben segmentieren hört sich das Kind ein Wort auf der CD an und muss es anschließend in Silben klatschen. Die ausgewählten Wörter in den Phonolino-Messungen haben zwischen einer und vier Silben, während im BISC Wörter mit

zwei oder drei Silben segmentiert werden und somit der häufig fälschlich segmentierte Fall eines einsilbigen Wortes ausgelassen wird (Bsp.: „Wu-hurst“).
Der Untertest „Laut-zu-Wort Vergleich“ aus dem BISC ist eine Überprüfung der phonologischen Bewusstheit im engeren Sinne, in dem das Kind Laute heraushören soll. Dies ist bei dem Item „Lautisolation“ in Phonolino ebenfalls der Fall. Die Ziellaute im BISC umfassen ausschließlich Vokale oder Vokalverbindungen (au, ei, e, u, i, a, o u), während bei Phonolino hauptsächlich Konsonanten herausgehört werden müssen. Zudem soll bei Phonolino die Lautposition (zu Beginn oder am Ende des Wortes) bestimmt werden, während im BISC ausschließlich entschieden werden muss, ob ein Laut in einem Wort enthalten ist. Damit ähneln sich die Übungen zwar, sind jedoch vom Anspruch her verschieden.
Aus dem Grund der Ähnlichkeit der Untertests des BISC und den Items von Phonolino liegt es auf der Hand auch die einzelnen Rohwertpunkte miteinander zu vergleichen.

Die Korrelationen zwischen den Risikopunkten (je mehr Risikopunkte ein Kind hat, umso schlechter sind seine Leistungen in der phonologischen Informationsverarbeitung) des BISC und den Rohwerten aus Phonolino (je mehr Punkte ein Kind erreicht, umso besser sind seine Leistungen in der phonologischen Informationsverarbeitung) werden vermutlich negativ ausfallen, da sie auf einem negativen kausalen Zusammenhang basieren (je mehr Rohwertpunkte ein Kind in Phonolino hat, umso weniger Risikopunkte sollte es im Bielefelder Screening haben). Eine gesamte Liste der Korrelationen zwischen dem BISC und deren Untertests sowie Phono 2, Phono 6 und Phono 7 ist im Anhang 12 zu finden.

Die Übereinstimmungsvalidität konnte für den ersten Zeitpunkt (BISC 1 - im November 2008) anhand einer Stichprobe zwischen N = 190 und N = 193 kontrolliert werden. Es ergaben sich für Phono 2 und den Gesamtwert des BISC (Risikopunkte) eine Korrelation von r = - .274. Der negative Zusammenhang ist wie vermutet eingetreten und signifikant. Dieser Zusammenhang ist als mittlere Korrelation zu bewerten. Ebenfalls der Zusammenhang zwischen Phono 2 und dem Untertest „Pseudowörter“ (r = .251), dem Untertest „Silben segmentieren“ (r = .231) und dem Untertest „Laut-zu-Wort Vergleich“ (r = .255) ist signifikant und als mittlerer Zusammenhang zu interpretieren (vgl. Tab. 45).

Für den zweiten Zeitpunkt (BISC 2 - im April/ Mai 2009) bestand die Stichprobe aus N = 167 bis N = 169 Kindern. Die Korrelation zwischen den Risikopunkten des BISC und Phono 6 (r = -.546) sowie Phono 7 (r = -.441) kann als hohe Korrelation interpretiert werden. Ebenfalls ist der antizipierte negative Zusammenhang eingetreten. Die Korrelationen aller

Untertests des BISC mit Phono 6 und Phono 7 sind hochsignifikant und liegen zwischen r = .313 und r = .457. Sie sind damit als mittlere bis hohe Zusammenhänge einzuschätzen. Eine Übersicht der Korrelationen befindet sich in Tabelle 68.

Tab. 65: Korrelationen zwischen den Messungen Phono 2 und BISC1 sowie Phono 6 und Phono 7 und BISC 2

	Phono 2
BISC 1 (Risikopunkte)	Gesamtrohwert N = 190 r = - .274**
BISC 1 (Pseudowörter)	Pseudowörter nachsprechen N = 192 r = .251**
BISC 1 (Silben segmentieren)	Silben klopfen N = 193 r = .231**
BISC 1 (Laut-zu-Wort Vergleich)	Lautisolation N = 192 r = .255**

	Phono 6	Phono 7
BISC 2 (Risikopunkte)	Gesamtrohwert N = 167 r = - .546**	Gesamtrohwert N = 169 r = - .441**
BISC 2 (Pseudowörter)	Pseudowörter nachsprechen N = 168 r = .447**	Pseudowörter nachsprechen N = 169 r = .457**
BISC 2 (Silben segmentieren)	Silben klopfen N = 167 r = .447**	Silben klopfen N = 169 r = .313**
BISC 2 (Laut-zu-Wort Vergleich)	Lautisolation N = 168 r = .411**	Lautisolation N = 169 r = .435**

Anmerkung: : **= Korrelation ist bei Niveau 0,01 signifikant, * Korrelation ist bei Niveau 0,05 signifikant, N= Anzahl der Kinder, r = Korrelationskoeffizient, BISC 1= Bielefelder Screening – erste Messung 10 Monate vor Einschulung, BISC 2= Bielefelder Screening – zweite Messung 4 Monate vor Einschulung

1.4.2 Prognostische Validität

Die *prognostische Validität* wird anhand von Produkt-Moment-Korrelationen nach Pearson und die Signifikanz durch Weber überprüft (vgl. Weber in Clauß und Ebner, 1968 aus Clauß, Finze & Partzsch 2004, S. 64f.). Als Außenkriterien der hier vorgestellten CBM Phonolino wurden ein standardisierter Leseverständnistest (Knuspel-L, vgl. Marx 1998) sowie der Deutsche Rechtschreibtest für 1. Klassen (DRT 1, vgl. Müller 2003a) herangezogen.

Für die Berechnung der Korrelationen mit dem Knuspel-L wurde hier der T-Wert der Klassenstufennorm[30] zugrunde gelegt. Die Gesamtauswertung bezieht sich in der vorliegenden Studie auf den „Knuspel-Score I“, der den Vorläuferfähigkeiten des Hörverstehens, Dekodierens und Rekodierens entspricht, da aufgrund des Alters der Kinder der Untertest zum Leseverstehen nicht durchgeführt werden konnte. Der Messwert des DRT1 wird in „Wortfehlern“ gemessen, sodass ein negativer Zusammenhang besteht: Es ist zu vermuten, dass je mehr Punkte die Kinder in den Phonolino-Messungen erreichen, desto weniger Wortfehler machen sie im DRT1. Aus diesem Grund müsste sich eine negative Korrelation ergeben.

Die Gesamtauswertung bezieht sich auf die Rohwerte der Wortfehler im gesamten Test. Die Messungen von Phonolino ergaben in Bezug auf das Kriterium „Lesen“ (vgl. Tab. 66) , gemessen durch den Knuspel-Score 1, mit einer Stichprobe von N = 162 bis N = 175 durchgehend hochsignifikante Korrelationen zwischen r = .284 (Phono 2) und r = .434 (Phono 6). Diese Zusammenhänge sind als mittlere bis hohe Korrelationen einzuschätzen.

Die Messungen von Phonolino ergaben in Bezug auf das Kriterium „Rechtschreiben“ (vgl. Tab. 66), gemessen durch den DRT1, mit einer Stichprobe von N = 164 bis N = 182 durchgehend hochsignifikante Korrelationen zwischen r = - .341 (Phono 2) und r = - .564 (Phono 6). Diese Zusammenhänge sind als mittlere bis hohe Korrelationen einzuschätzen.

Zur Überprüfung der prognostischen Validität des Verfahrens durch eine klassifikatorische Analyse und damit verbundenen Aussagen über Risiko- und Nicht-Risikokinder werden die Ergebnisse der CBM Phonolino als „Prädiktor“ und die Ergebnisse der standardisierten Testverfahren DRT 1 und Knuspel-L als „Kriterium“ bezeichnet. Anhand der Vierfeldertafel werden die vorher dichotomisierten Daten analysiert und ausgewertet, um die Güteindizes der klassifikatorischen Validität zu bestimmen und zu interpretieren. Die Kennwerte werden bzgl. ihrer Irrtumswahrscheinlichkeit auf ihre Signifikanz hin überprüft.

30 Für jede Subtestleistung liegen klassenstufenspezifische und klassenübergreifende Normen vor. Die Klassenstufennorm beschreibt die klassenspezifische Normierung.

Tab. 66: Produkte-Moment-Korrelationen der CBM „Phonolino“, Phono 2 bis Phono 9 (Prädiktor) und den Außenkriterien Lesen (Knuspel-L) und Rechtschreiben (DRT1)

	Lesen (Knuspel-L)	Rechtschreiben (DRT1)
Phono 2	r = .284**	r = - .341**
	N = 173	N = 174
Phono 3	r = .415**	r = - .498**
	N = 173	N = 174
Phono 4	r =.395**	r = - .495**
	N = 175	N = 177
Phono 5	r =.363**	r = - .448**
	N = 180	N = 182
Phono 6	r =.434**	r = - .564**
	N = 162	N = 164
Phono 7	r = .378**	r = - .461**
	N = 164	N = 166
Phono 8	r = .402**	r = - .507**
	N = 175	N = 177
Phono 9	r = .402**	r = - .481**
	N = 171	N = 173

Anmerkung: **= Korrelation ist bei Niveau 0,01 signifikant, * Korrelation ist bei Niveau 0,05 signifikant, N= Anzahl der Kinder, r = Korrelationskoeffizient

Zur Dichotomisierung der Rohwerte von Phonolino wurden diese je Messzeitpunkt in Perzentile mit $184 \leq N \leq 196$ eingeteilt (vgl. Tab. 67). Bei der CBM Phonolino wird das unterste Quartil als Grenzwert für die Einteilung in Risiko- und Nicht-Risikokinder anhand der insgesamt bearbeiteten Rohwerte der jeweiligen CBMs ermittelt und ist damit für die einzelnen Messzeitpunkte unterschiedlich. Die Grenzwerte des untersten Quartils zu jeder der Messungen ist der Tab. 67 in der Spalte zum Perzentil von 25 % zu entnehmen. Der jeweilige Grenzwert für die Vierfeldertabellen ist dann größer oder gleich der 25 %-Grenze, z.B. ist für Phono 2 der Grenzwert ≤ 9 Rohwertpunkten.

Tab. 67: Einteilung der Rohwerte von Phonolino in Quartil zu den Messzeitpunkten Phono 2 bis Phono 9

		Phono 2 Gesamt	Phono 3 Gesamt	Phono 4 Gesamt	Phono 5 Gesamt	Phono 6 Gesamt	Phono 7 Gesamt	Phono 8 Gesamt	Phono 9 Gesamt
N	Gültig	197	196	193	203	184	184	197	193
	Fehlend	17	18	21	11	30	30	17	21
Perzentile	25	9,00	10,00	11,00	12,00	13,00	13,00	14,00	15,00
	50	11,00	13,50	15,00	15,00	17,00	17,00	19,00	18,00
	75	15,00	17,75	18,00	19,00	20,00	21,00	22,00	22,00

Anmerkung: blau hervorgehoben: Cut-Off-Wert bei 25%

Im Anschluss wurden Kreuztabellen zu allen acht Messzeitpunkten zu den Kriterien „Lesen“ (gemessen über den Knuspel-L) und „Rechtschreiben“ (gemessen über den DRT1) anhand der Einteilung von Risiko- und Nicht-Risikokindern erstellt.

In der folgenden Darstellung der Ergebnisse werden bei der klassifikatorischen Analyse die Güteindizes der Sensitivität, Prädiktortrefferquote (positiver prädikativer Wert) und der RATZ-Wert nach den Empfehlungen von Marx und Lenhard (2010, S. 9ff.) und Walter (2010a, S. 34) in den Vordergrund gestellt, da diese Kennwerte bei einer späteren Beurteilung der prognostischen Validität der CBM eine bedeutende Rolle spielen.

Für die Vorhersage der Lesefähigkeit am Ende der ersten Klasse sind die berechneten besonders relevanten Kennwerte der Güteindizes in Tab. 68 für Messungen mit der CBM Phonolino als Prädiktor und Knuspel-L als Kriterium zusammengefasst (die zugehörigen Auswertungen befinden sich im Anhang 13).

Die Lesefähigkeit wird anhand des T-Werts der Klassennorm aufgeführt. Die angegebenen Größen in der Tabelle wurden auf ganze Prozent gerundet. Für das Kriterium Knuspel-L wird der T-Wert des Knuspel-Score I (Vorläuferfertigkeiten für das verstehende Lesen) als Gesamttest verwendet. Das unterste Quartil wird anhand der üblichen Verteilung bei T-Werten mit $T \leq 37$ bzw. $T > 37$ angesetzt (vergleichend befindet sich die Berechnungen für den Cut-Off $T = 40$ und $T = 43$ im Anhang 14). Für Phonolino wird das unterste Quartil durch Berechnung bestimmt (vgl. Tab. 67, S. 367). Die Festlegung dient der Dichotomisierung der Datensätze. 40 bis 52 % der Risikokinder können durch die CBM Phonolino richtig klassifiziert werden (SE[31]). Dies ist als moderate Vorhersage zu bewerten.

31 SE = Sensitivität

Die höchste Sensitivität lässt sich bei Phono 9 mit 52 % erzielen, die geringste bei Phono 4, 5, 7 und 8 mit je 40 %, was als moderat zu interpretieren ist (vgl. Tab. 51, S. 336).

Tab. 68: Güteindizes (gerundet auf volle Prozent) als klassifikatorische Analyse der CBM Phonolino zur Vorhersage von Risikokindern (PR ≤ 25) und Nicht-Risikokindern (PR > 25) im Bereich der Lesefähigkeit anhand des Kriteriums Knuspel-L (T-Wert der Klassennorm, Cut-Off: T-Wert = 37)

$Phono_k$	N	a	b	c	d	SE	SP	PPW	GQ	ZQ	LI+	LI-	Chi^2 5%	Chi^2 1%	p	RATZ
Phono 2	173	19	39	22	93	46	71	33	65	58	1.6	0.8	4.0	3.2	0.00	19
Phono 3	173	19	30	25	99	43	77	39	68	60	1.9	0.7	6.4	5.5	0.00	21
Phono 4	175	17	37	26	95	40	72	32	64	59	1.4	0.8	2.0	1.5	0.00	13
Phono 5	180	18	36	26	100	40	74	33	66	60	1.5	0.8	3.3	2.6	0.00	16
Phono 6	162	16	33	22	91	42	73	33	66	60	1.6	0.8	3.3	2.6	0.00	17
Phono 7	164	16	29	24	95	40	77	36	68	61	1.7	0.8	4.2	3.4	0.00	17
Phono 8	175	17	32	25	101	40	76	35	67	61	1.7	0.8	4.3	3.5	0.00	17
Phono 9	171	23	29	21	98	52	77	44	71	60	2.2	0.6	13.4	12.0	0.00	31

Anmerkung: N= Anzahl der Probanden, a = Anzahl richtig-positiver Screening-Befunde, b = Anzahl falsch-positiver Befunde, c = Anzahl falsch-negativer Befunde, d = Anzahl richtig-negativer Befunde, SE= Sensitivität, SP= Spezifität, PPW=Positive Korrektheit, GQ=Gesamttrefferquote, ZQ=Zufallstrefferquote, LI+/LI- = Likelihood Ratio, Chi² 5%= Chi² Test 5% krit.=3.84, Chi² 1%= Chi² Test 1% krit.=6.64, p = Fisher's p, RATZ= Relativer Steigerungsbetrag gegenüber dem Zufall

Die meisten Nicht-Risikokinder (SP) ließen sich durch Phono 8 (76 %), die wenigsten durch Phono 6 (73 %) klassifizieren. Diese Vorhersage lässt sich als stark interpretieren, muss aber aus bereits dargelegten Gründen mit Vorsicht betrachtet werden (vgl. Kap. VI.5.2.3.4, S. 335, nach Marx und Lenhard (2010, S. 10) ist die Spezifizität ein durch den Cut-Off leicht manipulierbarer Wert).

Insgesamt konnten 65 bis 71% der Kinder richtig klassifiziert werden (GQ), damit lässt sich die Gesamttrefferquote als stark beschreiben. Die Werte für die Likelihood Ratio, als Wert für die Testeffizienz, liegen zwischen $1.4 \leq LR^+ \leq 2.2$ und zwischen $0.6 \leq LR^- \leq 0.8$. Die Werte sind für LR^+ als schwach bis mäßig und für LR^- als schlecht zu beurteilen. Das negative Likelihood Ratio sollte möglichst klein sein, das symbolisiert, mit welcher Sicherheit durch einen negativen Befund die Lese-Rechtschreibschwierigkeiten ausgeschlossen werden können (vgl. Tröster 2009, S. 106).

Die RATZ-Indizes liegen zwischen 13 % und 31 % und sind damit für die Vorhersage von Lesefähigkeit unbrauchbar und nicht akzeptabel (vgl. Tab. 53, S. 337).

Wie in der Übersichtstabellen (vgl. Tab. 68) zu sehen ist, liegen die Werte der Gesamttrefferquote zwischen 64 % und 71 %, und sind damit nur etwas höher als die Werte der Zufallstrefferquote (zwischen 58 % und 61 %). Dies rechtfertigt nicht zwingend den Einsatz eines diagnostischen Verfahrens zur Vorhersage der Leseleistung durch die Einzelmessungen von Phonolino.

Für die Vorhersage der Rechtschreibfähigkeit am Ende der ersten Klasse sind die berechneten besonders relevanten Kennwerte der Güteindizes in Tab. 69 für Messungen mit der CBM Phonolino als Prädiktor und dem DRT1 als Kriterium zusammengefasst (die zugehörigen Auswertungen befinden sich im Anhang 15). Die Rechtschreibfähigkeit wird anhand des Prozentrangs des Gesamtwerts der DRT1-Testung aufgeführt. Die angegebenen Größen in der Tabelle wurden auf ganze Prozent gerundet. Das unterste Quartil wird mit PR ≤ 10 bzw. PR > 10 angesetzt (vergleichend befindet sich die Berechnungen für den Cut-Off PR = 25 und PR = 15 im Anhang 16, es wurde sich für PR = 10 aufgrund der relativ hohen Prävalenzrate der Studie entschieden, vgl. Tab. 43, S. 305). Für Phonolino wird das unterste Quartil durch Berechnung bestimmt (vgl. Tab. 67, S. 367). Die Festlegung dient der Dichotomisierung der Datensätze.

Tab. 69: Güteindizes (gerundet auf volle Prozent) als klassifikatorische Analyse der CBM Phonolino zur Vorhersage von Risikokindern (PR ≤ 25) und Nicht-Risikokindern (PR > 25) im Bereich der Rechtschreibfähigkeit anhand des Kriteriums DRT (Prozentrang, Cut-Off: PR = 10)

$Phono_k$	N	a	b	c	d	SE	SP	PPW	GQ	ZQ	LI+	LI-	Chi^2 5%	Chi^2 1%	p	RATZ
Phono 2	174	20	39	16	99	56	71	34	68	59	1.9	0.6	9.5	8.3	0.00	32
Phono 3	174	18	31	21	104	46	77	37	70	62	2.0	0.7	8.0	6.9	0.00	25
Phono 4	177	22	32	16	107	58	77	41	73	61	2.5	0.5	17.1	15.5	0.00	39
Phono 5	182	22	31	19	110	54	78	42	73	62	2.4	0.6	15.4	13.9	0.00	34
Phono 6	164	21	28	14	101	60	78	43	74	62	2.8	0.5	19.3	17.5	0.00	43
Phono 7	166	17	28	20	101	46	78	38	71	63	2.1	0.7	8.6	7.4	0.00	26
Phono 8	177	21	29	18	109	54	79	42	73	62	2.6	0.6	16.2	14.6	0.00	36
Phono 9	173	21	31	17	104	55	77	40	72	61	2.4	0.6	14.7	13.2	0.00	36

Anmerkung: N= Anzahl der Probanden, a = Anzahl richtig-positiver Screening-Befunde, b = Anzahl falsch-positiver Befunde, c = Anzahl falsch-negativer Befunde, d = Anzahl richtig-negativer Befunde, SE= Sensitivität, SP= Spezifität, PPW=Positive Korrektheit, GQ=Gesamttrefferquote, ZQ=Zufallstrefferquote, LI+/LI- = Likelihood Ratio, Chi^2 5%= Chi^2 Test 5% krit.=3.84, Chi^2 1%= Chi^2 Test 1% krit.=6.64, p = Fisher's p, RATZ= Relativer Steigerungsbetrag gegenüber dem Zufall

46 bis 60 % der Risikokinder können durch die CBM Phonolino richtig klassifiziert werden (Sensitivität). Dies ist als moderate Vorhersage zu bewerten. Die höchste Sensitivität lässt sich bei Phono 6 mit 60% erzielen, die geringste bei Phono 3 und 7 mit 46 %.

Die meisten Nicht-Risikokinder (SP) ließen sich durch Phono 8 (79 %), die wenigsten durch Phono 2 (72 %) klassifizieren. Diese Vorhersage lässt sich als stark interpretieren, muss aber aus bereits dargelegten Gründen mit Vorsicht betrachtet werden.

Insgesamt konnten 68 bis 74 % der Kinder richtig klassifiziert werden (GQ), damit lässt sich die Gesamttrefferquote als stark beschreiben. Die Werte für die Likelihood Ratio, als Wert für die Testeffizienz, liegen zwischen $1.9 \leq LR^{+} \leq 2.8$ und zwischen $0.5 \leq LR^{-} \leq 0.7$. Die Werte sind für LR^{+} und LR^{-} als schwach bis mäßig zu beurteilen.

Die RATZ-Indizes liegen zwischen 25 % und 39 %. Damit sind nur Phono 4, 5, 6, 8 und 9 für die Vorhersage von Rechtschreibleistungen im ersten Schuljahr gut, jedoch unspezifisch. Phono 2, 3 und 7 sind für die Vorhersage von Rechtschreibleistungen unbrauchbar und nicht akzeptabel (vgl. Tab. 53, S. 337).

Wie in der Übersichtstabelle (vgl. Tab. 69) zu sehen ist, liegen die Werte der Gesamttrefferquote zwischen 68 % und 74 % und sind damit geringfügig höher als die Werte der Zufallstrefferquote (zwischen 59 % und 63 %). Dies rechtfertigt, warum der Einsatz eines diagnostischen Verfahrens zur Vorhersage der Rechtschreibleistung eine Berechtigung findet. Bei Betrachtung der Sensitivität (SE) und der Spezifizität (SP) zeigt sich, dass sich bei beiden Tabellen die Nicht-Risikokinder besser vorhersagen lassen als die Risikokinder, da die Werte der Spezifizität durchgängig höher sind als die der Sensitivität.

Als Vergleich werden hier ebenfalls die in dieser Untersuchung erhobenen Daten zum Bielefelder Screening (BISC) zu zwei Messzeitpunkten (BISC1: 10 Monate vor der Einschulung und BISC2: 4 Monate vor der Einschulung) klassifikatorisch analysiert (vgl. Anhang 17, 18 und 19), um die Ergebnisse von Phonolino besser einordnen zu können. In Tab. 70 findet sich eine Übersicht über die klassifikatorischen Ergebnisse des BISC als Prädiktor und des DRT1 und Knuspel-Ls als Kriterium. Die Cut-Offs wurden ebenfalls bei PR = 10 für den DRT und bei T = 37 für Knuspel-L gesetzt. Für den BISC gilt der im Handbuch vermerkte Cut-Off von vier oder mehr Risikopunkten für die Klassifikation als Risikokind und für weniger als vier Risikopunkten als Nicht-Risikokind.

Wie in Tab. 70 sichtbar, können 30 bis 41 % der Risikokinder durch das Bielefelder Screening richtig klassifiziert werden (SE). Dies ist als moderate Vorhersage zu bewerten, sagt aber auch aus, dass 59 bis 70% der Risikokinder übersehen werden. Die höchste Sensitivität lässt sich beim BISC 2 in Bezug auf die Rechtschreibfähigkeit mit 60 % erzielen, die geringste beim BISC 1 in Bezug auf die Lesefähigkeit mit 46 %.

Die meisten Nicht-Risikokinder (SP) ließen sich durch den BISC 1 in Bezug auf die Rechtschreibfähigkeit (93 %), die wenigsten durch den BISC 1 und BISC 2 (91 %) in Bezug auf das Lesen klassifizieren. Diese Vorhersage lässt sich als stark interpretieren, muss aber aus bereits dargelegten Gründen mit Vorsicht betrachtet werden.

Tab. 70: Güteindizes (gerundet auf volle Prozent) als klassifikatorische Analyse des Bielefelder Screenings zur Vorhersage von Risikokindern und Nicht-Risikokindern im Bereich der Lese- und Rechtschreibefähigkeit anhand des Kriteriums Knuspel-L (T-Wert der Klassennorm, Cut-Off: T-Wert = 37) und DRT (Prozentrang, Cut-Off: PR = 10)

Kriterium	BISC	N	a	b	c	d	SE	SP	PPW	GQ	ZQ	LI+	LI-	Chi^2 5%	Chi^2 1%	p	RATZ
Lesen (Knuspel)	BISC 1	179	13	12	30	124	30	91	52	77	69	3.4	0.8	12.5	10.7	0.00	37
	BISC 2	172	13	12	28	119	31	91	52	77	69	3.5	0.8	12.8	11.0	0.00	37
Recht-schreiben (DRT1)	BISC 1	180	14	10	24	132	37	93	58	81	71	5.2	0.7	23.0	20.5	0.00	47
	BISC 2	173	14	11	20	128	41	92	56	82	72	5.2	0.6	24.4	21.8	0.00	45

Anmerkung: N= Anzahl der Probanden, a = Anzahl richtig-positiver Screening-Befunde, b = Anzahl falsch-positiver Befunde, c = Anzahl falsch-negativer Befunde, d = Anzahl richtig-negativer Befunde, SE= Sensitivität, SP= Spezifität, PPW=Positive Korrektheit, GQ=Gesamttrefferquote, ZQ=Zufallstrefferquote, LI+/LI- = Likelihood Ratio, Chi^2 5%= Chi^2 Test 5% krit.=3.84, Chi^2 1%= Chi^2 Test 1% krit.=6.64, p = Fisher's p, RATZ= Relativer Steigerungsbetrag gegenüber dem Zufall

Insgesamt konnten 77 bis 82 % der Kinder richtig klassifiziert werden (GQ), damit lässt sich die Gesamttrefferquote als stark beschreiben. Die Werte für die Likelihood Ratio, als Wert für die Testeffizienz, liegen zwischen $3.4 \leq LR^{+} \leq 5.2$ und zwischen $0.6 \leq LR^{-} \leq 0.8$. Die Werte sind für LR^{+} als mäßig bis gut und für LR^{-} als schwach zu beurteilen.

Die RATZ-Indizes liegen zwischen 37 % und 47 %. Damit ist das Bielefelder Screening für die Vorhersage von Lese- Rechtschreibleistungen im ersten Schuljahr gut, jedoch unspezifisch (vgl. Tab. 53, S. 337).

Wie in der Übersichtstabelle (vgl. Tab. 70) zu sehen ist, liegen die Werte der Gesamttrefferquote zwischen 77 % und 82 % und sind damit geringfügig höher als die Werte der Zufallstrefferquote (zwischen 69 % und 71 %). Dies rechtfertigt, warum der Einsatz eines diagnostischen Verfahrens zur Vorhersage der Rechtschreibleistung eine Berechtigung findet.

2 Darstellung der Ergebnisse zur Änderungssensibilität

Der Vorteil und auch das Alleinstellungsmerkmal der curriculumbasierten Messung gegenüber klassischen Testverfahren ist das Ziel einer Veränderungsmessung. Nach Petermann (2006, S. 821) ist das Ziel der Veränderungsmessung „[...] die Beschreibung, Erklärung und Prognose von Bedingungs- und Zustandsänderung in Abhängigkeit von Wirkfaktoren."

Zur Überprüfung der Hypothese H5.2 (*Die Mittelwerte des Gesamtwerts von Phonolino bzw. der drei Items steigen signifikant über die acht Messzeitpunkte hinweg mit Ausnahme direkt nebeneinanderliegender Messungen.*) muss zunächst verifiziert werden, ob die Mittelwerte der acht Messungen und ggf. auch der drei Items tatsächlich ansteigen. Danach muss die Voraussetzung überprüft werden, ob nebeneinanderliegende Items tatsächlich auch gleichschwer sind, da ansonsten ein Anstieg der Leistung auch auf einfachere nachfolgende Aufgaben zurückzuführen wäre und die Probanden sich nicht real in ihrer Leistung verbessert haben. Im Anschluss wird dann für einen Vergleich der Mittelwerte der einzelnen Phonolino-Messungen eine einfaktorielle Varianzanalyse mit Messwiederholung berechnet, um zu kontrollieren, ob sich diese Mittelwerte (nicht nebeneinanderliegender Messungen) signifikant unterscheiden. In Tab. 71 sind die Mittelwerte der Gesamtsumme der acht Messungen sowie der einzelnen Items zu sehen. Hierbei wird ebenfalls auf die für die Ermittlung der signifikanten Unterschiede durchgeführte einfaktorielle Varianzanalyse genutzte Stichprobe zurückgegriffen. Die Mittelwerte des Gesamtwerts liegen zwischen 11,96 und 18,95 Rohwertpunkten. Die Mittelwerte steigen kontinuierlich, nur zwischen Phono 3 und Phono 4 ist an der Nachkommastelle ein sinkender Wert von 0,02 Punkten Differenz zu erkennen.

Die Mittelwerte des Items „Pseudowörter nachsprechen" liegen zwischen 2,48 und 4,41 Rohwertpunkten. Die Mittelwerte steigen kontinuierlich, nur zwischen Phono 3 und Phono 4 (0,09 Punkte Differenz) und zwischen Phono 7 und Phono 8 (0,08 Punkte Differenz) ist an der Nachkommastelle ein sinkender Wert zu erkennen. Die Mittelwerte des Items „Silben klopfen" liegen zwischen 5,94 und 9,26 Rohwertpunkten. Die Mittelwerte steigen

kontinuierlich, nur zwischen Phono 8 und Phono 9 ist an der Nachkommastelle ein sinkender Wert von 0,07 Punkten Differenz zu erkennen.

Tab. 71: Deskriptive Statistiken der Mittelwerte des Gesamtwerts von Phonolino, sowie der Items "Pseudowörter nachsprechen", "Silben klopfen" und "Lautisolation" zu den acht Messzeitpunkten

Messzeitpunkt	Mittelwert Gesamtwert N = 91	Mittelwert Pseudowörter N = 94	Mittelwert Silben klopfen N = 91	Mittelwert Lautisolation N = 91
Phono 2	11,96	2,48	5,94	3,46
Phono 3	14,71	3,23	7,22	4,27
Phono 4	14,69	3,14	7,53	4,05
Phono 5	15,54	3,73	7,58	4,34
Phono 6	16,79	3,96	8,17	4,76
Phono 7	17,27	4,07	8,22	5,12
Phono 8	18,59	3,99	9,26	5,43
Phono 9	18,95	4,41	9,19	5,47

Anmerkung: Hervorhebung in grau: Abweichung bzgl. des Anstiegs der Mittelwerte

Die Mittelwerte des Items „Lautisolation“ liegen zwischen 3,46 und 5,47 Rohwertpunkten. Die Mittelwerte steigen kontinuierlich, nur zwischen Phono 3 und Phono 4 ist an der Nachkommastelle ein sinkender Wert von 0,22 Punkten Differenz zu erkennen. In Abb. 58 ist der Mittelwertvergleich des Gesamtwerts und in Abb. 59 der einzelnen Items grafisch nachzuverfolgen.

Die Itemschwierigkeit der drei Items zu den acht Messzeitpunkten nimmt bei den Items ab (vgl. Tab. 55, S. 352). Nur von der Messung Phono 3 zu Phono 4 bei der Lautisolation sowie von Phono 7 zu Phono 8 beim Nachsprechen der Pseudowörter ist eine leichte Zunahme der Schwierigkeit zu verzeichnen. Die Differenzen der Schwierigkeiten nebeneinanderliegender Messungen (vgl. Tab. 72) differiert bei dem Item „Pseudowörter nachsprechen“ zu den acht Messzeitpunkten zwischen 1,23 % und –5,89 %, bei dem Item „Silben klopfen“ zwischen -0,15 % und –5,79 % sowie bei dem Item „Lautisolation“ zwischen 1,45 % und -7,45 %.

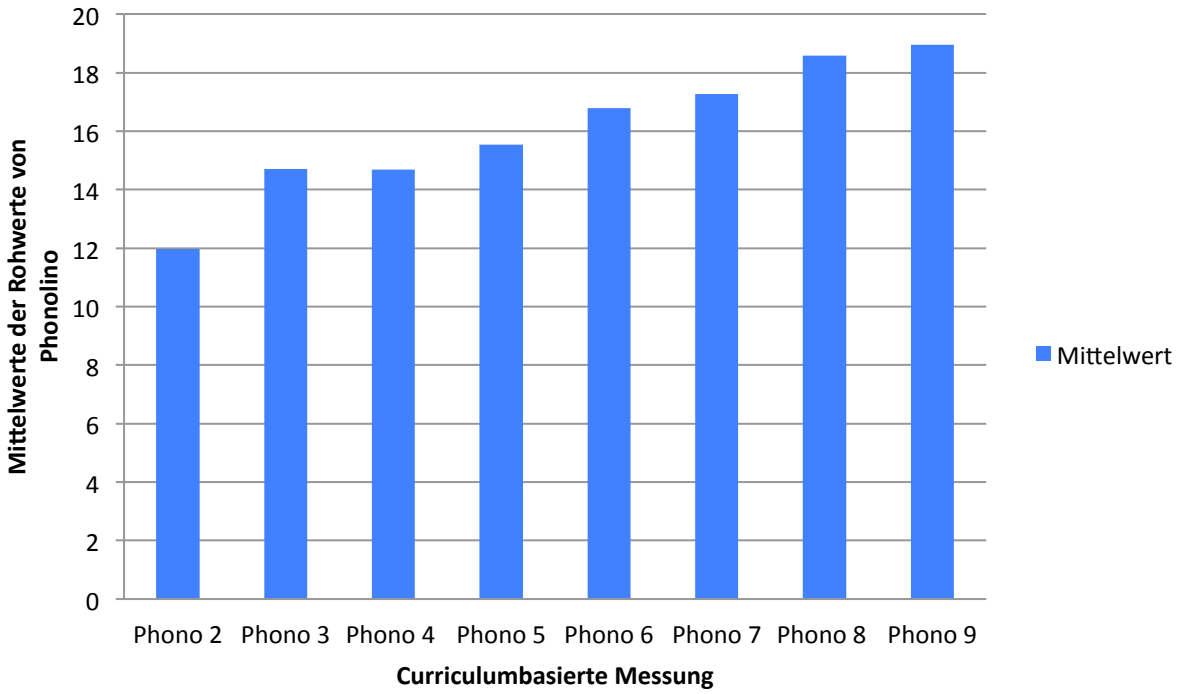


Abb. 58: Veränderung der Mittelwerte der längsschnittlich ermittelten Rohwerte aller Items von Phonolino über die acht curriculumbasierten Messungen im letzten Kindergartenjahr (2008/2009) hinweg, N= 91

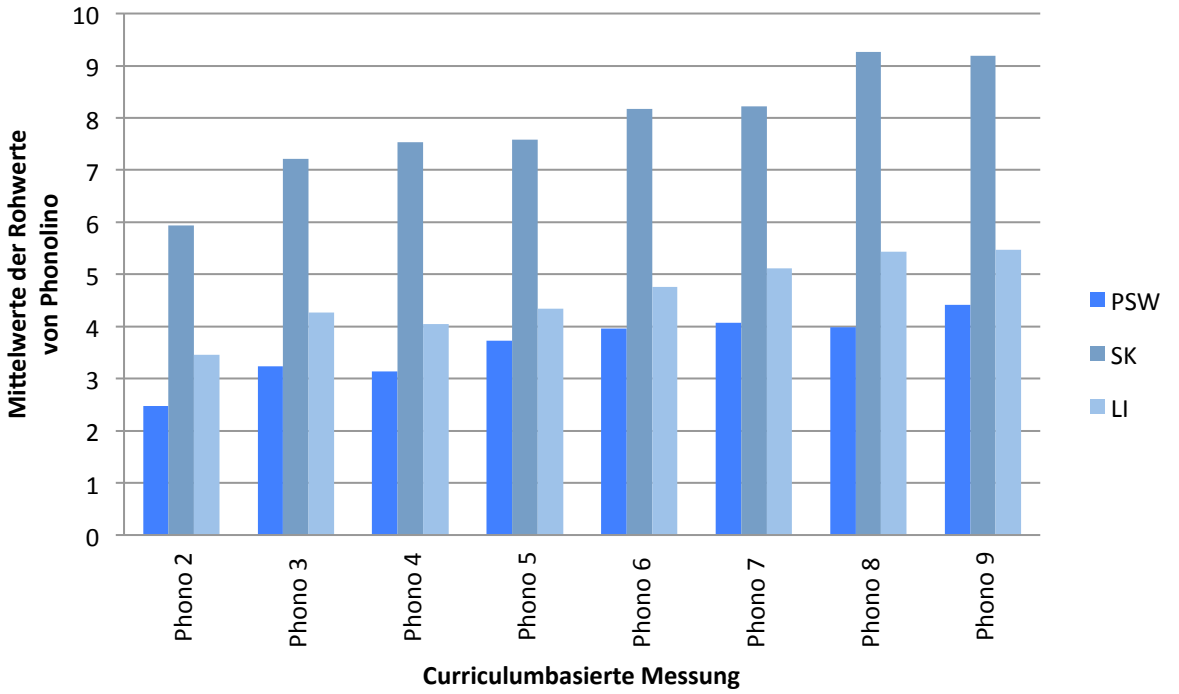


Abb. 59: Veränderung der Mittelwerte der längsschnittlich ermittelten Rohwerte der Items „Pseudowörter nachsprechen“ (PSW, N= 94), „Silben klopfen“ (SK, N= 91) und „Lautisolation“ (LI, N=91) von Phonolino über die acht curriculumbasierten Messungen im letzten Kindergartenjahr (2008/2009) hinweg

Insbesondere der Unterschied von der Messung Phono 2 zu Phono 3 scheint mit gewissen „Anfangsschwierigkeiten“ belegt zu sein. Die erste Messung weist die drei höchsten Differenzwerte in Bezug auf die Schwierigkeiten der Items auf. Dennoch liegen die Itemschwierigkeits-Differenzen zwischen allen nebeneinanderliegenden Messungen unter 10%. Die Voraussetzung zum Vergleich der Mittelwerte kann durch den Augenschein-Vergleich über die deskriptive Analyse als erfüllt angesehen werden.

Tab. 72: Differenzwerte der Itemschwierigkeit nebeneinanderliegender Messungen, k = maximale Punktzahl des jeweiligen Items über alle acht Messezeitpunkte hinweg

	Differenz der Itemschwierigkeit in %		
Messzeitpunkt	Pseudowörter (k = 9)	Silben klopfen (k = 14)	Lautisolation (k = 9)
Phono 2 – Phono 3	- 5,89 %	- 5,79 %	- 7,45 %
Phono 3 – Phono 4	-1,22 %	- 3,85 %	1,45 %
Phono 4 – Phono 5	- 4,34 %	- 0,15 %	- 4,67%
Phono 5 – Phono 6	- 1,22 %	- 5,28 %	- 3,89 %
Phono 6 – Phono 7	- 1,89%	- 1,34 %	- 3,33 %
Phono 7 – Phono 8	1,23 %	- 5,22 %	- 2,45 %
Phono 8 – Phono 9	- 4,67 %	0,58 %	- 1 %

Anmerkung: Die errechnete Differenz beruht auf der Differenz zwischen einer Itemschwierigkeit und der nachfolgenden Itemschwierigkeit, somit hat ein negatives Vorzeichen zu bedeuten, dass die nachfolgende Aufgabe leichter ist, als die vorherige.

Für die einfaktorielle Varianzanalyse mit Messwiederholung wurde die Leistung der phonologischen Informationsverarbeitung (abhängige Variable) in Bezug auf den Faktor Zeit (Messwiederholung ca. einmal pro Monat im letzten Kindergartenjahr) für den Gesamtrohwert der acht Messungen sowie für die einzelnen Items „Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“ betrachtet. Die Normalverteilung (vgl. deskriptive statistische Analyse, Kap. VII.1.2.1, S. 350) sowie die intervallskalierten abhängige Variable sind als Voraussetzung für die Durchführung der Varianzanalyse gegeben. Die Sphärizität für den Gesamtwert (vgl. Tab. 73) der je acht Messungen kann nicht angenommen werden, da der Mauchly-Test (Mauchly-W = .438, p = .000) sich als signifikant erwies. Somit wird im Folgenden die Huynh-Feldt-Korrektur angewendet, da der Greenhouse-Geisser-Epsilon-Wert .794 und damit größer als .75 ist (vgl. Empfehlung von Girden 1992, zit. nach Holling & Gediga 2016, S. 293).

Die Varianzanalyse mit Messwiederholung konnte zeigen, dass die Leistung in der phonologischen Informationsverarbeitung mit der Zeit zusammenhängt und damit ein Haupteffekt vorliegt (Huynh-Feldt Korr.: F = 67,961, p = .000, η^2=.430, N = 91). Das Partielle-Eta-Quadrat zeigt, dass 43 % der Variation der Leistung in der phonologischen Bewusstheit durch die Zeit erklärt werden kann. Die daraus errechnete Effektstärke f nach Cohen beträgt f = .868 und entspricht damit einem starken Effekt. Die Bonferroni-korrigierten paarweisen Vergleiche sind in Anhang 20 dargestellt. Bis auf eine Mittelwertdifferenz (Phono 3 und Phono 5) unterscheiden sich nur nebeneinanderliegende Messungen in der Leistung der phonologischen Informationsverarbeitung nicht (Phono 3 und Phono 4, Phono 4 und Phono 5, Phono 6 und Phono 7 sowie Phono 8 und Phono 9). Alle anderen Mittelwerte unterscheiden sich signifikant ($p < 0.05$).

Tab. 73: Mauchly-Test auf Sphärizität für die Gesamtmessung sowie die Items "Pseudowörter nachsprechen", "Silben klopfen" und "Lautisolation"

	Gesamtwert	Pseudowörter nachsprechen	Silben klopfen	Lautisolation
Mauchly-W	.438	.651	.567	.594
Signifikanz (p)	.000	.065	.004	.010
Epsilon (Greenhouse-Geisser)	.794	Betrachtung nicht notwendig	.850	.869
Korrektur	Huynh-Feldt	Sphärizität gegeben	Huynh-Feldt	Huynh-Feldt

Die Sphärizität für das Item „Pseudowörter nachsprechen“ (vgl. Tab. 73) der je acht Messungen kann angenommen werden, da der Mauchly-Test (Mauchly-W = .651, p = .065) sich als nicht signifikant erwies. Somit wird im Folgenden keine Korrektur angewendet.

Die Varianzanalyse mit Messwiederholung für das Item „Pseudowörter nachsprechen“ konnte zeigen, dass die Leistung in der phonologischen Informationsverarbeitung mit der Zeit zusammenhängt und damit ein Haupteffekt vorliegt (F = 22,111, p = .000, η^2=.192, N = 94). Das Partielle-Eta-Quadrat zeigt, dass 19,2 % der Variation der Leistung in der phonologischen Bewusstheit durch die Zeit erklärt werden kann. Die daraus errechnete Effektstärke f nach Cohen beträgt f = .49 und entspricht damit einem starken Effekt. Die Bonferroni-korrigierten paarweisen Vergleiche sind in Anhang 21 dargestellt. Wie vermutet unterscheiden sich die nebeneinanderliegenden Messungen Phono 3 und Phono 4, Phono 5 und Phono 6, Phono 6 und Phono 7, Phono 7 und Phono 8, Phono 8 und Phono 9 in Bezug auf die Leistung in der phonologischen Informationsverarbeitung nicht.

Außerdem unterscheiden sich die Messungen Phono 3 und Phono 5, Phono 5 und Phono 7, Phono 5 und Phono 8, Phono 6 und Phono 8, Phono 6 und Phono 9 sowie Phono 7 und Phono 9 in Bezug auf die Leistung in der phonologischen Informationsverarbeitung ebenfalls nicht.
Ein signifikanter Unterschied der Mittelwerte ($p < 0.05$) konnte für alle nachfolgenden Messungen von Phono 2 und auch von Phono 3 (bis auf den Unterschied zu Phono 4) und Phono 4 ermittelt werden. Bei den nachfolgenden Messungen zu Phono 5 besteht nur noch für Phono 9 ein signifikanter Unterschied, bei den nachfolgenden Messungen zu Phono 6, Phono 7 und Phono 8 konnte kein signifikanter Unterschied mehr festgestellt werden.

Die Sphärizität für das Item „Silben klopfen“ (vgl. Tab. 73) der je acht Messungen kann nicht angenommen werden, da der Mauchly-Test (Mauchly-W = .567, p = .004) sich als signifikant erwies. Somit wird im Folgenden die Huynh-Feldt-Korrektur angewendet, da der Greenhouse-Geisser-Epsilon-Wert .850 und damit größer als .75 ist (vgl. Empfehlung von Girden 1992, zit. nach Holling & Gediga 2016, S. 293).
Die Varianzanalyse mit Messwiederholung für das Item „Silben klopfen“ konnte zeigen, dass die Leistung in der phonologischen Informationsverarbeitung mit der Zeit zusammenhängt und damit ein Haupteffekt vorliegt (F = 35,450, p = .000, η^2=.278, N = 91). Das Partielle-Eta-Quadrat zeigt, dass 27,8 % der Variation der Leistung in der phonologischen Bewusstheit durch die Zeit erklärt werden kann. Die daraus errechnete Effektstärke f nach Cohen beträgt f = .62 und entspricht damit einem starken Effekt. Die Bonferroni-korrigierten paarweisen Vergleiche sind in Anhang 22 dargestellt. Wie vermutet unterscheiden sich die nebeneinanderliegenden Messungen Phono 3 und Phono 4, Phono 4 und Phono 5, Phono 5 und Phono 6, Phono 6 und Phono 7 sowie Phono 8 und Phono 9 in Bezug auf die Leistung in der phonologischen Informationsverarbeitung nicht.
Außerdem unterscheiden sich die Messungen Phono 3 und Phono 5, Phono 4 und Phono 6, Phono 4 und Phono 7 sowie Phono 5 und Phono 7 in Bezug auf die Leistung in der phonologischen Informationsverarbeitung ebenfalls nicht.

Die Sphärizität für das Item „Lautisolation“ (vgl. Tab. 73) der je acht Messungen kann nicht angenommen werden, da der Mauchly-Test (Mauchly-W = .594, p = .010) sich als signifikant erwies. Somit wird im Folgenden die Huynh-Feldt-Korrektur angewendet, da der Greenhouse-Geisser-Epsilon-Wert .869 und damit größer als .75 ist (vgl. Empfehlung von Girden 1992, zit. nach Holling & Gediga 2016, S. 293).
Die Varianzanalyse mit Messwiederholung für das Item „Lautisolation“ konnte zeigen, dass die Leistung in der phonologischen Informationsverarbeitung mit der Zeit zusammenhängt und damit ein Haupteffekt vorliegt (F = 25,128, p = .000, η^2=.213, N = 91). Das Partielle-

Eta- Quadrat zeigt, dass 21,3 % der Variation der Leistung in der phonologischen Bewusstheit durch die Zeit erklärt werden kann. Die daraus errechnete Effektstärke f nach Cohen beträgt f = .52 und entspricht damit einem starken Effekt. Die Bonferroni-korrigierten paarweisen Vergleiche sind in Anhang 23 dargestellt. Wie vermutet unterscheiden sich die nebeneinanderliegenden Messungen Phono 3 und Phono 4, Phono 4 und Phono 5, Phono 5 und Phono 6, Phono 6 und Phono 7, Phono 7 und Phono 8 sowie Phono 8 und Phono 9 in Bezug auf die Leistung in der phonologischen Informationsverarbeitung nicht.
Außerdem unterscheiden sich die Messungen Phono 3 und Phono 5, Phono 3 und Phono 6, sowie Phono 7 und Phono 9 in Bezug auf die Leistung in der phonologischen Informationsverarbeitung ebenfalls nicht.

Zur Überprüfung der Hypothese H5.2 (*Die Leistung der Probanden steigt durchschnittlich um mindestens einen Punkt zu jeder zweiten Messung.)* wurde zwischen jeder zweiten aufeinanderfolgenden Messung die Mittelwertdifferenz berechnet und anhand eines T-Tests für abhängige Stichproben auf Signifikanz überprüft (vgl. Tab. 74).

Tab. 74: T-Tests zu jeder zweiten aufeinanderfolgenden Phonolino-Messung

	M	SD	Min.	Max.	t	Sig.
Phono 2 – Phono 4 (N = 176)	-2,523	3,978	-3,115	-1,931	-8,412	.000
Phono 3 – Phono 5 (N = 186)	-1,210	3,382	-1,764	-0,655	-4,305	.000
Phono 4 – Phono 6 (N = 165)	-2,352	3,676	-2,917	-1,787	-8,218	.000
Phono 5 – Phono 7 (N = 174)	-1,977	3,707	-2,532	-1,422	-7,034	.000
Phono 6 – Phono 8 (N = 168)	-1,577	3,369	-2,090	-1,064	-6,069	.000
Phono 7 – Phono 9 (N = 164)	-1,543	3,452	-2,075	-1,010	-5,723	.000

Anmerkung: M = Mittelwertdifferenz zwischen den beiden Mittelwerten der Messungen (die errechnete Mittelwertdifferenz beruht auf der Differenz zwischen dem ersten und einem nachfolgenden Mittelwert, somit hat ein negatives Vorzeichen einen positiven Anstieg der Leistung zu verzeichnen), SD=Standardabweichung, Min.= Minimum, Max.=Maximum, t = Teststatistik des t-Wert-Vergleichs, Sig. = Signifikanz (2-seitig)

Der T-Test darf verwendet werden, da die Voraussetzung einer normalverteilten Stichprobe (vgl. Kap. VI.5.2.1.1, S. 319) sowie einer intervallskalierten Variable gewährleistet ist. Aufgrund der gleichen Stichprobe handelt es sich um einen T-Test für abhängige Stichproben (vgl. Anhang 24). Die Mittelwertdifferenzen weisen ein negatives Vorzeichen auf, da der zeitlich spätere Mittelwert vom zeitlich früheren Mittelwert subtrahiert wird (z. B. $M_{Phono3} - M_{Phono5}$). Für N = 164 bis N = 176 Probanden betragen die Mittelwertdifferenzen zwischen -2,523 und – 1,210, das bedeutet (gerundet) ein Anstieg zwischen 1 und 3 Punkten zwischen den jeweils zweiten aufeinanderfolgenden Messungen. Alle Mittelwertunterschiede sind statistisch signifikant.

Zur Überprüfung der Hypothese H5.3 (*Je größer der Abstand zwischen den Messzeitpunkten ist, umso kleiner sind die Korrelationen zwischen dem Gesamtwert von Phonolino bzw. zwischen den Items zum jeweiligen Messzeitpunkt.*) werden die Korrelationen zwischen den Phonolino-Messungen verglichen. Die Korrelationen zwischen den acht Phonolino-Messungen wurden bereits in Bezug auf die Paralleltestreliabilität verwendet (vgl. Tab. 59, S. 356). Nach Klauer (2011, S .218ff.) müssten, wenn eine Veränderungsmessung vorliegt, die Korrelationen der Messungen, je weiter der Messzeitpunkt entfernt ist, abnehmen. Dies würde zeigen, dass die Leistungen der Probanden sich über die Zeit hinweg verändern und Phonolino eine Veränderungsmessung in Bezug auf die phonologische Bewusstheit darstellt. In den unten stehenden Abbildungen (vgl. Abb. 60, Abb. 61, Abb. 62, Abb. 63, Abb. 64, Abb. 65) wurden die Korrelationskoeffizienten zu den nachfolgenden Phonolino-Messungen jeweils in einer Kurve abgebildet. Die orangefarbene Gerade beschreibt den Trend aller Messungen. Insgesamt ist anhand der Darstellung sichtbar, dass sich für Phono 2 bis Phono 7 grundsätzlich ein fallender Trend der Korrelationen beschreiben lässt (sichtbar durch die Gerade mit negativer Steigung). Besonders deutlich wird das bei Phono 2. Weniger sichtbar aufgrund eines Ausreißers und der geringeren Anzahl an nachfolgenden Messungen wird es bspw. bei Phono 6. Bei den Abbildungen (vgl. Abb. 60, Abb. 61, Abb. 62, Abb. 63, Abb. 64, Abb. 65) ist zu beachten, dass die y-Achse je nach gegebenen Werten eine unterschiedliche Skalierung aufweist und die Abbildungen somit grafisch nicht direkt miteinander verglichen werden können.

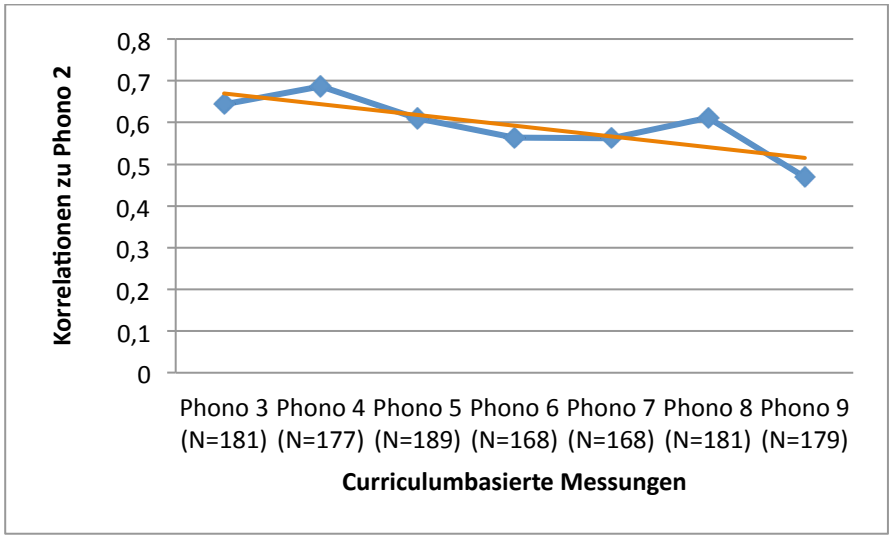


Abb. 60: Veränderung der Korrelationen zwischen Phono 2 und den nachfolgenden Messzeitpunkten
Anmerkung: orangefarbene Gerade: lineare Trendlinie.

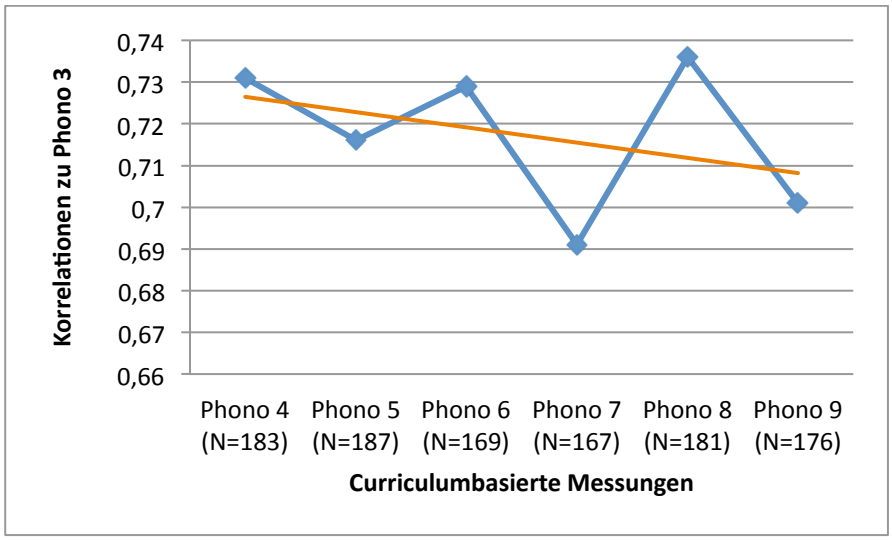


Abb. 61: Veränderung der Korrelationen zwischen Phono 3 und den nachfolgenden Messzeitpunkten
Anmerkung: orangefarbene Gerade: lineare Trendlinie, zur Verdeutlichung des Trends ist hier eine Skalierung auf das Hauptintervall 0,01 dargestellt.

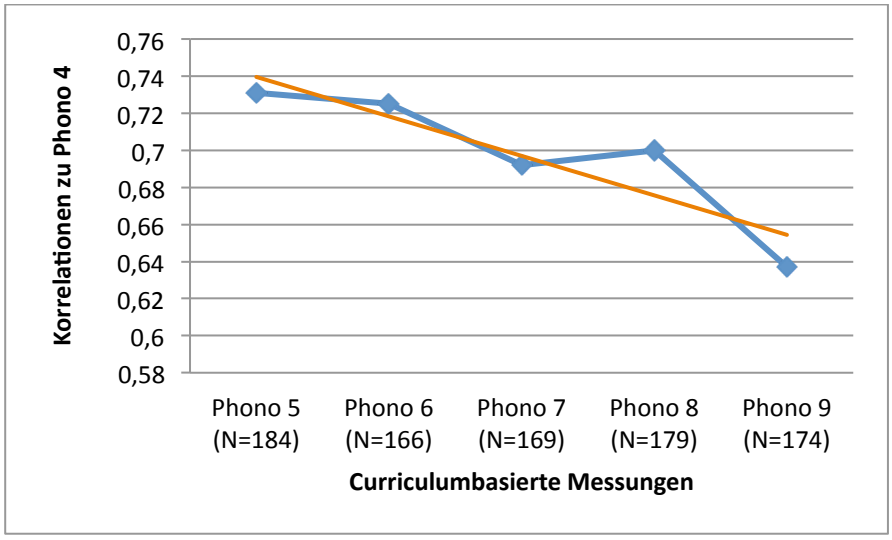


Abb. 62: Veränderung der Korrelationen zwischen Phono 4 und den nachfolgenden Messzeitpunkten
Anmerkung: orangefarbene Gerade: lineare Trendlinie, zur Verdeutlichung des Trends ist hier eine Skalierung auf das Hauptintervall 0,01 dargestellt.

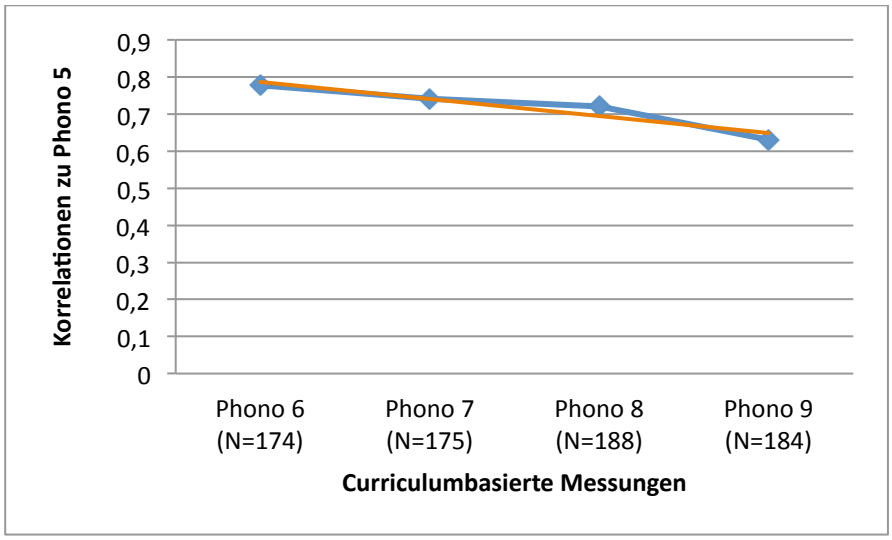


Abb. 63: Veränderung der Korrelationen zwischen Phono 5 und den nachfolgenden Messzeitpunkten
Anmerkung: orangefarbene Gerade: lineare Trendlinie.

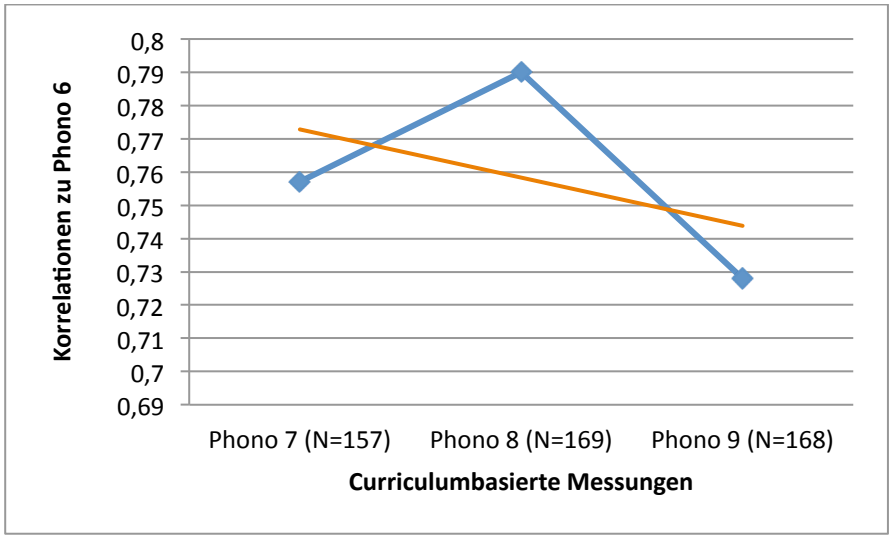


Abb. 64: Veränderung der Korrelationen zwischen Phono 6 und den nachfolgenden Messzeitpunkten
Anmerkung: orangefarbene Gerade: lineare Trendlinie, zur Verdeutlichung des Trends ist hier eine Skalierung auf das Hauptintervall 0,01 dargestellt.

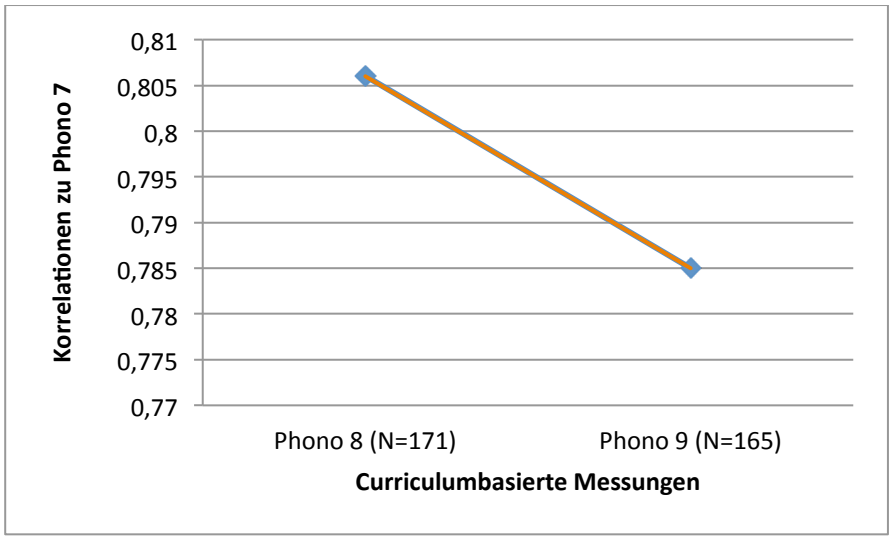


Abb. 65: Veränderung der Korrelationen zwischen Phono 7 und den nachfolgenden Messzeitpunkten
Anmerkung: orangefarbene Gerade: lineare Trendlinie, zur Verdeutlichung des Trends ist hier eine Skalierung auf das Hauptintervall 0,01 dargestellt.

3 Darstellung der Ergebnisse zur Erweiterung des Prognosemodells

Zur möglichen Erweiterung des Prognosemodells für die Vorhersage durch die phonologische Informationsverarbeitung (Modell 1) wird hier anhand der logistischen Regression das Modell 1 mit dem Prädiktor der nonverbalen kognitiven Leistungsfähigkeit der Kinder am Anfang des letzten Kindergartenjahrs (Modell 2) und mit der logistischen Regression überprüft. Die Voraussetzung einer Stichprobe mit mehr als 25 Probanden ist erfüllt, jedoch sind die geforderten „unabhängigen Variablen“ (Werte von Phonolino und

CFT) voneinander abhängig. Dies ist für die Interpretation bei der Wahl der Methodik zu beachten. Zunächst wurden die Kennwerte der logistischen Regression für Modell 1 (unabhängige Variable ist je eine der Phonolino-Messungen) und für Modell 2 (unabhängige Variablen sind je eine der Phonolino-Messungen sowie die kognitiven Leistungen im Kindergartenalter, gemessen mit dem CFT) berechnet. Die Korrelationen der Werte des CFT mit den Werten von Phonolino liegen zwischen .331 und .447 und beeinflussen damit vermutlich nur marginal das Ergebnis der logistischen Regression (Problematik der Multikollinearität - vgl. Kap. VI.5.4, Korrelationen - vgl. Anhang 25).

In Tab. 75 und Tab. 76 sind die logistischen Regressionsgleichungen mit den wichtigsten Parametern aufgelistet. Zur Überprüfung der Parameter wurde zum einen das Wald-Kriterium angewendet, das die berechneten Regressionskoeffizienten auf Signifikanz überprüft, zum anderen wurden die Odds-Ratio (Exp(B)) ebenfalls auf Signifikanz überprüft (Konfidenzintervall von Exp(B) schließt die 1 nicht mit ein). Die negativen Vorzeichen der Regressionskoeffizienten bei Phonolino und CFT kommen dadurch zustande, dass in der abhängigen Variable auf ein Risiko auf Misserfolg im Lesen/Rechtschreiben geschlossen wird.

Tab. 75 zeigt die Regressionskoeffizienten für die Prognose eines Risikos im Lesen (gemessen über den Knuspel-L). Nicht-signifikante Regressionskoeffizienten sind in der Tabelle daran zu erkennen, dass sie nicht mit Sternchen markiert sind. Es zeigt sich, dass im Modell 1 durchgängig die Konstante nicht signifikant ist (mit Ausnahme von Phono 9). Das bedeutet, dass in der Regressionsgleichung Z die Konstante β_0 auf Null gesetzt wird (z.B. für Phono 2 im Modell 1: Z = -0,97• Phono 2, vgl. Schwarz 2012, Folie 62). Im Modell 1 sind alle Phonolino-Werte signifikant bei N = 164 bis N = 180 Kindern. Der Odds-Ratio-Wert lässt auf die Wahrscheinlichkeit schließen. Im Beispiel Phono 2 (Z = -0,138• Phono 2 mit Exp(B)=.908) kann anhand des Odds-Ratio-Wertes ausgesagt werden, dass wenn der Phonolino-Wert um eine Einheit ansteigt, die relative Wahrscheinlichkeit, dass ein Kind ein Risiko im Lesen im ersten Schuljahr aufweist um 9,2 % sinkt, da der Regressionskoeffizient ein negatives Vorzeichen hat (Berechnung - Differenz aus Exp(B) und 1: 0,908 – 1 = - 0,092 → 9,2 %).

Tab. 75: Gleichungen der logistischen Regression für Modell 1 (Phonolino-Messungen) und Modell 2 (Phonolino-Messungen und CFT) zur Prognose der abhängigen Variablen „Schwierigkeiten im Lesen“ (dichotomisiert anhand des Cut-Offs T-Wert = 40)

Modell 1				Modell 2				
N	Z	$Phono_k$	β	N		$Phono_k$	CFT	β
173	Z =	Phono 2 -.097* Exp(B) = .908#	.140	166	Z =	Phono 2 -.055 Exp(B) = .946	-.060** Exp(B) = .918#	1.261*
173	Z =	Phono 3 -.138** Exp(B) = .871#	.931	164	Z =	Phono 3 -.107* Exp(B) = .899#	-.043 Exp(B) = .958	1.622*
175	Z =	Phono 4 -.124** Exp(B) = .883#	.798	165	Z =	Phono 4 -.074 Exp(B) = .929	-.053* Exp(B) = .948#	1.481*
180	Z =	Phono 5 -.120** Exp(B) = .887#	.819	170	Z =	Phono 5 -.079* Exp(B) = .924	-.051* Exp(B) = .950#	1.562*
162	Z =	Phono 6 -.129** Exp(B) = .879#	1.063	151	Z =	Phono 6 -.089* Exp(B) = .915#	-.038 Exp(B) = .963	1.383
164	Z =	Phono 7 -.113** Exp(B) = .893#	.925	153	Z =	Phono 7 -.058 Exp(B) = .943	-.054* Exp(B) = .947#	1.397
175	Z =	Phono 8 -.108** Exp(B) = .898#	.924	165	Z =	Phono 8 -.055 Exp(B) = .946	-.052* Exp(B) = .950#	1.324
171	Z =	Phono 9 -.124** Exp(B) = .883#	1.343*	160	Z =	Phono 9 -.076* Exp(B) = .927#	-.062** Exp(B) = .940#	2.079**

Anmerkung: N = Anzahl der Probanden, Z = abhängige Variable der Linearkombination der Prädiktorvariablen, $Phono_k$ = Regressionskoeffizient der jeweils acht Phonolino-Messungen auf Basis der Rohwerte von Phono 2 bis Phono 9, β = Konstante der Regressionsgleichung, CFT = Regressionskoeffizient des CFT1 auf der Basis der Rohwertsumme aller Untertests des CFT1, Exp(B) = Odds Ratio, * = p<.05 signifikant (Wald-Kriterium), ** = p<.01 (Wald-Kriterium), # = Odds Ratio ist signifikant (Konfidenzintervall schließt 1 **nicht** ein)

Im Modell 2 sind alle Phonolino-Messungen (bis auf Phono 5 und Phono 9) sowie einige Konstanten nicht signifikant. Dies bedeutet, dass das Modell 2 mit den Prädiktoren CFT und Phono bis auf „Phono 5 + CFT“ und „Phono 9 + CFT“ zur Vorhersage von Leseleistungen nicht brauchbar ist. Für „Phono 5 + CFT“ (Z = - 0,079• Phono 3 - 0,051• CFT + 1,562 für N = 170) und „Phono 9 + CFT“ (Z = - 0,076• Phono 3 - 0,062• CFT + 2,079 für N = 160) ergeben sich signifikante Regressionskoeffizienten. Für diese Gleichungen kann über die Odds-Ratio ausgesagt werden, dass wenn der Wert von Phono 5 um eine Einheit ansteigt, die relative Wahrscheinlichkeit, dass ein Kind ein Risiko im Lesen im ersten Schuljahr aufweist um 8,5 % und, wenn der CFT-Wert um eine Einheit ansteigt, um 5 % sinkt. Somit hat in diesem Modell die Phonolino-Messung einen geringfügig größeren Einfluss als der CFT-Wert.

Tab. 76 zeigt die Regressionskoeffizienten für die Prognose eines Risikos im Rechtschreiben (gemessen über den DRT1 im ersten Schuljahr). Die nicht-signifikante Konstante im Modell 1 bei Phono 2 kann durch $\beta = 0$ ersetzt werden (vgl. Schwarz 2012, Folie 62). Im Modell 1 sind alle Phonolino-Regressionskoeffizienten signifikant bei N = 164 bis N = 182 Kindern. Im Modell 2 sind alle Regressionskoeffizienten der Phonolino-Messungen (bis auf Phono 2) sowie die Regressionskoeffizienten des CFT und die der Konstanten signifikant. Dies bedeutet, dass das Modell 1 und auch das Modell 2 mit den Prädiktoren CFT und Phono bis auf „CFT und Phono 2“ zur Vorhersage von Rechtschreibleistungen (mit Ausnahme von „Phono 2 und CFT“) verwendbar sind. Im Modell 2 ist sichtbar, dass die Odds-Ratio-Werte (Exp(B)) für Phonolino durchgängig etwas niedriger sind als für den CFT und damit einen etwas höheren Einfluss auf die Vorhersageleistung haben.

Tab. 76: Gleichungen der logistischen Regression für Modell 1 (Phonolino-Messungen) und Modell 2 (Phonolino-Messungen und CFT) zur Prognose der abhängigen Variablen „Schwierigkeiten im Rechtschreiben“ (dichotomisiert anhand des Cut-Offs Prozentrang PR = 10)

Modell 1				Modell 2				
N	Z	$Phono_k$	β	N		$Phono_k$	CFT	β
174	Z =	Phono 2 -.100* Exp(B) = .905#	-.233	167	Z =	Phono 2 -.034 Exp(B) = .967	-.129** Exp(B) = .879#	2.224**
174	Z =	Phono 3 -.233** Exp(B) = .800#	1.581*	165	Z =	Phono 3 -.175** Exp(B) = .840#	-.106** Exp(B) = .899#	3.578**
177	Z =	Phono 4 -.220** Exp(B) = .802#	1.588**	166	Z =	Phono 4 -.170** Exp(B) = .843#	-.122** Exp(B) = .885#	3.965**
182	Z =	Phono 5 -.208** Exp(B) = .812#	1.679**	171	Z =	Phono 5 -.158** Exp(B) = .854#	-.108** Exp(B) = .897#	3.618**
164	Z =	Phono 6 -.257** Exp(B) = .773#	2.532**	152	Z =	Phono 6 -.222** Exp(B) = .801#	-.084* Exp(B) = .920#	4.065**
166	Z =	Phono 7 -.190** Exp(B) = .827#	1.776**	154	Z =	Phono 7 -.154** Exp(B) = .858#	-.102** Exp(B) = .903#	3.648**
177	Z =	Phono 8 -.197** Exp(B) = .822#	1.985**	166	Z =	Phono 8 -.126** Exp(B) = .882#	-.104** Exp(B) = .901#	3.375**
173	Z =	Phono 9 -.180** Exp(B) = .835#	1.826*	161	Z =	Phono 9 -.129** Exp(B) = .879#	-.117** Exp(B) = .889#	3.817**

Anmerkung: N = Anzahl der Probanden, Z = abhängige Variable der Linearkombination der Prädiktorvariablen, $Phono_k$ = Regressionskoeffizient der jeweils acht Phonolino-Messungen auf Basis der Rohwerte von Phono 2 bis Phono 9, β = Konstante der Regressionsgleichung, CFT = Regressionskoeffizient des CFT1 auf der Basis der Rohwertsumme aller Untertests des CFT1, Exp(B) = Odds Ratio, * = p<.05 signifikant (Wald-Kriterium), ** = p<.01 (Wald-Kriterium), #= Odds Ratio ist signifikant (Konfidenzintervall schließt 1 **nicht** ein)

Die Regressionsmodelle müssen für eine Interpretation Signifikanz aufweisen. Dies wurde im Omnibus-Test für Modellkoeffizienten durch einen Chi-Quadrat-Test überprüft. In Tab. 77 ist für die Prognose der Leseleistung sichtbar, dass sowohl bei Modell 1 als auch bei Modell 2 alle Chi-Quadrat-Tests auf dem .05- oder .01-Niveaus signifikant sind und somit interpretiert werden dürfen. Jedoch konnte bei der Signifikanz-Prüfung der Regressionskoeffizienten festgestellt werden (vgl. Tab. 75), dass diese nicht alle signifikant sind. Die zugehörigen Berechnungen sind in grau gefärbt, da diese Modelle mit Vorsicht interpretiert werden müssen. Es wurden zusätzlich die Effektstärken (f) nach Cohen berechnet, alle dargestellten Effektstärken des Modell 1 konnten als mittlerer Effekt bewertet werden. Für das Modell 2 für „Phono 5 + CFT“ konnte eine höhere Effektstärke von .39 berechnet werden, die dennoch als mittlerer Effekt zu interpretieren ist. Für das Modell 2 für „Phono 9 + CFT“ konnte eine etwas höhere Effektstärke von .43 berechnet werden, die als starker Effekt zu interpretieren ist, sich von .39 jedoch nicht wesentlich unterscheidet.

Tab. 77: Omnibus-Test der Modellkoeffizienten, Überprüfung des Modells auf Signifikanz zur Prognose von Kindern mit Schwierigkeiten im Lesen im ersten Schuljahr sowie Vergleich der beiden aufgestellten Modelle 1 und 2

	Modell 2 (Phono 2 bis Phono 9 und CFT)					Modell 1 (Phono 2 bis Phono 9)					Modellvergleich (Modell 2 – Modell 1)		
$Phono_k$	Chi^2	df	p	R^2	f	Chi^2	df	p	R^2	f	Chi^2	df	p
Phono 2	14.32	2	< .01	.119	.37*	7.19	1	< .01	.059	.25*	7.13	1	< .01
Phono 3	17.25	2	< .01	.143	.41**	14.23	1	< .01	.113	.36*	3.02	1	n.s.
Phono 4	14.07	2	< .01	.118	.37*	11.70	1	< .01	.093	.32*	2.37	1	n.s.
Phono 5	16.12	2	< .01	.130	.39*	12.09	1	< .01	.093	.32*	4.03	1	< .05
Phono 6	12.06	2	< .01	.112	.36*	12.70	1	< .01	.109	.35*	-0.64	1	n.s.
Phono 7	12.71	2	< .01	.115	.36*	11.20	1	< .01	.095	.32*	1.51	1	n.s.
Phono 8	12.86	2	< .01	.109	.35*	10.80	1	< .01	.086	.31*	2.06	1	n.s.
Phono 9	18.23	2	< .01	.153	.43**	13.83	1	< .01	.110	.35*	4.40	1	< .05

Anmerkung: Chi^2=Chi^2-Test des Regressionsmodells (Omnibus-Test), df=Freiheitsgrad des Chi^2-Tests, R^2= Nagelkerkes R-Quadrat, f= Effektstärke nach Cohen (berechnet aus Nagelkerkes R-Quadrat), *=mittlerer Effekt, **starker Effekt, grau hinterlegt: Modelle, die keine signifikanten Regressionskoeffizienten nach dem Wald-Kriterium aufwiesen, n.s. = nicht signifikant auf dem Signifikanzniveau von 0.05

In Tab. 78 ist für die Prognose der Rechtschreibleistung sichtbar, dass sowohl bei Modell 1 als auch bei Modell 2 alle Chi-Quadrat-Tests auf dem .05- oder .01-Niveaus signifikant sind und somit interpretiert werden dürfen. Jedoch konnte bei der Signifikanz-Prüfung der Regressionskoeffizienten festgestellt werden (vgl. Tab. 76), dass der Regressionskoeffizient

von Phono 2 im Modell 2 (Phono 2 und CFT) nicht signifikant ist. Dieses Modell ist in der Tabelle in grau gefärbt, da es mit Vorsicht interpretiert werden muss. Für alle Gleichungen wurden die Effektstärken (f) nach Cohen berechnet. In Modell 1 und 2 konnten alle Effektstärken als starker Effekt bewertet werden. Die Effektstärken des Modells 2 sind durchweg höher als die passenden Effektstärken im Modell 1 (bspw. bei „Phono 3 und CFT" im Modell 2 mit f = .72 hat einen höheren Effekt als „Phono 3" im Modell 1 mit f = .52).

Tab. 78: Omnibus-Test der Modellkoeffizienten, Überprüfung des Modells auf Signifikanz zur Prognose von Kindern mit Schwierigkeiten im Rechtschreiben im ersten Schuljahr sowie Vergleich der beiden aufgestellten Modelle 1 und 2

	Modell 2 (Phono 2 bis Phono 9 und CFT)					Modell 1 (Phono 2 bis Phono 9)					Modellvergleich (Modell 2 – Modell 1)		
$Phono_k$	Chi^2	df	p	R^2	f	Chi^2	df	p	R^2	f	Chi^2	df	p
Phono 2	30.88	2	< .01	.263	.59**	6.17	1	< .05	.055	.24	24.71	1	< .01
Phono 3	41.58	2	< .01	.340	.72**	26.48	1	< .01	.216	.52**	15.10	1	< .01
Phono 4	44.24	2	< .01	.361	.75**	26.26	1	< .01	.213	.52**	17.98	1	< .01
Phono 5	42.48	2	< .01	.337	.71**	26.96	1	< .01	.210	.52**	15.54	1	< .01
Phono 6	40.73	2	< .01	.366	.76**	34.95	1	< .01	.297	.64**	5.78	1	< .05
Phono 7	37.80	2	< .01	.334	.71**	23.91	1	< .01	.205	.51**	13.89	1	< .01
Phono 8	37.67	2	< .01	.313	.67**	27.96	1	< .01	.224	.54**	9.71	1	< .01
Phono 9	38.26	2	< .01	.326	.69**	22.03	1	< .01	.184	.47**	16.23	1	< .01

Anmerkung: Chi^2=Chi^2-Test des Regressionsmodells (Omnibus-Test), df=Freiheitsgrad des Chi^2-Tests, R^2= Nagelkerkes R-Quadrat, f= Effektstärke nach Cohen (berechnet aus Nagelkerkes R-Quadrat), *=mittlerer Effekt, **starker Effekt, grau hinterlegt: Modelle, die keine signifikanten Regressionskoeffizienten nach dem Wald-Kriterium aufwiesen

Aufgrund der aufgestellten Vermutung, dass sich das Prognose-Modell bei Hinzunahme der kognitiven Leistung (Modell 2) verbessert und damit dazu in der Lage ist die Lese- bzw. Rechtschreibleistung im ersten Schuljahr besser vorherzusagen als ein Modell, das ausschließlich die phonologische Informationsverarbeitung (Modell 1) berücksichtigt, werden in Tab. 77 und Tab. 78 im letzten Abschnitt „Modellvergleich" die beiden Modelle miteinander verglichen.

Für die abhängige Variable „Leseleistung" konnten nur für „Phono 2" (Modell 1) und „Phono 2 + CFT" (Modell 2), „Phono 5" (Modell 1) und „Phono 5 + CFT" (Modell 2) sowie „Phono 9" (Modell 1) mit „Phono 9 + CFT" (Modell 2) signifikante Unterschiede der Modelle festgestellt werden. Die Differenz der Chi-Quadrat-Werte beträgt 7.13 (Phono 2),

4.03 (Phono 5) bzw. 4.40 (Phono 9) und ist mit dem Freiheitsgrad von 1 als signifikanter Unterschied (.01-Niveau für Phono 2 und .05-Niveau für Phono 5 und Phono 9) zu interpretieren. Dies bedeutet, dass die Prognoseleistung signifikant durch die Hinzunahme der kognitiven Leistungen der Kinder als Prädiktor zu der phonologischen Informationsverarbeitung verbessert werden konnte. Alle anderen Vergleiche der beiden Modelle stellten sich als nicht signifikant auf dem .05-Niveau heraus. Zu bemerken ist, dass nur für Phono 5 und Phono 9 signifikante Regressionskoeffizienten in der logistischen Regression erzielt wurden.

Für die abhängige Variable „Rechtschreibleistung" zeigt sich, dass das Modell 2 ein deutlich besseres Modell-fit aufzeigt, als das Modell 1, die Chi-Quadrat-Werte sind durchgängig höher. Die Chi-Quadrat-Differenzen konnten als signifikant herausgestellt werden, sodass für die Prognose der Rechtschreibleistung im ersten Schuljahr die Hinzunahme von kognitiven Leistungen im Kindergartenalter zu dem Prädiktor der phonologischen Informationsverarbeitung zu allen Messzeitpunkten signifikant (auf dem .01 oder .05-Niveau) bessere Ergebnisse erzielt werden konnten.

Um die Prognosegüte der Modelle präziser bewerten zu können, wurde eine Klassifikationsanalyse zur Berechnung der Güteindizes durchgeführt (vgl. Tab. 79 und Tab. 80, sowie Anhang 26 und 27). Die Bewertung der Güteindizes ist insbesondere für die Praxis zur Beurteilung eines vorhandenen Risikos relevant. Hierfür wurden nur die Modelle genutzt, die signifikante Regressionskoeffizienten der unabhängigen Variablen aufwiesen.

Für die Vorhersage von Schwächen in der Lesefähigkeit (vgl. Tab. 79) zeigt sich, dass der Wert für die Sensitivität für alle Modelle zwischen 8 % und 78 % liegt. Das heißt, dass bis zu 78 % der Kinder, denen eine Leseschwäche prognostiziert worden ist, auch eine Leseschwäche entwickelt haben. Der Wert von 8 % ist dabei als extrem schwach zu interpretieren. Der Wert für die Spezifität liegt zwischen 72 % und 75 %, d.h. dass bis zu 75 % der Kinder, denen keine Schwierigkeiten im Lesen prognostiziert worden sind, auch keine Probleme entwickelt haben. Die Werte für die Likelihood Ratio, als Wert für die Testeffizienz, liegen zwischen $0.3 \leq LR^{+} \leq 3.0$ und zwischen $0.3 \leq LR^{-} \leq 1.3$. Die Werte sind für LR^{+} und LR^{-} als schwach bis mäßig zu beurteilen. Zur Beurteilung der Prognosequalität wird üblicherweise der RATZ-Index verwendet (vgl. Walter 2014b, S. 28 und Kap. VI.5.2.3.4. S. 335), der in der Tabelle in der rechten Spalte zu finden ist. Der RATZ-Wert kann eine Aussage darüber treffen, inwieweit das Screening dazu in der Lage ist, eine Prognose besser als der Zufall zu treffen. Für die Prognose von Lesefähigkeit konnte kein sehr guter RATZ-Wert (>66 %) (vgl. Jansen et al. 2002 in Tab. 53, S. 337) erzielt werden.

Die RATZ-Werte von Phono 6, Phono 9 und „Phono 9 + CFT“ konnten noch als gut, aber unspezifisch klassifiziert werden. Somit kann in Bezug auf die Güteindizes für die Prognose der Lesefähigkeit nicht gezeigt werden, dass der Einbezug der kognitiven Leistungen der Kinder einen Vorteil für die Prognose im Allgemeinen liefert.

Tab. 79: Güteindizes (gerundet auf volle Prozent) als klassifikatorische Analyse der Modelle 1 und 2 im Bereich der Lesefähigkeit anhand des Kriteriums Knuspel-L (T-Wert der Klassennorm, Cut-Off: T-Wert = 40)

	N	a	b	c	d	SE	SP	PPW	GQ	ZQ	LI+	LI-	Chi^2 5%	Chi^2 1%	p	RATZ
Phono 2	173	1	47	2	123	33	72	2	72	72	1.2	0.9	0.2	0.1	0.00	8
Phono 3	173	4	46	44	119	8	72	8	58	64	0.3	1.3	7.9	6.9	0.00	-2
Phono 4	175	7	43	2	123	78	74	14	74	69	3.0	0.3	11.2	8.8	0.00	7
Phono 5	180	3	48	6	123	33	72	6	70	70	1.2	0.9	0.11	0.00	0.00	7
Phono 5 + CFT	170	8	40	8	114	50	74	17	72	68	1.9	0.7	4.1	3.0	0.00	30
Phono 6	162	7	37	4	114	63	75	16	74	69	2.6	0.5	7.9	6.1	0.00	50*
Phono 7	164	4	42	4	114	50	73	9	71	70	1.9	0.7	2.0	1.0	0.00	31
Phono 8	175	2	47	4	122	33	72	4	71	70	1.2	0.9	0.1	0.0	0.00	7
Phono 9	171	6	45	5	115	55	72	12	71	68	1.9	0.6	3.4	2.3	0.00	35*
Phono 9 + CFT	160	11	36	9	104	55	74	23	72	65	2.1	0.6	7.2	5.9	0.00	36*

Anmerkung: N= Anzahl der Probanden, a = Anzahl richtig-positiver Screening-Befunde, b = Anzahl falsch-positiver Befunde, c = Anzahl falsch-negativer Befunde, d = Anzahl richtig-negativer Befunde, SE= Sensitivität, SP= Spezifität, PPW=Positive Korrektheit, GQ=Gesamttrefferquote, ZQ=Zufallstrefferquote, LI+/LI- = Likelihood Ratio, Chi² 5%= Chi² Test 5% krit.=3.84, Chi² 1%= Chi² Test 1% krit.=6.64, p = Fisher's p, RATZ= Relativer Steigerungsbetrag gegenüber dem Zufall, **RATZ-Index wird als sehr gut bewertet (> 66%), *RATZ-Index wird als gut, aber unspezfisch bewertet (34% < RATZ ≤ 66%)

Für die Vorhersage von Schwächen in der Rechtschreibfähigkeit (vgl. Tab. 80) zeigt sich, dass der Wert für die Sensitivität für alle Modelle zwischen 36 % und 81 % liegt. Das heißt, dass bis zu 81 % der Kinder, denen eine Rechtschreibschwäche prognostiziert worden ist, auch eine Rechtschreibschwäche entwickelt haben. Der Wert für die Spezifität liegt zwischen 79 % und 87 %, d.h. dass bis zu 87 % der Kinder, denen keine Schwierigkeiten im Rechtschreiben prognostiziert worden sind, auch keine Probleme entwickelt haben. Die Werte für die Likelihood Ratio, als Wert für die Testeffizienz, liegen zwischen $1.8 \leq LR^{+} \leq 6.2$ und zwischen $0.2 \leq LR^{-} \leq 0.8$. Die Werte sind für LR^{+} als schwach bis gut und LR^{-} als schwach bis mäßig zu beurteilen.

Tab. 80: Güteindizes (gerundet auf volle Prozent) als klassifikatorische Analyse der Modelle 1 und 2 im Bereich der Rechtschreibfähigkeit anhand des Kriteriums DRT1 (Prozentrang, Cut-Off: PR = 10)

	N	a	b	c	d	SE	SP	PPW	GQ	ZQ	LI+	LI-	Chi^2 5%	Chi^2 1%	p	RATZ
Phono 2 + CFT	167	12	23	3	129	80	85	34	84	74	5.2	0.2	34.7	30.1	0.00	75**
Phono 3	174	6	33	8	127	43	79	15	76	73	2.1	0.7	3.7	2.5	0.00	26
Phono 3 + CFT	165	15	22	6	122	71	84	41	83	71	4.7	0.3	33.2	30.1	0.00	63*
Phono 4	177	10	28	7	132	59	83	26	80	73	3.4	0.5	15.6	13.2	0.00	48*
Phono 4 + CFT	166	17	19	4	126	81	87	47	86	71	6.2	0.2	49.7	45.8	0.00	76**
Phono 5	182	6	35	9	132	40	79	15	76	73	1.9	0.8	2.9	1.9	0.00	23
Phono 5 + CFT	171	16	22	5	128	76	85	42	84	71	5.2	0.3	40.3	36.9	0.00	70**
Phono 6	164	10	25	7	122	59	83	29	81	73	3.5	0.5	15.9	13.5	0.00	48*
Phono 6 + CFT	152	13	19	9	111	59	85	41	82	71	4.0	0.5	22.4	19.8	0.00	48*
Phono 7	166	5	32	8	121	39	79	14	76	73	1.8	0.8	2.1	1.2	0.00	21
Phono 7 + CFT	154	15	19	6	114	71	86	44	84	70	5.0	0.3	34.4	31.2	0.00	63*
Phono 8	177	7	32	12	126	37	80	18	75	72	1.8	0.8	2.7	1.8	0.00	19
Phono 8 + CFT	166	15	21	9	121	63	85	42	82	70	4.2	0.4	27.5	24.8	0.00	52*
Phono 9	173	4	34	7	128	36	79	11	76	75	1.7	0.8	1.4	0.7	0.00	19
Phono 9 + CFT	161	14	21	5	121	74	85	40	84	72	5.0	0.3	34.2	30.8	0.00	66.4**

Anmerkung: N= Anzahl der Probanden, a = Anzahl richtig-positiver Screening-Befunde, b = Anzahl falsch-positiver Befunde, c = Anzahl falsch-negativer Befunde, d = Anzahl richtig-negativer Befunde, SE= Sensitivität, SP= Spezifität, PPW=Positive Korrektheit, GQ=Gesamttrefferquote, ZQ=Zufallstrefferquote, LI+/LI- = Likelihood Ratio, Chi^2 5%= Chi^2 Test 5% krit.=3.84, Chi^2 1%= Chi^2 Test 1% krit.=6.64, RATZ= Relativer Steigerungsbetrag gegenüber dem Zufall, **RATZ-Index wird als sehr gut bewertet (> 66%), *RATZ-Index wird als gut, aber unspezifisch bewertet (34% < RATZ ≤ 66%)

Für die Prognose der Rechtschreibfähigkeit konnte für vier Modelle ein sehr guter RATZ-Wert (> 66 %, vgl. Jansen et al. 2002 in Tab. 53, S. 337) erzielt werden: „Phono 2 + CFT“ mit 75 %, „Phono 4 + CFT“ mit 76 % , „Phono 5 + CFT“ mit 70 % sowie „Phono 9 + CFT“ mit 66,4 %. Die RATZ-Werte von „Phono 3 + CFT“, Phono 4, Phono 6, „Phono 6 + CFT“,

„Phono 7 + CFT“ und „Phono 8 + CFT“ konnten noch als gut, aber unspezifisch, klassifiziert werden. Somit konnte in Bezug auf die Güteindizes für die Prognose der Rechtschreibfähigkeit gezeigt werden, dass der Einbezug der kognitiven Leistungen der Kinder einen deutlichen Vorteil für die Prognose erbringt.

VIII. Diskussion der Ergebnisse

Nach einer umfangreichen Betrachtung der Ergebnisse der Evaluation der curriculumbasierten Messung Phonolino werden diese nun interpretiert und anhand der im theoretischen Teil (vgl. Kap. I - III) dargestellten wissenschaftlichen Befunde eingeordnet. Dabei wird die Güte von Phonolino beurteilt. Hierbei werden die der vorliegenden Studie zugrunde gelegten Fragestellungen, Annahmen und Hypothesen beantwortet, die zusammenfassend im Überblick in Kap. VIII.4, S. 408 dargestellt werden.

1 Interpretation der Ergebnisse zu den klassischen Testgütekriterien

Im folgenden Abschnitt werden die klassischen Testgütekriterien Objektivität, Reliabilität und Validität der curriculumbasierten Messung Phonolino anhand der vorliegenden Berechnungen und Ergebnisse (vgl. Kap. VIII, S. 393ff.) bewertet.

1.1 Objektivität

Wie die oben beschriebenen Auswertungen und Ergebnisse zeigen, ist die Instruktions-, Durchführungs- und Auswertungsobjektivität der Phonolino-Messungen gegeben und nicht vom Testleiter abhängig. Die Objektivität dieser Messungen wird aufgrund der einfachen Durchführung und Auswertung somit als hoch eingeschätzt. Auch andere Untersuchungen zeigten, dass die Interrater-Reliabilität von CBM als hoch eingeschätzt wird (vgl. bspw. Elliott et al. 2001, S. 41 mit $.87 \leq r \leq .94$ im Bereich der Vorläuferfertigkeiten Lesen im Programm DIBELS). So ist aufgrund der hohen standardisierten Vorgaben die Objektivität erfüllt und die Annahmen A1.1, A1.2 und A1.3 können bestätigt werden.

1.2 Itemanalyse

Die Häufigkeitsverteilungen der einzelnen Items sind weitestgehend normalverteilt, wie es durch das Konstrukt der phonologischen Bewusstheit bei Kindern gleichen Alters (im letzten Kindergartenjahr) zu vermuten war. Damit kann die Annahme A2.1 bestätigt werden.

Die Itemschwierigkeit konnte zeigen, dass die Items von Phonolino im Allgemeinen im mittleren Bereich zwischen 29,11% und 64,29% liegen und damit der Forderung von Bortz und Döring (2006, S. 218) unterliegen, dass die bevorzugten Itemschwierigkeiten im Bereich

von 0,2 und 0,8, also zwischen 20% und 80% liegen. Dies ist hiermit gewährleistet. Die Items liegen insgesamt im mittleren Bereich, das Item „Pseudowörter nachsprechen“ ist etwas schwieriger als die anderen beiden Items zu beurteilen. Dies ist für die Konzeption der Messungen von Vorteil, da durch eine breitere Streuung innerhalb einer Messung besser differenziert werden kann.

Die einzelnen Items selbst sind über die Messezeitpunkte hinweg ähnlich schwierig (zu sehen anhand der innerhalb eines Items nur wenig differierenden Mittelwerte). Es konnte anhand der vorliegenden Werte dennoch gezeigt werden, dass sich die Vorschulkinder über die Zeit hinweg verbessern. Dieser Effekt war zu erwarten, da die Kinder im letzten Kindergartenjahr im Bereich der phonologischen Bewusstheit noch deutliche Fortschritte machen und dieser Fachbereich in den Kindertagesstätten auch aktiv gefördert wird (bzw. werden soll). Die Itemschwierigkeit bringt damit zwei zufriedenstellende Ergebnisse: die Itemschwierigkeit liegt im geforderten mittleren Bereich und die Items werden über die Zeit gesehen für die Probanden immer leichter, vermutlich aufgrund des eintretenden Lerneffekts, der im letzten Kindergartenjahr antizipiert wurde. Die Annahme A2.2 kann damit bestätigt werden.

Die als gut und mittelmäßig zu bewertenden Trennschärfekoeffizienten zeigen, dass Phonolino in allen Items dazu in der Lage ist, eine differenzierte Aussage über die Leistung der Kinder im Bereich der phonologischen Informationsverarbeitung zu treffen und kann damit Kinder mit guten und schlechten Leistungen trennen. Insofern sind die Trennschärfen von Phonolino zufriedenstellend, die Annahme A2.3 ist bestätigt.

Die Faktorladungen der Faktorenanalyse zeigen, dass keines der Items aussortiert werden muss, sondern wie durch die inhaltliche Konzeption von Phonolino vermutet, die Items den drei Faktoren „Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“ zugeordnet werden können. Phonolino lässt sich damit anhand der Items in drei Teilkonstrukte operationalisieren. Hiermit kann die Annahme A2.4 bestätigt werden.

1.3 Reliabilität

Die Überprüfung der Reliabilität von curriculumbasierten Messverfahren wurde in US-amerikanischen Studien (vgl. im Überblick Walter 2008b und Klauer 2006 sowie Kap. III zur CBM) mehrfach erfolgreich kontrolliert, indem die Korrelation der zeitlich nacheinander erhobenen Daten der Einzelmessungen berechnet wurde. Einige Studien konnten auch in

Deutschland sehr gute Ergebnisse für die Paralleltestreliabilität von curriculumbasierten Messverfahren vorlegen (vgl. bspw. Walter 2010a, S. 30f. im Bereich Lesen). Gleichermaßen ist die Paralleltestreliabilität der Phonolino-Messungen fast durchgängig durch hohe Korrelationen geprägt. Strathmann, Klauer und Greisbach (2010, S. 74) erhielten in ihrer Untersuchung nur mittlere Reliabilitäten. Sie vermuten, dass die Leistungen der Kinder im Laufe der Zeit stärker als ursprünglich angenommen variieren. Die variierenden Ergebnisse der Reliabilitätsberechnungen anhand der CBM-Daten von Phonolino zu den Messzeitpunkten Phono 2 bis Phono 9 lassen vermuten, dass die phonologische Entwicklung im Vorschulalter nicht beständig stabil zu sein scheint. Es könnte ein Anzeichen dafür sein, dass sich das Merkmal unstet entwickelt. Nach Diekmann (2010, S. 253) sollte der Zuverlässigkeitskoeffizient möglichst nicht kleiner als r = .80 sein, was in dieser Studie nur teilweise erfüllt ist. Ebenfalls in den USA und auch in den Studien von Walter (2008b, 2009a, 2009b) wurde eine Paralleltestreliabilität über r = 0.75 erzielt (vgl. Strathmann, Klauer & Greisbach 2010, S. 67), die hier in dem Sinne nicht repliziert werden konnte.

Bezüglich der Paralleltestreliabilität liegen alle Korrelationskoeffizienten nebeneinanderliegender Messungen im mittleren bis hohen Bereich ($r \geq .50$). Besonders positiv fallen die Vergleiche zwischen den Messungen kurz vor Schuleintritt auf (Phono 8 und Phono 9). Auch in den vorhergehenden Berechnungen (vgl. Kap. VII.1.2, S. 350 zur Itemanalyse) scheint die Messung Phono 2 Anfangsschwierigkeiten zu haben und weist etwas schwächere Zusammenhänge auf, als die anderen Messungen. Dabei sei zu berücksichtigen, dass die für viele der Probanden und Erzieher erste Messung Phono 2 und dessen gemessene Kennwerte unter Vorbehalt zu interpretieren sind, da davon ausgegangen wird, dass sich die Kinder in der ersten Messung zunächst an die neue Form der Diagnostik gewöhnen müssen und es bei einigen empfindlicheren Kindern dazu kommen kann, dass sie sich während der ersten Messung nicht wohl fühlen. Außerdem ist es möglich, dass den Erzieherinnen die Handhabung von Protokollierung, Durchführungsanweisung und Nutzung der Handpuppe noch nicht leicht von der Hand geht. So könnte die Unerfahrenheit der Erzieherinnen den ersten Testdurchlauf stärker beeinflusst haben. Dennoch liegt der Wert noch im mittleren Bereich und ist signifikant. Somit hat sich die Annahme, dass nebeneinanderliegende Messungen mit einem Monat Abstand als Parallelmessungen gehandhabt werden können und mindestens mittlere Korrelationen aufweisen, bestätigt. Es war zu vermuten, dass keine sehr hohen Korrelationen ($r > .80$) zu verzeichnen sind, da innerhalb eines Monats zumindest eine geringe Entwicklung im Bereich der phonologischen Informationsverarbeitung der Kinder zu erwarten ist und die Messungen nicht genau zeitlich parallel stattgefunden haben. Es wäre nützlich die Messung in einer anderen Untersuchung an zwei verschiedenen Stichproben zeitlich parallel zu testen. Dies lag aufgrund des zu

hohen Aufwands in dieser Untersuchung nicht im Rahmen des Möglichen. Grundsätzlich kann die Annahme A3.1 bestätigt werden.

Die Split-Half-Reliabilitäten liegen im mittleren Bereich. Der Split-Half-Wert beschreibt den Zusammenhang zwischen der ersten Hälfte der Messungen und der zweiten Hälfte der Messungen. Die mittleren Reliabilitätskoeffizienten sind möglicherweise in der Dreidimensionalität der Messungen zu begründen (vgl. Kap. VII.1.2.4, S 353 zur Dimensionalität). Die Messung besteht aus den drei Dimensionen „Pseudowörter nachsprechen“, „Silben klopfen“ und „Lautisolation“, die theoretisch betrachtet nicht vollkommen zusammenhängen. Somit ist zu erwarten, dass einige Kinder in der Dimension „Pseudowörter nachsprechen“ deutlich höhere Leistung zeigen, als beim Item „Lautisolation“. Daher war bei der Split-Half-Reliabilität für diese Messungen ein geringer Reliabilitätskoeffizient zu erwarten. Es konnte dennoch gezeigt werden, dass die Phonolino-Messung auch über die verschiedenen Dimensionen hinweg gut misst. Die Messung ist dazu in der Lage die theoretischen Annahmen über den Zusammenhang der drei Komponenten zu erfassen. Bei einer inhaltlichen Weiterentwicklung der Messung wäre zu überlegen, ein weiteres Item mitaufzunehmen, um anhand dieses Items eine Split-Half-Reliabilität über die Items (je zwei pro Testhälfte) hinweg berechnen zu können oder alternativ die Messung zeitlich innerhalb eines Items zu halbieren. Zudem ist zu berücksichtigen, dass weniger als die Hälfte der Fälle in den Berechnungen verwendet werden konnten und dies das Ergebnis ebenfalls beeinflussen könnte.
Aufgrund der oben begründeten Annahme, dass die Dreidimensionalität der Messungen die Reliabilitäten beeinflusst, sind mittlere Split-Half-Reliabilitäten akzeptabel. Die Annahme A3.2 ist damit bestätigt.

Die Reliabilitätskoeffizienten nach Cronbachs Alpha als Maß der *internen Konsistenz* liegen für die acht curriculumbasierten Messungen zwischen $\alpha = .485$ und $\alpha = .625$ (vgl. Tab. 52). Diese Werte sind verhältnismäßig niedrig, allerdings ist bei einer Messung mit nur drei Items eine niedrige interne Konsistenz zu erwarten, sodass die Werte als entsprechend gut bewertet werden können (vgl. Cratzius 2003, S. 165: „Bei Messverfahren mit nur zwei oder drei Items können auch Werte von mindestens 0.4 akzeptiert werden.“). Eine Skala wird bezüglich auf Cronbachs Alpha immer „schlechter“, je weniger Items sie hat. Phonolino hat nur drei Items pro Messung. Zudem wächst Cronbachs Alpha mit der Anzahl an Probanden. Diese liegen in dieser Berechnung zwischen N = 184 und N = 204. Somit kann die Annahme A3.3 mit einer Einschränkung ($\alpha = .485$ bei Phono 4) bestätigt werden.

1.4 Validität

Im Folgenden werden die Ergebnisse zur Überprüfung der Validität der Phonolino-Messungen interpretiert und eingeordnet. Zudem werden teilweise die Ergebnisse der Untersuchungen zum Bielefelder Screening (vgl. Marx & Weber 2006, S. 251ff.) sowie die in der eigenen Untersuchung angestellten Berechnungen zwischen den Kennwerten von Phonolino sowie den Ergebnissen des Bielefelder Screening in der vorliegenden Stichprobe verglichen.

Die Inhaltsvalidität der Messungen ist ausreichend durch die Beschreibung der Konzeption von Phonolino (vgl. Kap. IV.2.5, S. 265) begründet worden. Somit kann die Annahme A4.1 bestätigt werden.

Die Annahme A4.2 zur Übereinstimmungsvalidität des Messverfahrens ist nur in Teilen bestätigt. Die Überprüfung der Zusammenhänge hat gezeigt, dass Phonolino nur in Bezug auf die zweite Testung des BISC vier Monate vor der Einschulung der Kinder mittlere bis hohe Zusammenhänge aufweisen kann. Alle anderen Zusammenhänge in Bezug auf die Kennwerte der ersten Messung des BISC zehn Monate vor der Einschulung konnten nur mittlere, aber dennoch signifikante Korrelationen aufzeigen. Die anderen Überprüfungen (vgl. Kap. VII.1.2, S. 350 zur Itemanalyse sowie Kap. VII.1.3, S. 355 zur Reliabilitätsanalyse) haben gezeigt, dass Phono 2 einige Schwächen aufzuweisen hat. Es ist wieder zu vermuten, dass auch für die Übereinstimmungsvalidität mit der ersten Testung des Bielefelder Screenings (BISC 1) die Begründung anzuführen ist, dass es sich um die erste Messung handelt und die Erzieher und Kinder sich zunächst auf die Form der Überprüfung einstellen mussten. Die Annahme A4.2 kann damit in Bezug auf die zweite Testung durch das BISC bestätigt werden. Ist der Abstand zur Einschulung größer, sind die Korrelationen auch deutlich geringer.

In Bezug auf die prognostische Validität der Phonolino-Messungen wurden sowohl die Korrelationen zwischen Prädiktor (Phonolino) und Kriterium (Lesen oder Rechtschreiben) überprüft, als auch eine genauere klassifikatorische Analyse durchgeführt. Die Validität angesichts der Konkordanz mit einem Außenkriterium konnte durch die Korrelationen mit dem Leseverständnistest Knuspel-L und dem Rechtschreibtest DRT 1 überprüft werden. Für das Außenkriterium „Lesen“, gemessen durch den Knuspel-L, konnten durchweg signifikante Korrelationen berechnet werden, die mittlere bis hohe Zusammenhänge nachwiesen. Damit konnte die Annahme A4.3 mit leichten Einschränkungen ($r = .28$ für Phono 2) bestätigt werden.

Für das Außenkriterium „Rechtschreiben“, gemessen durch den DRT1, sind durchweg signifikante Korrelationen zu Phonolino im mittleren bis hohen Bereich zu verzeichnen. Die Annahme A4.4 ist damit bestätigt.

Es müsste in einer weiteren Untersuchung gezeigt werden, dass ggf. bei einer Aufsummierung der einzelnen Werte der Phonolino-Messungen bessere Aussagen getroffen werden können, wenn eine deutlich größere Stichprobe untersucht wird. Es ist zu vermuten, dass mehrfache Messungen Messfehler und tagesformabhängige Leistungen nivelliert werden. Hierbei müsste dann auch eine Reliabilitätsanalyse der Summenwerte erfolgen. Es ist zu vermuten, dass die Ergebnisse von mehreren Messungen deutlich bessere Aussagen über die Leistung der Kinder treffen und damit ggf. bessere Vorhersagen für die Leistungen im Schriftspracherwerb geleistet werden können.

Als Vergleichswert wurden die in dieser Untersuchung erhobenen Werte vom BISC zehn (BISC 1) und vier Monate (BISC 2) vor der Einschulung mit den Lese- und Rechtschreibleistungen der Probanden in Beziehung gesetzt (vgl. Tab. 81).

Tab. 81: Produkte-Moment-Korrelationen des BISC 1 und des BISC 2 (Prädiktor) und den Außenkriterien Lesen (Knuspel-L) und Rechtschreiben (DRT1)

	Lesen (Knuspel-Score 1: T-Wert, Klassenstufen-norm)	Rechtschreiben (Wortfehler gesamt - Rohwert)
BISC 1 (10 Monate vor der Einschulung)	r = - .391**	r = .375**
	N = 179	N = 180
BISC 2 (4 Monate vor der Einschulung)	r = - .394**	r = .472**
	N = 172	N = 173

Anmerkung: **= Korrelation ist bei Niveau 0,01 signifikant, * Korrelation ist bei Niveau 0,05 signifikant, N= Anzahl der Kinder, r= Korrelationskoeffizient

Hierbei wurden im BISC die Risikopunkte erhoben, so dass sich in der Korrelation mit dem Lesen ein negativer Zusammenhang ergibt. Die Korrelationen mit dem Knuspel-L (BISC 1: r = -.391, BISC2: r = -.394) sind als mittlere Korrelationen einzuschätzen und sind nur geringfügig niedriger als die Korrelationen der Messungen Phono 8 (r = .402) und Phono 9 (r = .402). Die Korrelationen mit dem DRT1 (BISC 1: r = -.375, BISC2: r = -.472) sind

ebenfalls als mittlere bis hohe Korrelationen einzuschätzen und sind niedriger als die Korrelationen der Messungen Phono 3 (r = .498), Phono 4 (r = .495), Phono 6 (r = .564), Phono 8 (r = .507) sowie Phono 9 (r = .481). Die Phonolino-Messungen können damit für diese Stichprobe deutlich bessere Korrelationen für die Rechtschreibleistungen bzw. teilweise bessere Korrelationen für die Leseleistungen im ersten Schuljahr aufweisen als der BISC in beiden Testungen. Die Ergebnisse zum BISC zeigen, dass das bereits mehrfach evaluierte Verfahren in dieser vorhandenen Stichprobe scheinbar auch zu keiner besseren Vorhersage in der Lage ist. Eine mögliche Interpretation der moderaten Vorhersageleistung ist, dass die Leistungen der Schüler im Lesen und Rechtschreiben stark abhängig sind von dem jeweiligen Anfangsunterricht (vgl. Schabmann et al. 2009, die einen Unterschied zwischen systematisch phonologisch-orientiertem und nicht systematisch phonologisch-orientiertem Unterricht und den jeweiligen Vorhersageleistungen des BISC feststellen konnten, systematisch phonologisch-orientierter Unterricht führte zu einem höheren Anteil an falsch-positiven Kindern), oder dass der Prädiktor der phonologischen Informationsverarbeitung nur einen Teil der Leistung beschreibt und für die Identifikation von Risikokindern weitere Faktoren von Interesse sind. Zudem scheinen die Messungen von Phonolino, die näher an der Einschulung sind, besonders im Lesen bessere Kennwerte zu liefern.

Um die Korrelationswerte interpretieren zu können, ist es sinnvoll mit den Werten eines weiteren standardisierten Verfahrens zu vergleichen. Die Basiskompetenzen für Lese-Rechtschreibleistungen (BAKO 1-4 von Stock, Marx & Schneider 2003) weisen beispielsweise nur gering höhere Korrelationskoeffizienten in Bezug auf die kriterienbezogene Validität auf. Die Korrelationskoeffizienten bewegen sich zwischen r = .42 und r = .68 und wurden ebenfalls durch einen Vergleich mit einem standardisierten Lese-Rechtschreibtest erfasst. Die Autoren von BAKO 1-4 interpretieren diese Korrelationen als hoch, da der BAKO 1-4 keine primäre Schulleistung erfasst, sondern (wie auch bei Phonolino) mit der phonologischen Bewusstheit eine Leistung, die mit dem Lesen und Rechtschreiben im Zusammenhang steht, jedoch nicht die gleichen Fähigkeiten erfasst. Somit können die für Phonolino gefundenen Korrelationen mit dieser Begründung auch mehr Gewicht erlangen. Auch das Verfahren zur formativen Evaluation von Leistungen im ersten Schuljahr (Fähigkeitsindikatoren Primarschule – FIPS) kann zwischen der Skala „Lautbewusstheit“ und der WLLP[32] eine Korrelation von r = .34 in Bezug auf die Lesefähigkeiten und zwischen der „Lautbewusstheit“ und dem DERET 1-2+[33] eine

32 WLLP = Würzburger Leise Leseprobe (Küspert & Schneider 1998)
33 DERET 1-2+ = Deutscher Rechtschreibtest für das erste und zweite Schuljahr (Stock & Schneider 2008a)

Korrelation von r = - .40[34] in Bezug auf die Rechtschreibfähigkeiten erreichen. Diese sind als mittlere Korrelationen zu bewerten.

Marx (1997, S. 101) hebt für korrelationsstatistische Auswertungen außerdem zwei wichtige Aussagen hervor:

- Die Korrelation ist umso höher, je geringer der zeitliche Abstand zwischen Prädiktor- und Kriteriumsmaß ist (vgl. Tramontana, Hooper & Selzer 1988).
- Die Korrelation ist umso höher, je ähnlicher die Fertigkeiten sind, die Prädiktor- und Kriteriumsvariablen messen (vgl. Marx 1992a, c).

Es ist also zum einen anhand dieser Aussagen zu vermuten, dass sich die Korrelationswerte bei fortgeschrittener Klassenstufe noch mehr verringern, was in Untersuchungen wie z. B. zum Bielefelder Screening auch schon gezeigt worden ist. Zum anderen erklärt die zweite Aussage, warum die Werte von Phonolino und dem Lese- beziehungsweise Rechtschreibtest nicht extrem hoch korrelieren, da die abgefragten Fähigkeiten und Fertigkeiten im Vergleich zu beispielsweise zwei Lesetests inhaltlich nicht sehr nah beieinander liegen.

Die aufgeführten korrelativen Zusammenhänge sind zusammengefasst ein Indiz für die Übereinstimmungsvalidität der hier betrachteten curriculumbasierten Messung. Dieses Indiz lässt sich für die konzipierte curriculumbasierte Messung Phonolino besonders auf die mögliche Berechnung von Summenwerten der einzelnen Messungen übertragen, die ggf. eine bessere Vorhersage durch eine größere Datengrundlage zu einem Kind gewährleisten können.

Die prognostischen Aussagen in Bezug auf die klassifikatorische Analyse von Phonolino sind weniger erfolgsversprechend. Die von Marx und Lenhard (2010, S. 10) als besonders relevant hervorgehobenen RATZ-Werte messen größtenteils keine speziellen Vorhersagen. Die Annahmen A4.5 und A4.6 müssen diesbezüglich abgelehnt werden.

Die Einzelmessungen scheinen in Bezug auf die Lese- und Rechtschreibfähigkeiten eine geringe (Leseleistung) oder eine gute, jedoch unspezifische Aussage (Rechtschreibleistung) im ersten Schuljahr tätigen zu können. Das Bielefelder Screening als umfassendes Screeningverfahren scheint ggf. durch die weiteren Komponenten des Tests, die auch gewisse kognitive Prozesse, wie die Gedächtnisspanne oder visuelle Aufmerksamkeit

34 Die negativen Korrelationen entstehen aufgrund dessen, dass im DERET die Rechtschreibfehler erhoben werden.

umfassen, einen höheren jedoch auch keinen sehr guten Aussagewert treffen zu können. In Tab. 82 sind die RATZ-Werte der Phonolino-Messungen und des Bielefelder Screenings im Vergleich dargestellt.

Tab. 82: Vergleich der RATZ-Werte der klassifikatorischen Analyse der Phonolinomessungen und des Bielefelder Screenings (BISC) zu den Kriterien Lesen und Rechtschreiben

	Lesen (Knuspel-L, RATZ-Wert)	Rechtschreiben (DRT 1, RATZ-Wert)
BISC 1 Risikopunkte (10 Monate vor Einschulung)	37 %	47 %
BISC 2 Risikopunkte (4 Monate vor Einschulung)	37 %	45 %
Phono 2 bis Phono 9	17 - 31 %	25 - 43 %

Die Befunde zur klassifikatorischen Validität in Bezug auf das Lesen bestätigen die schwachen Korrelationen der Kriterien mit der CBM Phonolino. Bezüglich des Rechtschreibens zeigte sich, dass sich die RATZ-Werte mit Ausnahme der drei unbrauchbaren Werte (Phono 2, Phono 3 und Phono 7) als moderat einstufen ließen. Die dargestellten Werte liefern ein Indiz für die prognostische Vorhersage von Phonolino in erster Linie in Bezug auf das Rechtschreiben in der Schule.

Der Grund für den Anteil an Kindern, die als „falsch positiv“ klassifiziert wurden, kann sein, dass die Erzieherinnen, die die Messungen durchgeführt haben, besonders motiviert waren, die Vorläuferfertigkeiten der Vorschulkinder zu verbessern. Obwohl sie keine genauen Ergebnisse erhalten haben, ist es möglich, dass sie ein besonderes Augenmerk auf die Fähigkeiten in der phonologischen Bewusstheit gelegt haben und sie aufgrund ihrer beruflichen Vorerfahrungen die Schwierigkeiten dieser Kinder in der Messung richtig interpretiert und gezielt gefördert haben. Die Lehrkräfte der ersten Klasse wurden über die Studie im Kindergarten nur wenig informiert und haben genauso wie die Erzieherinnen in den Kindergärten auch keine Ergebnisse erhalten, so ist hier nicht davon auszugehen, dass sie Kinder mit schlechten Ergebnissen gesondert gefördert haben. Die mögliche Förderung durch die Erzieherinnen kann aber eine Auswirkung auf die klassifikatorischen Testgütekriterien gehabt haben. Ein Verzicht auf Förderung der durch Erzieherinnen

während der CBM nach Augenschein identifizierten Risikokinder wäre jedoch aus moralischen Gründen nicht zu akzeptieren gewesen.

Insgesamt lässt sich mit der hier dargestellten curriculumbasierten Messung eine bessere Vorhersage über die Lese- und Schreibfähigkeiten der Kinder in der Schule treffen, als dies bei der wiederholten Untersuchung des BISC der Fall war (vgl. Marx &Weber 2006, S. 256), in der sich für das Kriterium Rechtschreibung im ersten Schuljahr ein RATZ-Wert von nur 16 %, für die Lesegeschwindigkeit von 33 % und für das Leseverständnis von 25 % einstellte, wobei dies auch an der Zusammensetzung der Stichprobe liegen kann, die insgesamt eine sehr hohe Prävalenzrate für Lese- und Rechtschreibschwierigkeiten aufweist (vgl. Kap. VI.2.1).

Der Gruppentest zur Früherkennung von Lese– und Rechtschreibschwierigkeiten beispielsweise weist dagegen deutlich höhere klassifikatorische Werte (z.B. eine Sensitivität von 63 %) auf als die hier vorgestellten Ergebnisse, wobei kein Hinweis auf den RATZ-Wert zu finden ist. Der nachberechnete RATZ-Wert von Marx und Lenhard (2010, S. 12) beträgt 55 % (vgl. Kap. I.2.2.3 zum PB-LRS) und ist damit deutlich höher als die Ergebnisse des BISC und Phonolino in der vorliegenden Studie.

Zu beachten ist, dass insbesondere bei der klassifikatorischen Validität der Cut-Off-Wert in den Auswertungen eine besondere Rolle einnimmt, die nicht zu unterschätzen ist. Bei der Klassifikation von Risiko- und Nicht-Risikokindern wurde für die Phonolino-Werte jeweils ein bestimmter Cut-Off-Wert angenommen (vgl. Tab. 67, S. 367). Punktuelle Cut-Off-Werte bergen immer das Risiko, dass ein Kind mit 12 Punkten ein Risikokind ist, ein Kind mit zum Beispiel 13 Punkten jedoch nicht, obwohl auch dieses Kind später Schwierigkeiten im Lesen und Schreiben entwickeln könnte.

In den Ergebnissen wird deutlich, dass die phonologische Informationsverarbeitung scheinbar für die Rechtschreibleistung eine bessere Vorhersage treffen kann, als für die Leseleistung. Zusammenfassend konnten für die vorgestellten curriculumbasierten Messungen keine herausragenden Kennwerte in allen Belangen ermittelt, jedoch eine schwache bis moderate klassifikatorische Validität durch einzelne Messungen gewährleistet werden. Insgesamt scheint das Bielefelder Screening in dieser Untersuchung geringfügig und dennoch nicht zufriedenstellende bessere Vorhersagen treffen zu können.

2 Interpretation der Ergebnisse zur Änderungssensibilität

Die Gütekriterien Objektivität, Validität und Reliabilität sind als Grundvoraussetzung eines Messverfahrens anzusehen, das Lernzuwachs bzw. Leistungsveränderungen misst. Objektivität, Reliabilität und teilweise Validität (bis auf die Vorhersageleistung speziell im Lesen) konnten für Phonolino (vgl. Kap. VII.1) nachgewiesen werden. Es folgt eine weitere Betrachtung der Änderungssensibilität der curriculumbasierten Messungen.

Petermann (2006, S. 822) beschreibt folgende Ziele der Veränderungsmessung in Bezug auf die Lernforschung:

- das Abbilden reliabler und valider Lernverläufe,
- die Identifikation und Bewertung von Unterschieden in Lernverläufen,
- die Optimierung von Lehr- und Lernprozessen

Die annähernde Normalverteilung der Einzelwerte von Phonolino (vgl. Anhang 8) lässt sich als Indiz für eine mögliche Änderungssensibilität ansehen, dessen Ergebnisse im nachfolgenden Abschnitt ausgewertet werden.

Die deskriptive Beschreibung der Mittelwerte konnte bis auf einige Differenzen zeigen, dass die Mittelwerte über die acht Messzeitpunkte hinweg ansteigen. Die beschriebenen Differenzen liegen im Zehntel- bis Hundertstel-Bereich, so dass diese, insgesamt betrachtet, zu vernachlässigen sind. Zusammenfassend wird in Tab. 83 anhand der signifikanten Haupteffekte und der starken Effektstärken sichtbar, dass die Leistung in der phonologischen Informationsverarbeitung mit der Zeit zwischen den Messwiederholungen im Zusammenhang steht.

In Bezug auf die paarweisen Vergleiche kann nur für den Gesamtwert von Phonolino für alle bis auf eine nachfolgende Messung festgestellt werden, dass sich die Mittelwertdifferenzen signifikant unterscheiden. Hierbei sind die direkt nachfolgenden Messungen (also z.B. Phono 3 und Phono 4) ausgenommen. Für das Item „Pseudowörter nachsprechen“ waren sechs Mittelwertdifferenzen (mit Ausnahme der nebeneinanderliegenden Messungen), bei dem Item „Silben klopfen“ waren vier und bei dem Item „Lautisolation“ waren drei Items nicht signifikant.

Tab. 83: Test der Innersubjekteffekte zur Überprüfung des Haupteffekts durch eine einfaktorielle Varianzanalyse mit Messwiederholung – Vergleich des Gesamtwerts, sowie der drei Items über die acht Messzeitpunkte Phono 2 bis Phono 9 hinweg

	Gesamtwert	Pseudowörter nachsprechen	Silben klopfen	Lautisolation
N	91	94	91	91
F	67,961	22,111	35,450	25,128
p	.000	.000	.000	.000
η^2	.430	.192	.278	.213
f	.868	.49	.62	.52

Anmerkung: N = Anzahl der Probanden, F= Teststatistikwert für F in der Varianzanalyse, p = Signifikanzwert η_p^2= Partielles Eta-Quadrat, f = Effektstärke nach Cohen

Die Bonferroni-korrigierten paarweisen Vergleiche der einzelnen Items konnten aufzeigen, dass viele nachfolgende Messungen nicht mehr signifikant unterschiedlich waren. Bei dem Item „Pseudowörter nachsprechen“ zeigte sich, dass in der zweiten Hälfte des letzten Kindergartenjahrs ab der Messung Phono 5 sich keine signifikanten Unterschiede in den Leistungen der Kinder abzeichnen ließen. Es ist zu vermuten, dass bei den Messungen Deckeneffekte aufgetreten sind, sodass die Kinder, die vorher das Item eher schlecht bewältigt haben, die Kinder, die bereits zu Anfang gute Leistungen hatten, teils eingeholt haben.

Insgesamt kann die Hypothese H5.1 für den Gesamtwert von Phonolino mit der Einschränkung einer Mittelwertdifferenz bei Phono 3 und Phono 5 angenommen werden, da sowohl die Mittelwerte bis auf wenige Differenzen im Zehntel-/Hunderstelbereich ansteigen, als auch die Mittelwertdifferenzen aller nicht direkt nachfolgender Messungen sich signifikant unterscheiden. Es kann damit angenommen werden, dass die Zeit ein bedeutender Faktor für die Leistungsveränderung der Kinder im Bereich der phonologischen Informationsverarbeitung darstellt. Für die einzelnen Items kann die Hypothese nur teilweise angenommen werden, da sich einige Mittelwertdifferenzen nicht signifikant unterscheiden. Es ist zu vermuten, dass die Rohwertpunktdifferenzen innerhalb eines Items zwischen den Messzeitpunkten nicht groß genug sind. Hierzu müsste die Messung zeitlich (z.B. auf zwei Minuten pro Item) verlängert werden, dies würde jedoch wieder Einbußen in Bezug auf die Praktikabilität des Verfahrens darstellen.

Es konnte durch die Mittelwertdifferenzen jeder zweiten aufeinanderfolgenden Messung gezeigt werden, dass die Probanden im Mittel einen Punktanstieg zwischen einem und drei Punkten zu erwarten haben. Dies bedeutet, dass Phonolino dazu in der Lage ist, den Lernverlauf zu erfassen und dass zwischen den Messungen bedeutsame Unterschiede auftreten. Lernentwicklungen und Leistungsveränderungen somit mit dem Gesamtwert von Phonolino erfasst werden. Die Hypothese H5.2 kann angenommen werden.

Die Trendgeraden der dargestellten Vergleiche der Korrelationen weisen alle durchgängig eine negative Steigung auf. Die Verringerung der Korrelation gilt nicht für jeden einzelnen Vergleich. Es treten teilweise Schwankungen auf, jedoch ist zu erkennen, dass sich die Korrelation zwischen dem ersten Wert (Phono 2) und dem letzten Wert (Phono 9) deutlich verringert.
Hiermit ist die Hypothese H5.3, dass je größer der Abstand zwischen den Messzeitpunkten ist, desto kleiner die Korrelationen sind, aufgrund der hier dargestellten Trends angenommen.

Petermann (ebd.) beschreibt vier Arten von Veränderungsmessung:

- *Intensitätsveränderung*: Dies ist die Veränderung eines Merkmals anhand seiner Ausprägung.
- *Qualitätsveränderung*: Das Merkmal verstärkt sich nicht, sondern es werden erweiterte Kompetenzen erworben. Die Qualität des Merkmals hat sich verändert.
- *Frequenzänderung*: Die Auftretenshäufigkeit vermehrt oder verringert sich.
- *Änderung der Dauer*: Ein Merkmal wird langfristig erworben und kann langfristig aufrechterhalten werden.

Bei der hier gemessenen Veränderung wurde davon ausgegangen, dass sich die Intensität verändert, also dass ein Kind mehr Rohpunkte erlangt, wenn es die Fähigkeit besser beherrscht. Es wurde jedoch nicht berücksichtigt, dass vielleicht in bestimmten Stadien der Entwicklung zunächst eine Veränderung der Qualität auftritt. Diese kann durch Phonolino nur schwer abgebildet werden (dies wäre bspw. möglich, wenn beim Nachsprechen der Pseudowörter auch einzelne Silben ausgewertet würden).

Die Ergebnisse konnten zeigen, dass die CBM Phonolino insgesamt den Lernfortschritt der Kinder im Bereich der phonologischen Informationsverarbeitung abbilden kann. Um eine genaue Aussage darüber treffen zu können, müsste in einer Folgestudie ein allgemeines lineares Modell berechnet werden, anhand dessen die Leistung von Schülern vorhergesagt

werden kann und mit dem eine systematisch formative Evaluation von Unterricht bzw. Förderung im Kindergarten anhand einer Zielsetzung stattfinden kann.

3 Interpretation der Ergebnisse zur Erweiterung des Prognosemodells

Anhand der Ergebnisse der logistischen Regressionsanalyse und der darauf folgenden klassifikatorischen Analysen konnte gezeigt werden, dass für die Prognose von Schwächen im Lesen weder die phonologische Informationsverarbeitung allein noch die phonologische Informationsverarbeitung in Kombination mit der kognitiven Leistung der Kinder in den meisten Fällen eine zufriedenstellende Vorhersage treffen kann. Aufgrund der häufig nicht signifikanten Regressionskoeffizienten mussten viele Gleichungen aus dem Modell 2 ($Phono_k$ + CFT) verworfen werden. Für die zwei Gleichungen (Phono 5 + CFT und Phono 9 + CFT) konnte dagegen ein signifikanter Unterschied zu den Gleichungen (Phono 5 und Phono 9), bei denen nur die phonologische Informationsverarbeitung als Prädiktor verwendet wurde, festgestellt werden. Dies gilt auch für Phono 2, obwohl diese Gleichung bzgl. der Regressionskoeffizienten keine Signifikanz aufwies.
Somit kann die Hypothese H6.1 nur für die Gleichungen „Phono 5 + CFT“ und „Phono 9 + CFT“ angenommen werden, für die anderen ist aufgrund der mangelnden Signifikanz keine Aussage möglich, bzw. zeigt selbst bei Vernachlässigung der fehlenden Signifikanz, dass die Modelle im Modell-Vergleich keinen signifikanten Unterschied aufweisen konnten.

Beurteilt nach den Güteindizes für klassifikatorische Analysen konnte der oben benannte Prognose-Vorteil (in einigen Fällen) nicht bestätigt werden. Hier standen die Gleichungen von Phono 6 und Phono 9 im Vordergrund, in Bezug auf den RATZ-Index interpretiert, der diesbezüglich sehr gute Werte lieferte. Ebenfalls konnten Sensitivität, Spezifität und positive Korrektheit als hoch bewertet werden, sodass Phonolino mit diesen beiden Messungen in der Lage ist, eine zuverlässige Prognose in Bezug auf die Leseleistung treffen zu können. Die klassifikatorischen Analysen konnten in Bezug auf das Prognosemodell 2 nur mit den Gleichungen „Phono 5 + CFT“ und „Phono 9 + CFT“ durchgeführt werden.
Somit muss die Hypothese H6.2 für alle Gleichungen des Prognosemodells 2 zur Vorhersage der Leseleistung abgelehnt werden. Es bleibt jedoch zu erwähnen, dass der RATZ-Wert der Gleichung „Phono 9 + CFT“ immerhin im guten, jedoch unspezifischen Bereich liegt und den anderen Leistungen bzgl. der klassifikatorischen Kennwerte (mit Ausnahme von Phono 6) deutlich überlegen ist.

Für die Vorhersage der Rechtschreibleistung der Kinder konnte die logistische Regression deutlich positivere Ergebnisse offenlegen. Das Prognosemodell 2 ($Phono_k$ + CFT) konnte bei allen Gleichungen einen signifikanten Unterschied zum Prognosemodell 1 aufweisen. Die Chi-Qudrat-Differenzen waren signifikant (Phono 6 + CFT) bis hochsignifikant (Phono 3, 4, 5, 7, 8, 9 + CFT). Dies zeigt, dass die kognitiven Kompetenzen der Kinder einen wichtigen Prädiktor zur Vorhersage der Rechtschreibleistung im ersten Schuljahr darstellen und das bisher durch die phonologische Informationsverarbeitung beschriebene Prognosemodell 1 verbessert werden konnte. Die Hypothese H6.3 kann für alle Gleichungen angenommen werden. Für den Vergleich von Phono 2 und „Phono 2 + CFT“ ist anzumerken, dass der Modellvergleich zwar einen signifikanten Unterschied ergeben hat, die Regressionskoeffizienten in der Gleichung jedoch nicht signifikant waren. Dies kann durch die kleine Stichprobengröße verursacht werden und die Gleichungen wurde daher dennoch zu einem Modellvergleich herangezogen.

Die klassifikatorischen Kennwerte können den Vorteil einer Vorhersage unter Hinzunahme der kognitiven Leistung der Kinder bestätigen. Für „Phono 2 + CFT“, „Phono 4 + CFT“, „Phono 5 + CFT“, „Phono 9 + CFT“ konnte ein RATZ-Index > 66 % berechnet werden. Für „Phono 3 + CFT“, „Phono 6 + CFT“, „Phono 7 + CFT“ und „Phono 8 + CFT“ wurden gute (jedoch unspezifische) RATZ-Indizes erzielt. Somit kann die Hypothese H6.4 für die Gleichungen „Phono 2 + CFT“, „Phono 4 + CFT“, „Phono 5 + CFT“, „Phono 9 + CFT“ angenommen werden. Es zeigt sich, dass sich die Hinzunahme der kognitiven Leistung als Prädiktor einen Prognosevorteil für die Rechtschreibleistung im ersten Schuljahr darstellt.

Interessant zu beobachten ist, dass die Prognoseleistungen insbesondere für die Rechtschreibung deutlich besser ausfallen als es die unter dem Bereich „prognostische Validität“ berechneten Kennwerte hergeben (vgl. Kap. VII.1.4.2, S. 364). Dies ist vermutlich der Effekt der logistischen Regression anhand derer Vierfeldertafeln entstanden sind, bei der kein von der Autorin händisch gesetzter Cut-Off-Wert (bei 25 %) gewählt wurde, sondern die Berechnung den optimalen Wert ermittelt hat. Dies beschreibt wiederum ein Vorteil dieser Methodik. Außerdem ist zusammengefasst zu beobachten, dass die kognitiven Leistungen für die Vorhersage von Leseleistungen im ersten Schuljahr nicht annähernd so sehr von Bedeutung sind, wie es für die Vorhersage von Rechtschreibleistungen der Fall ist. Dies lässt vermuten, dass die Rechtschreibleistungen auf anderen Grundlagen als die Leseleistungen basieren.

Insgesamt lassen die Ergebnisse vermuten, dass für die Vorhersage von Leseleistungen und Rechtschreibleistungen verschiedene Prädiktoren in unterschiedlicher Ausprägung

verantwortlich sind. Ggf. sind für die Leseleistung neben der phonologischen Informationsverarbeitung noch weitere Prädiktoren maßgeblich, die bereits auch in anderen Untersuchungen eine Rolle gespielt haben, z.B. Sprachleistungen (vgl. Hagtvet 1997, S. 174, Scarborough 1991, S. 207, Schulte-Körne, Deimel, Bartling & Remschmidt 1998, S. 28, Roth, Speece & Cooper 2002, S. 262ff. oder Goldammer, Mähler, Bockmann & Hasselhorn 2010, S. 54). Die Rechtschreibleistung lässt sich dagegen durch die kognitiven Leistungen der Kinder in Kombination mit der phonologischen Informationsverarbeitung besser aufklären. Es ist damit zu vermuten, dass die nonverbale Intelligenz im Sinne eines Resilienzfaktors (vgl. Walter & Schulz 2013, S. 37) zugunsten sprachlicher Fähigkeiten wirkt.

Nach einer genaueren Betrachtung, Analyse und Interpretation des in der Längsschnittstudie gewonnenen Datenmaterials werden nun im folgenden Abschnitt die Ergebnisse der Untersuchung zusammengefasst und die zentralen Fragestellungen beantwortet.

4 Zusammenfassende Beantwortung der Fragestellungen

Anhand der Untersuchungsergebnisse wurden die dargelegten Annahmen geprüft und Hypothesen verifiziert beziehungsweise falsifiziert. Dieser Abschnitt zeigt nochmals einen Überblick über alle Annahmen und Hypothesen der Untersuchung sowie eine zusammenfassende Beantwortung der Fragestellungen zu folgenden Bereichen:

F1 Objektivität
F2 Differenzierungsfähigkeit/ Itemanalyse
F3 Reliabilität
F4 Validität
F5 Änderungssensibilität
F6 Erweiterung des Prognosemodells

F1	Ist Phonolino ein objektives Messverfahren?

A1.1 Die Durchführungsobjektivität ist sichergestellt. ✓
A1.2 Die Auswertungsobjektivität ist sichergestellt. ✓
A1.3 Die Interpretationsobjektivität ist sichergestellt. ✓

Die Annahmen zur Objektivität des Verfahrens konnten alle bestätigt werden, somit ist Phonolino ein objektives Messverfahren.

F2 Sind die Messungen von Phonolino brauchbar, also trennscharf und normalverteilt, ist die Schwierigkeit der Aufgaben angemessen und bestehen sie aus den drei Dimensionen „Pseudowörter nachsprechen", „Silben klopfen" und „Lautisolation"? (Itemanalyse)

A2.1 Die Häufigkeitsverteilungen der Items zu den acht Messzeitpunkten sind normalverteilt, sichtbar anhand einer glockenartigen, symmetrischen Kurve. ✓

A2.2 Die Itemschwierigkeiten p_i der Items von Phonolino liegen zwischen p = .20 und p =. 80 (vgl. Bortz und Döring 1995, S. 199). ✓

A2.3 Die Items aller Messungen über die acht Messzeitpunkte hinweg sind trennscharf und es gilt $r \geq .30$ (vgl. Bortz & Döring 2006, S. 220). ✓

A2.4 Die Items laden auf die drei Faktoren „Pseudowörter nachsprechen", „Silben klopfen" und „Lautisolation". ✓

Alle vier Annahmen zur Differenzierungsfähigkeit des Verfahrens konnten bestätigt werden, keines der Items musste eliminiert werden. Die curriculumbasierten Messungen von Phonolino können als trennscharf, normalverteilt und bzgl. der Schwierigkeit der Aufgaben als angemessen betrachtet werden. Außerdem handelt es sich um ein dreidimensionales Verfahren, dass drei sehr unterschiedliche Aspekte der phonologischen Informationsverarbeitung abbildet: die Kapazität des Arbeitsgedächtnisses („Pseudowörter nachsprechen"), die phonologische Bewusstheit im weiteren Sinne („Silben klopfen") sowie die phonologische Bewusstheit im engeren Sinne („Laute isolieren").

F3 Ist Phonolino ein zuverlässiges Messverfahren, d.h. misst es exakt und genau das, was es messen soll?

A3.1 Die Koeffizienten der Paralleltestreliabilität des Gesamtwerts von Phonolino sowie der Items zu den acht Messzeitpunkten liegen im mindestens mittleren Bereich ($r \geq .30$). (Paralleltestreliabilität) ✓ **(mit kleiner Einschränkung für Phono 2)**

A3.2 Die Koeffizienten der Testhalbierungsreliabilität des Gesamtwerts von Phonolino sowie der Items zu den acht Messzeitpunkten liegen im mindestens mittleren Bereich ($r \geq .50$). (Split-Half-Reliabilität) ✓

A3.3 Phonolino weist eine interne Konsistenz zu allen acht Messzeitpunkten mit einem Reliabilitätskoeffizienten im mindestens mittleren Bereich (Cronbachs Alpha $\propto \geq .50$) auf. (Interne Konsistenz) ✓ **(mit einer Einschränkung bei Phono 4 ($\alpha = .485$))**

Die Ergebnisse konnten zeigen, dass Phonolino ein reliables Messverfahren ist, das bis auf kleine Einschränkungen bei Phono 2 und bei Phono 4 exakt misst.

F4	Sind die curriculumbasierten Messungen von Phonolino valide?

A4.1 Phonolino ist inhaltsvalide. (Inhaltsvalidität) ✓

A4.2 Es bestehen mindestens mittlere Zusammenhänge ($r \geq .30$) zwischen den Messungen von Phonolino und der phonologischen Informationsverarbeitung gemessen durch das Bielefelder Screening (BISC), die im gleichen Zeitraum erhoben worden sind. (Übereinstimmungsvalidität) ✓ **für BISC 2, ✖ für BISC 1 liegen die Ergebnisse leicht unter .30**

A4.3 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten korrelieren signifikant im mindestens mittleren Bereich ($r \geq .30$) mit den Leseleistungen der Kinder im ersten Schuljahr, gemessen durch den Knuspel-L von Marx (1998). (Prognostische Validität - Lesen) ✓ (**mit einer Einschränkung bei Phono 2)**

A4.4 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten korrelieren signifikant im mindestens mittleren Bereich ($r \geq .30$) mit den Rechtschreibleistungen der Kinder im ersten Schuljahr, gemessen durch den DRT1 von Müller (2003a). (Prognostische Validität - Rechtschreiben) ✓

A4.5 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten tragen signifikant zur Varianzaufklärung im Lesen, gemessen durch den Knuspel-L von Marx (1998), bei. Für den RATZ-Wert gilt RATZ > 66%. (Prognostische Validität - Lesen) ✖ **RATZ-Wert deutlich niedriger**

A4.6 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten tragen signifikant zur Varianzaufklärung im Rechtschreiben, gemessen durch den DRT1 von Müller (2003a) bei. Für den RATZ-Wert gilt RATZ > 66%. (Prognostische Validität - Rechtschreiben) ✖ (**RATZ-Werte sind gut, jedoch unspezifisch)**

In Bezug auf die Inhaltsvalidität und der Übereinstimmungsvalidität (mit Rücksichtnahme auf die Messung Phono 2, die ggf. noch Anfangsschwierigkeiten aufweist), kann Phonolino als valides Messverfahren angesehen werden. Problematisch ist jedoch die prognostische Validität des Verfahrens, die hier deutlich macht, dass die curriculumbasierten Messungen

nur in Ansätzen im Lesen, dafür essentieller in der Rechtschreibung eine Prognose zu den schulischen Leistungen der Literalität machen können. Durch den angeführten Vergleich zu den Werten des Bielefelder Screenings lässt sich eine leicht abschwächende Wirkung begründen, da das BISC in dieser Studie ebenfalls keine hervorragenden Vorhersageleistungen treffen konnte. Das kann sowohl in der Stichprobe, die eine recht hohe Prävalenzrate aufweist, begründet sein, als auch in dem Fakt, dass die phonologische Informationsverarbeitung ggf. nicht hinreichend eine Begründung für Lese- und Rechtschreibleistungen ist und andere Komponenten hinzugezogen werden müssen. In dieser Studie wurde zudem die Form des Anfangsunterrichts nicht kontrolliert. Viele Schulen beginnen zunächst mit einem phonologischen Training, bevor sie mit einem Lehrwerk oder dem allgemeinen Lese- und Schreiblehrgang starten. Dies könnte dazu beitragen, dass sich die Leistungen der Kinder nochmals aufgrund der veränderten Intervention in der Schule verändern. Auch Walter (2014b, S. 37) beschreibt, dass „die Validität von Screening-Verfahren offensichtlich auch von dem abhängt, was Kinder zwischen dem Prognosezeitpunkt und der Kriteriumsmessung an „Erfahrungen“ sammeln können (vgl. Schabmann et al., 2009).“

F5 Erfasst Phonolino Veränderungen der Leistungen der phonologischen Bewusstheit zu den acht Messzeitpunkten?

H5.1 Die Mittelwerte des Gesamtwerts von Phonolino bzw. der drei Items steigen signifikant über die acht Messzeitpunkte hinweg mit Ausnahme direkt nebeneinanderliegender Messungen. ✓ **Gesamtmessung (außer Phono 3 und Phono 5), für die drei einzelnen Items nur teilweise angenommen**

H5.2 Die Leistung der Probanden steigt durchschnittlich um mindestens 1 Punkt zu jeder zweiten Messung. ✓

H5.3 Je größer der Abstand zwischen den Messzeitpunkten ist, umso kleiner sind die Korrelationen zwischen dem Gesamtwert von Phonolino bzw. zwischen den Items zum jeweiligen Messzeitpunkt. ✓

Einfache Überprüfungen bzgl. der Änderungssensibilität des Messverfahrens Phonolino konnten zeigen, dass Phonolino sensibel die Veränderungen zu jeder zweiten Messung messen kann, dass sich die Leistung der einzelnen Probanden um mindestens einen Punkt sichtbar vergrößert und dass die Messungen, je weiter sie auseinanderliegen, weniger miteinander korrelieren, was auf die Zeit und die erhöhte Leistung in der phonologischen Informationsverarbeitung schließen lassen kann. Es wird deutlich, dass pro Messung nur der Gesamtwert von Phonolino veränderungssensibel ist. Die einzelnen Items lassen nicht

präzise genug die Darstellung der Veränderung der Leistung zu. In weiteren Untersuchungen muss gezeigt werden, dass Phonolino auch in Bezug auf einzelne individuelle Entwicklungsverläufe Veränderungen exakt misst bzw. Verläufe genau beschreiben kann und ob sich diesbezüglich gewisse Subgruppen und deren Lernverläufe unterschiedlich darstellen lassen. Dies wäre ein bedeutsamer Vorteil für die Arbeit in der Praxis, da damit anhand der einzelnen individuellen Verläufe auf Intervention in bestimmten Bereichen geschlossen werden kann.

F6 Kann die Prognose von Lese- und Rechtschreibschwierigkeiten am Ende des ersten Schuljahrs durch die curriculumbasierten Messungen von Phonolino bei Hinzunahme des Prädiktors der nonverbalen kognitiven Leistung verbessert werden? (Prognosemodell 2)

H6.1 Die Güte des Prognosemodells 2 (mit den Prädiktoren der phonologischen Informationsverarbeitung und der nonverbalen kognitiven Leistung im Kindergarten) zeigt in Bezug auf das Modell-Fit (Chi-Quadrat-Wert) zur Vorhersage der Leseleistung am Ende des ersten Schuljahres eine signifikant ($p < .05$) bessere Passung im Vergleich zu einer Vorhersage ausschließlich über die phonologischen Vorläuferfertigkeiten. (Prognosemodell 2 - Lesen) ✓ **nur für die Gleichungen Phono 2 + CFT (mit Einschränkung der nicht signifikanten Regressionskoeffizienten), Phono 5 + CFT und Phono 9 + CFT**

H6.2 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten sowie den kognitiven Leistungen der Kinder tragen mit einem RATZ-Wert > 66% signifikant zur Varianzaufklärung im Lesen bei. (Prognosemodell 2, klassifikatorische Analyse) ✖

H6.3 Die Güte des Prognosemodells 2 (mit den Prädiktoren der phonologischen Informationsverarbeitung und der kognitiven Leistung im Kindergarten) zeigt in Bezug auf das Modell-Fit (Chi-Quadrat-Wert) zur Vorhersage der Rechtschreibleistung am Ende des ersten Schuljahres eine signifikant ($p < .05$) bessere Passung im Vergleich zu einer Vorhersage ausschließlich über die phonologischen Vorläuferfertigkeiten. . (Prognosemodell 2 - Rechtschreiben) ✓ **(mit Einschränkungen bei Phono 2 + CFT aufgrund der fehlenden signifikanten Regressionskoeffizienten)**

H6.4 Die Gesamtwerte von Phonolino zu den acht Messzeitpunkten sowie den kognitiven Leistungen der Kinder tragen mit einem RATZ-Wert > 66% signifikant zur Varianzaufklärung im Rechtschreiben bei. (Prognosemodell 2, klassifikatorische Analyse) ✓ **für „Phono 2 + CFT“, „Phono 4 + CFT“, „Phono 5 + CFT“, „Phono 9 + CFT“ (für „Phono 3 + CFT“, „Phono 6 + CFT“, „Phono 7 + CFT“ und „Phono 8 + CFT“ gut, jedoch unspezifisch)**

Zusammenfassend zeigen die hier dargelegten Ergebnisse der Untersuchung, dass die Hinzunahme kognitiver Faktoren für die Prognose von Rechtschreibleistungen im ersten Schuljahr einen bedeutenden Vorteil bringt und die Prognoseleistung insgesamt verbessert. Für die Vorhersage der Leseleistung konnten nur die Gleichungen Phono 5 + CFT und Phono 9 + CFT als passenderes Modell identifiziert werden, sie lieferten dennoch nur unspezifische oder sogar unzureichende RATZ-Werte. Es ist zu vermuten, dass die nonverbale kognitive Leistung der Kinder für die Rechtschreibleistung eine Art Resilienzfaktor darstellt (vgl. Walter 2014b, S. 19), der defizitäres Vorwissen zwar nicht ausgleichen, jedoch bspw. Instruktionsmängel kompensieren kann. Für die Prognose von Leseleistungen müssen ggf. noch weitere Prädiktoren gefunden werden, die neben der phonologischen Informationsverarbeitung und der nonverbalen Kognition spezifische Aussagen treffen können. Es zeigt sich, dass Phonolino für die Rechtschreibleistung insbesondere in Kombination mit den kognitiven Leistungen ein valides Messverfahren darstellt.

Im nächsten Kapitel folgt eine kritische Diskussion über die Methodik, die inhaltliche Tiefe der vorgestellten Studie, den Umgang mit den Ergebnissen, sowie über eine mögliche Übertragung in die Praxis. Im Ausblick werden dann weitere Forschungsmöglichkeiten sowie prognostische Entwicklungen und Möglichkeiten in der Zukunft erläutert.

IX. Fazit

Zusammenfassend wird die Untersuchung kritisch betrachtet, es werden mögliche Potenziale, Hürden und weitere Einflussfaktoren analysiert. Die Folgen der theoretischen Überlegungen sollen dann in Kombination mit den Ergebnissen dieser Studie für vorhandene Früherkennungsverfahren diskutiert werden. Im Ausblick werden die Grenzen und Möglichkeiten beschrieben, wie die in der Studie erarbeiteten Ergebnisse in den aktuellen Forschungsstand einzuordnen sind und wie sie für mögliche Nachfolgeuntersuchungen genutzt werden können.

1 Kritische Diskussion der Untersuchung

Für die Schlussfolgerungen aus der durchgeführten Studie ist zu berücksichtigen, dass trotz sorgfältiger Planung, Durchführung und Analyse der Daten einige weitere wichtige Aspekte auf die Generalisierbarkeit der Endergebnisse restriktive Auswirkungen haben können. Bezüglich des hier vorliegenden Untersuchungsaufbaus, der Auswahl der Stichprobe und der Untersuchungsinstrumentarien und Kennwerte zur Berechnung, sowie zu der Zusammensetzung der Stichprobe und der Durchführung der Testungen lassen sich einige methodische Punkte benennen, die das Untersuchungsergebnis beeinflusst haben könnten.

Die Stichprobe der vorliegenden Studie ist mit 258 (bereinigt N = 214) für eine quantitative Untersuchung akzeptabel, jedoch bzgl. der Größe nicht zwangsläufig repräsentativ. Aufgrund der Fülle der verschiedenen Testungen und Messungen über einen langen Zeitraum, die insbesondere durch die acht Messungen von Phonolino im letzten Kindergartenjahr zustande kommen, verringerte sich die Zahl der Probanden pro Messzeitpunkt stark. Dies erschwerte eine statistische Auswertung und beeinflusste auch die Ergebnisse. Zudem ist es möglich, dass in der Ausfallquote (Drop-Out) der Schüler, die an weiteren Testungen aufgrund einer Zurückstufung in den Kindergarten, eines Umzugs oder einer Verweigerung an der Teilnahme an weiteren Testungen, in höherem Maße die Kinder enthalten waren, die auch schon in den Vorläuferfertigkeiten Schwierigkeiten hatten und bei denen eine valid positive Klassifikation möglich gewesen und damit die Sensitivitätsrate der Untersuchung noch höher gewesen wäre. Dies beschreibt die Problematik einer Veränderungsmessung bei wiederholter Erfassung einer Stichprobe und der Gefahr einer positiven bzw. negativen Stichprobenselektion, die die Ergebnisse potenziell verzerren könnten (vgl. Kap. VI.2.3, S. 300, sowie Klauer 2001, S. 824f.).).

Eine weitere Problematik könnte der zufällige Anteil an schwächeren Kindern in der vorliegenden Stichprobe sein. Es ist aufgrund der hohen Prävalenzrate der Studie zu überprüfen, ob bei einer anderen Stichprobe ähnliche Ergebnisse replizierbar sind, da dies bspw. Auswirkungen auf die klassifikatorischen Kennwerte hat. Die RATZ-Werte lassen sich durch hohe Selektionsquoten künstlich erhöhen, hohe Selektionsquoten entstehen, je höher die Prävalenzrate einer Studie ist. Bei einer Gesamtzahl von 258 bzw. 214 Kindern ist es möglich, dass die Prävalenzrate nicht der Realität entspricht, da die Stichprobe zu klein ist. Normalerweise liegt der Anteil der Kinder mit Schwierigkeiten im Lesen und Rechtschreiben bei 8 – 12 %, je nach Messinstrument vermuten einige Autoren auch bis zu 20 % (vgl. Kap. I - III). Bei der vorliegenden Untersuchung sind es, gemessen mit dem DRT 1 mit mindestens 19,2 % (mit PR ≤ 10), deutlich mehr Kinder, die Schwierigkeiten im Rechtschreiben aufweisen. Gemessen mit Knuspel-L von mindestens 17,3 % (bei T ≤ 37) gilt dies gleichfalls auch im Lesen (vgl. Tab. 43, S. 305). Das entkräftet die Ergebnisse zur prognostischen Validität in mancher Hinsicht. In einer weiteren Studie ist zudem zu überprüfen, ob sich die Ergebnisse der vorliegenden Stichprobe mit einer relativ hohen Prävalenzrate (vgl. dazu Tab. 43, S. 305) an einer anderen Stichprobe mit einer ggf. niedrigeren Prävalenzrate wiederholen lässt. Es könnte sein, dass insbesondere der hohe Einfluss der kognitiven Komponenten bei einer relativ schwachen Stichprobe sich als Risiko- bzw. Schutzfaktor erweist, was ggf. bei einer Stichprobe mit einer geringeren Prävalenzrate nicht der Fall wäre.

Zur Absicherung der Repräsentativität müsste zudem auf bundesweite Daten zurückgegriffen werden, in diesem Fall spielen sozio-politische Hintergründe (dass bspw. in Schleswig-Holstein schon lange Fortbildungen von Erziehern im Bereich der phonologischen Bewusstheit etabliert sind) länderspezifisch eine Rolle.
In den USA konnten Irvin, Alonzo, Nese & Tindal (2013, S. 3) bei einer Überprüfung der CBM im Kindergartenalter sowie in der ersten Klasse erforschen, dass für die Prognose von Leseleistungen neben den akademischen Fähigkeiten Indikatoren des Sozialverhaltens der Kinder beim Übergang in die Schule eine Rolle spielen. Sie konnten feststellen, dass einige Kinder mit einem sehr niedrigen Niveau in den vorschriftsprachlichen Fähigkeiten im ersten Schuljahr ihre Leistung dramatisch steigern konnten. Die Wirkung des Sozialverhaltens im Übergang vom Kindergarten in die Schule erfordert eine Replikation auf deutschem Terrain, die Autoren merken zudem an, dass das Konstrukt „readiness“ für die Schule genauer definiert werden müsse (vgl. ebd.). Das Sozialverhalten der Kinder konnte in dieser Untersuchung nicht maßgeblich kontrolliert werden.

Das Alter der Kinder birgt wiederum die Gefahr, dass die Ergebnisse der Kinder im letzten Kindergartenjahr aufgrund unterschiedlicher Motivation oder Konzentration nicht immer aussagekräftig waren und dadurch möglicherweise die Befunde verfälscht wurden. Diese Komponenten müssten in einer Folgeuntersuchung möglicherweise zusätzlich erfasst und ausgewertet werden.
Des Weiteren ist es ebenfalls möglich, dass ergänzende Informationen zum Sprachstand der Kinder mit und ohne Migrationshintergrund und zum deutschen Sprachanteil im Alltag der Kinder, die aufgrund mangelnder personeller und finanzieller Ressourcen nicht erhoben worden sind, die Aussagekraft der dargestellten Ergebnisse hätten verstärken können. Ebenfalls wären genauere Informationen zum Erstlesewerk oder zur Methodik des Erstunterrichts oder möglicher angebotener Fördermaßnahmen für schwache oder begabte Schüler in den Schulen von Relevanz gewesen. Dies zeigte auch die Interpretation der Ergebnisse, die insbesondere bei der Lesefähigkeit vermuten ließen, dass eine schlechte Prognose auch von der unterschiedlichen Unterrichtsmethodik abhängen könnte.

Zudem ist, wie bei allen empirischen Studien, die mit Kindern durchgeführt werden, die soziale Komponente zu berücksichtigen. Klauer (2001, S. 95) beschreibt dies folgendermaßen:

> *„Pädagogisch-psychologische Forschung hat es häufig mit Kindern und Jugendlichen als Versuchspersonen zu tun. Aus diesem Umstand ergeben sich in zweifacher Hinsicht besondere Verpflichtungen. Sie betreffen einerseits die Zustimmungsproblematik, andererseits den möglichen Eingriff der Forschung in die Erziehung.“*

Die von Klauer (vgl. a. a. O.) genannte Zustimmungsproblematik tangiert diese Studie nicht direkt, da sich die Kindergärten vorwiegend mit der gesamten Kindergartengruppe für die Untersuchung angemeldet haben. Es ist jedoch zu erwarten, dass bei der Auswahl der Kindergärten indirekt eine spezielle Selektion vorliegt, die nicht berechenbar ist. Möglicherweise melden nur die Erzieher ihre Gruppe für eine Untersuchung an, die bereits wissen, dass die meisten Eltern der Kinder der Gruppe keine Kritik an der Teilnahme üben. Somit würde allein durch die Anmeldung eine spezielle Gruppe selektiert. Da die Kindergärten jedoch unterschiedlichen Einzugsgebieten zuzuordnen sind und die Kinder damit aus unterschiedlichen sozialen Verhältnissen stammen, sollten bezüglich der Ergebnisse keine zu großen Verfälschungen aufgetreten sein. Der „Eingriff der Forschung in die Erziehung“ kann auch unter dem Aspekt beurteilt werden, dass die Erzieher anhand augenscheinlich erhobener Ergebnisse der Messungen durch Phonolino besonders

„auffällige“ Kinder gezielt im Bereich der phonologischen Bewusstheit nach ihrem Ermessen gefördert haben. Das konnte durch die Autorin im Nachhinein nicht überprüft werden. Die Schulung der Erzieher der teilnehmenden Kindertagesstätten kann diese Sensibilität für das Thema der Früherkennung und insbesondere der Frühförderung hervorrufen und damit die Realität der Untersuchung verändern (vgl. Klauer 2001, S. 96). Die Elternbefragung oder auch lediglich die Tatsache, dass im Kindergarten eine Messung zur Früherkennung von Problemen im Lesen und Schreiben in der Schule stattfindet, kann die Eltern dafür sensibilisieren und sie dazu anspornen mit ihren Kindern bestimmte Fähigkeiten zu üben und zu trainieren. Es ist möglich, dass dies die Untersuchungsergebnisse massiv verändert. Diese Aspekte der Untersuchung sind jedoch kaum zu kontrollieren, da den Kindern aus sozialen Gründen niemals eine Förderung verwehrt werden kann, insbesondere nicht durch die Eltern. Marx (1992a, S. 249) beschreibt, dass die nach der Durchführung des Screenings wirkenden Einflüsse (er benennt sie „unvermeidliche Unzulänglichkeiten“) der Familie oder der Schule die Prognose erheblich beeinflussen. Diese Vermutung gehört zu einer der drei Säulen der möglichen Klassifikationsproblematik dieser Arbeit, die aus den Faktoren „Inhalt des Screenings“, „exogene und endogene Einflussfaktoren“ sowie aus der „Methodik der Verfahren“ besteht. Obwohl in der vorliegenden Untersuchung die Inhalte weitestgehend konstant gehalten wurden und als essentieller Bestandteil die Methodik des Verfahrens verändert wurde, ist es dennoch sehr wahrscheinlich, dass endogene und exogene Faktoren und somit Schutz- und Risikofaktoren auf die Kinder zwischen der Messung des Prädiktors und des Kriteriums eintreten und wirken und somit die Leistungen der Kinder entgegen der Prognose verändern.

Die hier dargelegten Berechnungen basieren auf der bereits in den USA genutzten ersten der drei Stufen zur Evaluation von CBM von L. S. Fuchs (2004), die vorwiegend in amerikanischen Studien für Veränderungsmessungen eingesetzt worden sind.
Ein Kritikpunkt ist der Umgang mit fehlenden Werten bei der Durchführung der Evaluation der vorliegenden Studie. Es wurde entschieden Fälle in Bezug auf die Messungen von Phonolino fallweise auszuschließen. Gleichzeitig findet bei den Berechnungen auch ein paarweiser Ausschluss statt, um die Stichprobengröße nicht zu stark zu verringern. Damit werden für unterschiedliche Berechnungen unterschiedliche Teilstichproben verwendet. Durch die MCAR-Analyse wird versucht Verzerrungen der Ergebnisse zu vermindern. Die Form des paarweisen Ausschlusses bringt jedoch auch aufgrund der beständig wechselnden Stichprobengröße Schwierigkeiten für die Interpretierbarkeit der Daten mit sich. Für eine Folgestudie wäre es sinnvoll ggf. nur die Kinder zu berücksichtigen, die an allen Messungen von Phonolino teilgenommen haben.

Zu kritisieren ist des Weiteren die allgemeine Problematik, dass die Evaluation einer curriculumbasierten Messung über eher klassische statistische Methoden ausgewertet wird. Klauer (2011, S. 211) beschreibt, dass eine Konzeption einer CBM gemäß der klassischen Testtheorie eigentlich kaum möglich ist, es jedoch bisher keine wirklichen Alternativen gibt. Eine einfache Begründung dieser Schwierigkeit liegt darin, dass bei einer curriculumbasierten Messung Veränderungen erwartet werden, diese Veränderungen bei der klassischen Testtheorie jedoch als Messfehler wahrgenommen werden (vgl. bspw. bei der Paralleltestreliabilität dieser Untersuchung). Wilbert und Linnemann (2011, S. 230) stellen eine eigens entwickelte Vorgehensweise vor, die weitere Berechnungen durch andere Messverfahren benötigt, und ergänzen in vier Schritten die klassische Testtheorie mit der Item-Response-Theorie[35]:

- Schritt 1: Durchführung einer Itemanalyse auf Basis der klassischen Testtheorie (Trennschärfen, Schwierigkeiten, Interne Konsistenz, Homogenität und wenn möglich Retest-Reliabilität)
- Schritt 2: Dimensionalität (über eine konfirmatorische Faktorenanalyse)
- Schritt 3: polytome Raschmodellierung zur genaueren Bestimmung der Itemschwierigkeiten
- Schritt 4: Überprüfung der Testfairness der relevanten Subgruppen durch Berechnung von DIF[36]-Werten (mit Hilfe ordinal logistischer Regressionsanalysen), zusätzlich wird Typ, Richtung und Effektstärke berechnet

Diese Form der Berechnungen waren auf der hier vorgelegten Datenbasis so nicht möglich, könnten aber in Folgeuntersuchungen als alternative teststatistische Methodik eingesetzt werden. Es wäre jedoch auch in einem nächsten Schritt sinnvoll, den Verlauf des Merkmals der phonologischen Informationsverarbeitung an einem individuellen Verlauf einer konkreten Person zu modellieren (vgl. Wilbert & Linnemann 2011, S. 240). Bei der Interpretation der Lernverläufe merkt Klauer (2011, S. 25) jedoch an, dass die Lernfortschrittskurven der Kinder zwar Verläufe darstellen, jedoch durch eine Veränderung der Verläufe nicht automatisch auf die Ursachen zu schließen sei. Die Veränderung kann

35 Psychologische Test unterscheiden im Wesentlichen zwischen den beiden unterschiedlichen statistischen Testtheorien: der klassischen Testtheorie und der Item Response Theorie. Die klassische Testtheorie hat sich über Jahrzehnte als Standard etabliert und ist als „Messfehlertheorie“ bekannt, die ganze Tests analysiert. Bei der Item-Response-Theorie wurden erstmals auch mathematische und stochastische Modelle hinzugezogen (Rasch 1960, Birnbaum 1968), durch sie werden eher itembezogene Analysen durchgeführt (vgl. Becker 2004, S. 33ff.).

36 DIF = Differential item functioning bzw. gruppenabhängiger Indikator, der Gruppenabhängigkeiten aus verschiedenen Gruppen mit denselben Eigenschaften angibt, basiert auf der logistischen Item-Response-Theorie

nicht zwingend auf die Effektivität einer Intervention zurückgeführt werden. Es wäre auch sinnvoll, die Verläufe von Kindern mit bestimmten Merkmalen und deren Lernzuwachs zu beobachten. Ggf. könnte die Betrachtung der Leistungsquartile von Phonolino einen Hinweis auf einen unterschiedlichen Verlauf der Leistungsentwicklung der vier verschiedenen Leistungsgruppen geben. Voß und Hartke (2014, S. 93) beschreiben, dass die Sensibilität einer CBM über einen „Vergleich der Leistungsentwicklungen verschiedener Probandengruppen“ untersucht werden kann. Dies sollte auch für dieses Instrument noch angestrebt werden. Anzumerken ist, dass die Veränderungssensibilität der vorliegenden CBM relativ gering ausfällt und erst über zwei Monate hinweg sich bedeutende Unterschiede ergeben. Dies könnte durch eine Erhöhung der Itemanzahl oder eine Verlängerung der Zeit (bspw. 2 Minuten pro Item) ggf. verändert werden.

Eine weitere statistische Schwierigkeit (wie es die Berechnung der Trennschärfe oder auch der Split-Half-Reliabilität gezeigt hat) liegt in der geringen Anzahl von Items in der curriculumbasierten Messung. Eine Erweiterung dieser kann zeitökonomische Nachteile zur Folge haben, kann aber auch in Maßen einen Vorteil für die statistische Auswertbarkeit des Verfahrens mit sich bringen.

Ein weiterer statistischer Kritikpunkt ist die Berechnung der Veränderungsmessung, die in dieser Arbeit mit einfachen statistischen Mitteln über Korrelationen und die Differenzen von Mittelwerten erhoben wurde. Die erstellten Differenzwerte können einem doppelten Messfehler unterliegen, da sie aus zwei Mittelwerten erstellt wurden und dadurch die Messfehler beider Tests einbeziehen. Die Differenzwerte weisen eine mangelnde Reliabilität auf (vgl. Bortz & Döring 2006, S. 532, Klauer 2001, S. 91). Bortz und Döring (ebd.) argumentieren, dass das Differenzmaß nicht ausschließlich von der Reliabilität abhängig ist, sondern von der Unterschiedlichkeit der wahren individuellen Veränderung, von der Genauigkeit und der Verteilung der Messungen sowie von der Anzahl der Messzeitpunkte (vgl. a.a.O., S. 532ff.). Es wird empfohlen, wenn die Reliabilität des Messinstruments vorliegt, dass auch eine Berechnung der Differenzmaße stattfinden und zur verbesserten Schätzung der wahren individuellen Veränderungen verwendet werden kann. Außerdem sollten in einer Folgeuntersuchung, wie Strathmann (2014, S. 214) es vorschlägt, die Itemschwierigkeiten zufallskritisch untersucht werden. Hierfür empfiehlt er (ebd.) für die Durchführung der Signifikanzprüfungen die Anwendung der Bonferroni-Holm-Prozedur. Die Voraussetzung der gleichen Schwierigkeiten wurde in dieser Untersuchung nur durch eine deskriptive Beschreibung der Itemschwierigkeitsdifferenzen betrachtet und wurde nicht zufallskritisch geprüft.

Eine Schwierigkeit könnte auch die „Regression zur Mitte“ darstellen, die einen Messfehler beschreibt, der deutlich macht, dass sich schwache Probanden vergleichsweise gut verbessern, während sich gute nur mäßig verbessern oder sogar abrutschen. Da es sich bei

der Studie jedoch nicht um ein klassisches Prä-Post-Design handelt, sondern mit acht Messwerten mehr als zwei vorliegen, könnte sich der Fehler etwas relativieren. Eine gute Möglichkeit zur Prüfung in Folgeuntersuchungen zeigt auch Voß (2013, S. 299) in seiner Dissertation zu einer curriculumbasierten Messung, der zur Auswertung der Änderungssensibilität seines Messverfahrens auf die Hierarchische Lineare Modellierung (HLM) zurückgreift. Der Vorteil dieses probabilistischen Verfahrens ist, dass sich die Veränderung innerhalb der Probanden auf die Merkmale zwischen den Probanden schätzen lässt (vgl. Klauer 2001, S. 92).

Die Wiederholung der Messungen zu verschiedenen Messzeitpunkten im letzten Kindergartenjahr sollte die Stabilität von Veränderungen registrieren (vgl. Klauer 2001, S. 825). Möglicherweise wäre eine größere Anzahl an Messungen vonnöten gewesen, um die Zufallsveränderungen von den substantiellen Effekten zu unterscheiden (vgl. ebd.), dies hätte andererseits aber die Testungseffekte, wie z.B. den Sättigungseffekt oder einen Trainings- bzw. Lerneffekt mit sich bringen können. Diese Effekte müssen ebenfalls in der vorliegenden Studie Beachtung finden. Es ist durchaus möglich, dass bei einigen Kindern aufgrund der vorliegenden Messungen Sättigungseffekte (Differenzierung nach oben) aufgetreten sind, da die noch zeitliche schnellere Beantwortung der geforderten Items aus methodischen Gründen (Sprechtempo, ...) nicht möglich war.

Eine weitere Schwierigkeit der Veränderungsmessung und damit der Überprüfung der Veränderungssensitivität stellt das Reliabilitäts-Validitäts-Dilemma dar, das für alle Arten der Veränderungsmessung eine bisher unlösbare Problematik bildet (vgl. a. a. O., S. 825f.).

Das Dilemma beschreibt ...

> *„[...] die Schwierigkeit zu begründen, ob ein später gemessener, vom Vortest verschiedener Wert eine quantitative oder qualitative Veränderung bedeutet und gleiche Werte keine Veränderung ausdrücken oder wegen mangelnder Reliabilität doch eine Veränderung anzeigen.“ (a. a. O., S. 826)*

Petermann (ebd.) konstatiert, dass dieses Dilemma jedoch in der klassischen Testtheorie kaum lösbar ist.

Durch die Auswahl der Messverfahren können ebenfalls Variablen in der Studie einwirken, die die Ergebnisse beeinflussen. Der DRT 1, Knuspel-L und der CFT 1 können den

Leistungsstand der Kinder nur zu einem einzigen Messzeitpunkt erfassen und sind aufgrund dieser Tatsache für äußere Einflüsse, wie z. B. die Tagesform des Kindes, anfällig. Da dies in anderen empirischen Studien ebenfalls der Fall ist, sind sie vergleichbar. Allerdings ist gerade bei Kindergartenkindern der Vertrautheitsaspekt nicht zu unterschätzen. Trotz einer Aufwärmphase können Kinder verängstigt oder durch die Aufregung nicht dazu in der Lage sein, die geforderten Aufgaben adäquat zu beantworten, obwohl sie es in einer normalen Alltagssituation könnten. Außerdem ist es möglich, dass die Normen der Tests veraltet sind. Innerhalb der Gesamtpopulation kommt es bei neu aufgelegten Testnormierungen zu Leistungsverschiebungen. Meistens liegt die Leistung der aktuellen Population unterhalb der Leistung der älteren Population. Dies bedeutet nicht zwangsläufig, dass sich die Leistungen unserer heutigen Schüler drastisch verschlechtert haben, sondern dass das Testverfahren beispielsweise Wörter enthält, die den Schülern nicht mehr geläufig sind (vgl. Moll, Landerl & Kain 2008, S. 141ff.). Dies würde auch die oben erwähnte recht hohe Prävalenzrate begründen. Die letzte Normierung des DRT 1 erfolgte im Jahr 1994, die letzte Normierung von Knuspel-L in den Jahren 1994 bis 1997 und der CFT 1 wurde 1995 zum letzten Mal genormt. Die Neunormierung der WRT-Reihe (Weingartner Grundwortschatz-Rechtschreib-Test für die Klassen 1 bis 4 der Grundschule und der Klasse 5 der Hauptschule) zeigte beispielsweise bei einer Gegenüberstellung der Eichung 1986-1990 mit den Ergebnissen aus den Jahren 2003-2005, dass ältere Normen von Rechtschreibtests durchaus noch für jüngere Schüler der Klassen 1 bis 3 brauchbar sind, während sich die Normen für ältere Schüler ab Ende Klasse 3 drastisch verändert haben (vgl. Birkel & Stammet 2008, S. 61ff.). Es ist fraglich, ob sich diese Ergebnisse auf die hier verwendeten Tests ausweiten lassen, es besteht aufgrund der Ähnlichkeit des Inhalts die Möglichkeit dazu. Der DRT 1 ist ein reiner Rechtschreibtest, während sich die hier verwendeten „Knuspels Leseaufgaben“ mit den Leseleistungen der Schüler befassen. Nach der DIN-Norm 33430 wurde festgelegt, dass eine Testnorm, die älter als acht Jahre ist, nicht mehr geeignet sei (vgl. Kubinger & Proyer 2004, S. 191). Hemminger, Roth, Schneck, Jans und Warnke (2000, S. 191) empfehlen ebenfalls eine Erneuerung der Normwerte nach ungefähr zehn Jahren, da sich zeigte, dass es bei Rechtschreibtests langfristig zu einer Verschlechterung der Werte kommt. In der vorliegenden Untersuchung sind die Normen der angewendeten Testverfahren älter als acht Jahre. Nicht für alle Berechnungen wurden die Normen verwendet, sondern solange es möglich war, wurde auf die Rohwerte zurückgegriffen. Ein weiterer Kritikpunkt an den ausgewählten Verfahren wurde schon von Marx und Weber (2006, S. 257f.) zu der BISC-Replikationsstudie angemerkt: Knuspels Leseaufgaben sind stark auf phonologische Inhalte aufgebaut. Daher kann es sein, dass die Wahl des Untersuchungsinstruments das Vorhersageergebnis beeinflusst. Dies könnte auch bei dieser Studie zutreffen, da die Kennwerte für die Vorhersage durch das Bielefelder Screening wieder besser ausfielen als in

der Replikationsstudie von 2006. Dies könnte ein Indiz dafür darstellen, dass die Wahl von Knuspel-L den phonologischen Messungen näher kommt. Hinzu kommt, dass vom Knuspel-L der letzte Test zur Lesefertigkeit in der ersten Klasse noch nicht durchgeführt werden konnte und somit nur die Aufgaben zu den Vorläuferfertigkeiten im Lesen abgefragt wurden, die noch stärker auf phonologischen Komponenten beruhen, als der nicht durchgeführte Untertest.

Die Klassifikation und Vorhersage der Kinder mit Schwierigkeiten im Lesen und Rechtschreiben ist zudem nicht vollständig, da eine Diagnostik der Lese- und Rechtschreibschwäche im ersten Schuljahr noch nicht mit der notwendigen Zuverlässigkeit gemessen werden kann. D. h. es können zwar die ersten Schwierigkeiten durch den DRT 1 und Knuspel-L festgestellt werden, es kann zu dem Zeitpunkt dagegen nicht mit der notwendigen Korrektheit eine Aussage getroffen werden, da die meisten Lese- und Rechtschreiblehrgänge erst Mitte/Ende des zweiten Schuljahres beendet werden. Somit ist die Interpretation der Vorhersageleistung der CBM durch eine geringe Distanz zur Feststellung der Lese- und Rechtschreibprobleme (im ersten Schuljahr) geschwächt. Aufgrund der Einschränkung der Untersuchung auf einen Zeitraum von zwei Jahren wurden die nachfolgenden Fähigkeiten der Schüler in höheren Schuljahren nicht mehr gemessen.

Ein bedeutsamer Faktor ist gleichfalls die Einschulung, da die Weiterentwicklung der Kinder ferner in Abhängigkeit zum jeweiligen Erstunterricht zu sehen ist. Zum einen ist es möglich, dass der Erstunterricht an den über 40 Schulen in einem unterschiedlichen Tempo vorangeschritten ist und einige Kinder Vorteile bzw. Nachteile aufgrund der Geschwindigkeit des Lehrgangs haben und sich die Verhältnisse von schwachen und leistungsfähigen Lesern und Rechtschreibern innerhalb einer Klasse in Bezug auf die gesamte Stichprobe verschiebt. Zum anderen hat jedes einzelne Kind andere Schutzfaktoren (Motivation, Konzentration, Fleiß, familiäre Unterstützung, …), durch die es seine Schwächen kompensieren kann. Somit ist nicht zu erwarten, dass jedes klassifizierte Risikokind in der Schule eine Lese- Rechtschreibschwäche entwickelt bzw. ein klassifiziertes „Nicht-Risikokind“ nicht doch aufgrund persönlicher anderer Risikofaktoren (familiär, schulisch, psychisch, …) eine Lese- Rechtschreibschwäche entwickeln kann. Die Risiko- und Schutzfaktoren sind bereits in Kapitel II.2 beschrieben.

Auch ein strukturierter Erstunterricht kann einen enormen Einfluss auf die Vorläuferfertigkeiten der Schüler haben (vgl. bspw. Schabmann et al. 2009). Die Schwierigkeiten in der phonologischen Informationsverarbeitung können durch diesen aufgehoben werden. Dieser Einfluss wurde in der vorliegenden Untersuchung nicht

kontrolliert. Es ist jedoch bekannt, dass immer mehr Erstlesewerke und Fibeln die phonologische Bewusstheit in ihren Konzepten berücksichtigen und auch immer mehr Lehrkräfte dem Erstlese- und Rechtschreibunterricht eine Phase zur Übung der phonologischen Bewusstheit voranstellen (vgl. Valtin 2010, S. 8). Die Vermutung liegt nahe, dass der Grund für das schlechtere Abschneiden der phonologischen Informationsverarbeitung bei der Vorhersage von Lese- und Rechtschreibleistungen in der Schule in jüngeren Untersuchungen der veränderte Anfangsunterricht ist, in dem explizite Graphem-Phonem-Korrespondenzen gelehrt werden und dadurch Defizite einzelner Schüler in der phonologischen Bewusstheit ausgeglichen werden können. Viele Lehrer haben vor dem Lernen der Buchstaben und vor den ersten Leseerfahrungen eine Einheit zur phonologischen Bewusstheit durchgeführt, sodass die Unterschiede heute nicht mehr auffallen. Zum Zeitpunkt der ersten BISC-Untersuchungen wurde stattdessen in höherem Maße der ganzheitliche Ansatz im Erstunterricht für den Schriftspracherwerb verwendet, bei dem weniger Graphem-Phonem-Korrespondenzen gelehrt wurden. Dies verdeutlicht, warum bei der vorliegenden Untersuchung auch der Erstunterricht einen Einfluss hatte und damit von Relevanz ist.

Zudem konnten in dieser Untersuchung weitere Faktoren, wie Motivation, sozio-ökonomischer Hintergrund und insbesondere die sprachlichen Fähigkeiten aus ressourcentechnischen Gründen nicht erhoben werden. Anhand der bereits beschriebenen Befunde (vgl. Kap. II.1.4, S. 156) ist zu vermuten, dass sprachliche Leistungen im Kindergarten ebenfalls ein wichtiger Prädiktor für Lese- und Rechtschreibleistungen in der Schule darstellen. Die Messfaktoren könnten sich hierbei insbesondere an den sehr vielversprechenden Ergebnissen des Simple-View-Readings-Ansatzes orientieren, die bereits umfassend bestätigt wurden (vgl. ebd.). Das Ignorieren dieser Aspekte könnte auch die Relevanz der nonverbalen Intelligenz in dieser Untersuchung erhöht haben, da mangelhaft ausgeprägte schriftsprachliche Leistungen ggf. durch den IQ kompensiert wurden, die besser durch die sprachlichen Leistungen hätten vorhergesagt werden können. Walter (2014b, S. 34) konnte in seiner Studie zur Vorhersage von globalen schulischen Leistungen unter Hinzunahme des nonverbalen IQs eine deutlich bessere Prognose erstellen, als nur auf der Grundlage von basalen sprachlichen Kompetenzen in Kombination mit Aspekten des Arbeitsgedächtnisses. Dennoch betont er, dass kognitive Ansätze eine gute Prognose liefern, die praktische Relevanz und die Wirksamkeit bereichsspezifischer Förderansätze aber noch bestätigt werden müssen. Die reine Übung allgemeiner kognitiver Kompetenzen trägt nicht unbedingt zu einer Verbesserung in der Lese- und Rechtschreibleistung bei. Er beschreibt (2014b, S. 37):

> *„Ob und in welchem Mischungsverhältnis sowohl domänenspezifische als auch allgemein-kognitive Variablen im Rahmen von Screening-Verfahren ihren Platz haben sollten, werden letztlich (auch im Hinblick auf die jeweilige spezielle Diagnose-Intention) weitere Untersuchungen zeigen müssen.“*

Dies gilt im selben Maße wie in dieser Untersuchung. Nur aufgrund eines positiven Prognoseergebnisses darf nicht auf ein allgemein-kognitives Intelligenztraining geschlussfolgert werden, das die schriftsprachlichen Leistungen vermeintlich verbessern kann.

Ein weiterer wichtiger Punkt in den Ergebnissen sollte betrachtet werden: Die Praxisrelevanz eines Erhebungsverfahrens, also dessen ökonomischer Einsatz spielt eine erhebliche Rolle. Die Messungen von Phonolino sind aufgrund der Einsetzbarkeit parallel zum Alltag sehr zeitökonomisch anwendbar, da eine Messung durchschnittlich nur zehn Minuten dauert. Ein Gruppentest wäre noch effizienter, dieser konnte aufgrund der mangelnden Lese- und Schreibfertigkeiten und aufgrund der Berücksichtigung der motivatorischen Komponenten nicht realisiert werden. Die parallele Überprüfung der kognitiven Fähigkeiten durch den CFT 1 dagegen nimmt sehr viel Zeit in Anspruch und kann nur von einem sehr gut geschulten Testleiter durchgeführt werden. Die Handhabung der hier entwickelten CBM Phonolino wäre also aufgrund der zusätzlich geforderten Komponenten (kognitive Leistung und Informationen über die Frühförderung der Kinder) damit sehr eingeschränkt, wenn nicht gar unmöglich, da zusätzlich das eigentliche Ziel der Evaluation einer Lernfortschritts-messung verfehlt würde. Die Nützlichkeit der Messungen, die im Kindergarten von den Erziehern parallel zum Alltag durchgeführt werden können, ist sehr hoch, da eine spezifische Klassifikation der Kinder und eine Förderung bereits im Vorschulalter sowohl in der Forschung, als auch in der Praxis als sinnvoll erachtet wird. Die Ressourcen in Bezug auf die Fördermittel sind meistens begrenzt und es ist, wie die vorliegenden Ergebnisse zeigen, schwierig, diejenigen Kinder zu klassifizieren, die tatsächlich einen erhöhten Bedarf der Förderung haben. Insgesamt ist es wichtig, allgemeine Konventionen für Interventionen zu finden, sodass Interventionen individuell an die Kinder angepasst und bei Misserfolg, gemessen durch ein veränderungssensibles Messverfahren wie eine CBM, verändert werden können. Dafür muss die vorliegende curriculumbasierte Messung Phonolino noch in Zusammenhang mit Schülerprofilen und individuellen Leistungsverläufen im Detail untersucht und die Frage gestellt werden, inwieweit sich diese individuellen Verläufe in Bezug auf eine individuelle Förderung verändern. Dies kann aus wissenschaftlicher Hinsicht in Bezug auf eine unterschiedliche Entwicklung von Schülerprofilen je nach Intervention

von Interesse sein. Das hier vorgestellte Messverfahren kann diesen Ansprüchen insbesondere hinsichtlich der prognostischen Validität nicht genügen. Weiterrhin ist die ökonomische Handhabbarkeit durch Hinzunahme eines IQ-Tests nicht einzuhalten. Die Komplexität der Umwelteinflüsse sowie weiterer Variablen, wie beispielsweise die kognitiven Komponenten, darf nicht vernachlässigt werden. Die phonologische Informationsverarbeitung ist als einziger Faktor zur Früherkennung von Schwierigkeiten in der Schriftsprache nicht ausreichend. Insbesondere die kognitiven Variablen, sowie auch ggf. sprachliche Fähigkeiten, wie es die Untersuchungen von Goldammer, Mähler, Bockmann und Hasselhorn (2010, S. 54), Ennemoser, Marx, Weber und Schneider (2012, S. 63) sowie Marx und Jungmann (2000, S. 90) gezeigt haben, sollten als weitere Faktoren mit ins Feld geführt werden. Im Folgenden werden diesbezüglich weitere Möglichkeiten der Veränderung der Diagnostik zur Früherkennung von Lese- und Rechtschreibschwierigkeiten ins Feld geführt.

2 Erkenntnisse für Diagnostikverfahren zur Prädiktion von Lese- und Rechtschreibfähigkeiten

Grundsätzlich sind Fehlklassifikationen bei Screening-Verfahren unvermeidbar (vgl. Marx & Lenhard 2010, S. 14). In der Dichotomisierung in Risiko- und Nicht-Risikokinder sowie der Einteilung der schriftsprachlichen Leistungen, zu deren Vorhersage die vorherigen Variablen zuträglich sind, verbirgt sich ein großes Fehlerpotenzial. Besonders die Kinder, die auf der Grenze zwischen den beiden Einteilungsbereichen liegen, können kaum genau vorhergesagt werden. Der Cut-Off-Score spielt dabei eine zentrale Rolle und stellt bei der Dichotomisierung nicht ein Intervall, sondern einen einzelnen Wert (z. B. PR = 25) dar. Nur ein Prozentrangpunkt mehr oder weniger hätte eine andere Einstufung des Kindes zur Folge. Dabei kommen sowohl der Kombination der unterschiedlichen Tests, die das Kriterium messen, als auch dem Zeitpunkt eine hohe Relevanz zu für die Vorhersageleistung (vgl. ebd.). Der Cut-Off-Score wird je nach Screening-Verfahren unterschiedlich gewählt. Es muss entschieden werden, ob möglichst wenige falsch-positive Testergebnisse gewünscht sind, wobei riskiert wird, dass einige Kinder mit späteren Schwierigkeiten nicht entdeckt werden, oder ob möglichst viele Personen, die das Problem entwickeln, erkannt werden sollen (so erhält man voraussichtlich eine größere Gruppe von Kindern mit einem falsch-positiven Ergebnis). Wichtig ist bei Nutzung von Screening-Verfahren in der Praxis, dass Eltern über diese Irrtumswahrscheinlichkeiten aufgeklärt werden. Von gleicher Bedeutsamkeit ist, dass grundsätzlich Präventionsmaßnahmen, insbesondere im Kindergarten, nicht nur als Maßnahme für „Risikokinder“ angeboten werden, sondern als

Frühförderung für alle Kinder auf ihrem derzeitigen Niveau und Entwicklungsstand (vgl. Hartke & Koch 2008, S. 42).

Ein Screening im Vorschulalter ist für eine grobe Erfassung und Einordnung der Leistungen scheinbar ein sinnvolles Instrument. Es sollten jedoch auch die kritischen Stimmen zu Screening-Verfahren gehört werden. Marx und Weber (2006, S. 258) stellen in Frage, dass es die Möglichkeit gibt, ein Screening-Verfahren zur Früherkennung von Schwierigkeiten im Lesen und Schreiben zu entwickeln, das mehr als die Hälfte der Risikokinder identifizieren kann. Sie gehen davon aus, dass nur die phonologische Bewusstheit im weiteren Sinne sinnvoll erfasst werden kann, aber die Zusammenhänge zwischen dieser und der phonologischen Bewusstheit im engeren Sinne und deren Transferleistung auf den Schriftspracherwerb eher mäßig ist. Eine weitere Schwierigkeit bestehe darin, dass die Zuordnung der Kinder im untersten Quartil bzgl. ihrer Rechtschreibleistung innerhalb des ersten Schuljahres nicht stabil ist. Die Sensitivität liegt auch hier nur knapp über 50 %, so ist nach Marx und Weber (2006, S. 258): „[...] nicht zu erwarten, dass ein vorschulisches Verfahren Sensitivitäten und Prädiktortrefferquoten von 50 % oder mehr aufweisen kann." Dies gilt auch für die vorliegende Studie. Für den alleinigen Einsatz der Messungen von Phonolino als Prädiktor lag die Sensitivitätsrate zwischen 46 und 60 %. Die Werte, durch die logistische Regression ermittelt, lagen bei bis zu 78% und für die Erweiterung des Prognosemodells um den kognitiven Faktor konnten sogar höhere Sensitivitätsraten von bis zu 80 % berechnet werden. Einige Autoren empfehlen gar die Aufgabe der Unterteilung in Risiko- und Nichtrisikokinder durch Screening-Verfahren (vgl. Leach 1983, S. 47ff., in Marx 1992a, S. 263) und schlagen stattdessen vor allen Kindern eine Förderung zukommen zu lassen. Brügelmann (2005, S. 147) verfasste ein vernichtendes Fazit für vorschulische Screening-Verfahren:

> *„Die Kriterien für Schulerfolg lassen sich nicht unabhängig vom außerschulischen Kontext bestimmen – und auch die Suche nach den Ursachen von Misserfolg darf nicht auf individuelle ‚Voraussetzungen' und auf innerschulische ‚Bedingungen' verkürzt werden. Nur im Bewusstsein dieser Einschränkungen ist eine Diskussion über Risikoprognosen des Schulerfolgs angemessen zu führen."*

Den durchführenden pädagogischen Fachkräften sollte stets bewusst sein, dass bei jedem Screening-Verfahren immer ein Anteil an Fehlklassifikationen auftreten wird, der negative Konsequenzen für falsch-negativ wie auch für falsch-positiv klassifizierte Kinder nach sich ziehen kann (vgl. Marx & Lenhard 2010, S. 14f.). Die Konsequenzen für falsch-negative Kinder durch eine nicht genutzte Chance der Frühförderung liegen auf der Hand, doch auch

falsch-positiv klassifizierte Kinder können aufgrund einer sich selbst erfüllenden Prophezeiung, die durch Eltern oder Erzieher provoziert werden kann, negative Konsequenzen wie beispielsweise ein negatives Selbstbild davontragen (vgl. ebd.). Marx und Lenhard (2010, S. 15) empfehlen aus diesem Grund ein Screening ausschließlich als Grobsiebverfahren, dem eine Feindiagnostik folgen sollte. Wird eine CBM als interventionsprüfendes Verfahren genutzt, ist ein Screening mit anschließenden Folgemessungen von Vorteil. Dadurch ist es dann vielleicht auch möglich, falsch-positive Risikokinder zu späteren Messzeitpunkten zu identifizieren. Marx (1992a, zit. in Hippmann 2008, S. 194) unterstreicht, dass ein Unterlassen der Frühdiagnostik sich in höheren Kosten und dramatischeren Folgen niederschlagen würde, als es die Gefahr einer Fehlklassifikation tun würde. So sollte ein alternatives Messverfahren evaluiert werden, das dazu in der Lage ist, nicht nur Risiko- und Nichtrisikokinder zu klassifizieren, sondern vor allem den Lernprozess zu begleiten und die Förderung darauf aktiv abzustimmen. Aufgrund der Kritik am Cut-Off-Wert und der daraus folgenden Dichotomisierung der Kinder in Risiko- und Nichtrisikokindern würde sich eine curriculumbasierte Messung als lernverlaufsbegleitendes alternatives Messverfahren anbieten.

Fraglich bleibt, ob die Effekte in der vorliegenden Studie bereits zum Ende des ersten Schuljahres messbar sind. Wie in Kapitel II.1.2.1 erörtert, konnten die Studien von Mayringer, Wimmer und Landerl (1998) bei einer Stichprobe von 560 Jungen zum Ende der ersten Klassenstufe nur geringfügige Zusammenhänge zwischen phonologischen Fertigkeiten und den Lese-Rechtschreibleistungen im ersten Schuljahr entdecken. Ihre Vermutung ist, dass die Überprüfung der Leistungen im ersten Schuljahr an einem zu frühen Zeitpunkt stattgefunden hat und der phonemorientierte Erstunterricht zu langsam vorangeschritten ist (vgl. a. a. O., S. 66f.). Dies könnte auch ein Kritikpunkt an der vorliegenden Studie sein. Prognostisch gesehen könnte dies für die vorliegende Studie bedeuten, dass bei einer Überprüfung der Lese- und Rechtschreibleistungen zu einem späteren Zeitpunkt die Ergebnisse besser ausgefallen wären. Eine weitere Möglichkeit wäre, dass lese-rechtschreibschwache Schülerinnen und Schüler erst bei komplexeren Leistungen (sinnerfassendes Lesen, orthografisch korrektes Schreiben, ...) in höheren Schuljahren an ihre Grenzen stoßen und die Prädiktoren diese Schwierigkeiten vorhersagen, während die Schwierigkeiten in der ersten Klasse unerkannt bleiben. Dies spricht auch für die in Kapitel I.1.3.2 (vgl. S. 50) erwähnte Instabilität der Lese-Rechtschreibschwäche bis zum Ende des zweiten Schuljahres.

Die vorliegende Untersuchung konnte einen deutlichen Unterschied bei der Vorhersage von Lese- zu den Rechtschreibleistungen feststellen. Dies steht im Einklang mit der Studie von

Wimmer und Mayringer (2002, S. 272ff.), die aufzeigten, dass die phonologische Bewusstheit besser die Rechtschreibleistung als die Leseleistung vorhersagen konnte. Bereits in der Trainingsstudie von Bus & Ijzendoorn (1999) wurden bei einem Training der phonologischen Bewusstheit nur 12 % der Varianz der Lesefähigkeit aufgeklärt (vgl. Kap. II.1.2.1.2, S. 118), während Blaser et al. (2007, 2010) identifizierten, dass sich Effekte eines Trainings der phonologischen Bewusstheit nur noch für die Rechtschreibleistung am Ende des zweiten Schuljahres nachwiesen ließen, nicht für das Lesen (vgl. Kap. II.1.2.1.2, S. 118). In einer neuen Studie zur Evaluation eines Diagnose- und Fördermaterials konnten Hippmann, Jambor-Fahlen und Becker-Mrotzek (2016, S. 62ff.) ebenfalls zeigen, dass die vorschulische Diagnostik sich als problematisch für den Bereich der Leseentwicklung erwies. Nach der Dichotomisierung in Risiko- und Nicht-Risikokindern konnten die Autoren für die Vorhersage der Lesefähigkeiten zur Mitte und zum Ende des ersten Schuljahrs durch die Prädiktoren vorschulischen Schriftwissens (das die phonologische Bewusstheit inkludierte) nur einen RATZ-Index von jeweils 16 % berechnen (vgl. a. a. O., S. 64), der als sehr niedrig einzustufen ist. Nach der Erstellung von vier unterschiedlichen Lesegruppen und der Interpretation der Verläufe schließen die Autoren darauf, dass die Leseentwicklung ein dynamischer Prozess ist, in dessen Verlauf die Kinder nicht stabil einer Lesegruppe zugeordnet werden können. Somit könne die „Dynamik der Leseentwicklung besonders zu Beginn des Schriftspracherwerbs (...) zur Erklärung der schlechten Vorhersagewerte beitragen“ (a. a. O., S. 58). Dies beschreibt auch die Schwierigkeiten der Vorhersage der Leseleistung (im Vergleich zur Rechtschreibleistung) in der vorliegenden Studie. Die Einteilung in verschiedene Lesegruppen wäre auch für die Phonolino-Messungen ein interessanter Aspekt für weiterführende Forschungsvorhaben. Clemens, Hagan-Burke, Luo, Cerda, Blakely, Frosch, Gamez-Patience und Jones (2015, S. 76ff.) konnten im amerikanischen Raum dagegen feststellen, dass vorschulische Fähigkeiten 35 bis 37 % der Varianz der Leseleistungen des ersten Schuljahrs aufklärten. Hierbei wurden nicht nur die phonologische Bewusstheit, sondern auch das Benennen von Buchstaben sowie Phonem-Graphem-Korrespondenzen überprüft. Hierbei ist bei amerikanischen Untersuchungen darauf hinzuweisen, dass die Kinder bereits im Kindergarten erste Buchstaben kennenlernen und Phonem-Graphem-Korrespondenzen trainieren und damit den Schriftspracherwerbsprozess bereits früher beginnen. Ennemoser, Marx, Weber und Schneider (2012, S. 53ff.) erforschten verschiedene Prädiktoren für die Lesegeschwindigkeit, das Satz- und Textverständnis und die Rechtschreibleistung. Sie stellten, wie auch in dieser Studie nachgewiesen, fest, dass die phonologische Bewusstheit der konsistenteste Prädiktor zur Vorhersage von Rechtschreibleistungen ist, während die Leistungen für das phonologische Arbeitsgedächtnis nur für die erste Jahrgangsstufe einen signifikanten Beitrag leisten konnte (vgl. a. a. O., S. 63). Außerdem konnten Effekte für die nonverbale

Intelligenz, das schnelle Benennen und in höheren Klassenstufen für die linguistische Kompetenz nachgewiesen werden. Diese Studie zeigt zudem, dass ggf. die Fähigkeiten der Schriftsprache für eine eindeutige Vorhersage noch unterschieden werden müssen. Die Autoren differenzieren bspw. die Lesegeschwindigkeit und das Satz- und Textverständnis aus dem übergeordneten Bereich des Lesens heraus und konnten feststellen, dass die nonverbale Intelligenz, die linguistische Kompetenz und das phonologische Arbeitsgedächtnis für die Lesegeschwindigkeit keine aussagekräftigen Prädiktoren sind, während das schnelle Benennen als stärkster Prädiktor identifiziert werden konnte (vgl. a. a. O., S. 61). Beim Satzverständnis konnte der nonverbalen Intelligenz und dem phonologischen Arbeitsgedächtnis ebenfalls keine große Bedeutung als Prädiktor beigemessen werden, der stärkste Prädiktor war das schnelle Benennen; die phonologische Bewusstheit und die linguistische Kompetenz konnten gleichwohl einen Teil der Varianz aufklären. Da in der vorgelegten Studie eine CBM mit der Kombination aus der phonologischen Bewusstheit (Item 2 und 3) sowie des phonologischen Arbeitsgedächtnisses (Item 1) zur Vorhersage sowohl von Lese- als auch von Rechtschreibleistungen erstellt wurde, liegt die Vermutung nahe, dass die Leistungen ggf. genauer differenziert werden müssen. Die Forschungsergebnisse von Ennemoser et al. (2012) zeigten, dass verschiedene Prädiktoren auf unterschiedliche Teilfertigkeiten des Lese- und Schreibprozesses einen Einfluss haben. Es wäre eine Option, die vorhandene CBM anzupassen und Items zur Vorhersage von Rechtschreibleistungen und Items zur Vorhersage von Lesegeschwindigkeit und Leseverständnis zu etablieren.

Weitere Autoren (vgl. bspw. Schnitzler 2008, S. 72; Bus & Ijzendoorn 1999, S. 403ff.; Rothe 2008, S. 21f.; Walter 2002a, S. 48f. oder auch Kap. II.1.2.1.1) vermuteten bereits, dass noch andere Variablen zur Aufklärung der Varianz von schriftsprachlichen Fähigkeiten beitragen. Die Relevanz der phonologischen Bewusstheit wird von Schmid-Barnkow (2003, S. 40) folgendermaßen beschrieben:

> *„Phonologische Bewusstheit lässt sich nicht als Teilfunktion isolieren, dazu ist das Bedingungsgefüge des Schriftlern-prozesses zu komplex und zu individuell bei den einzelnen Kindern. Sie ist nur eine – unbestreitbar wichtige und mit anderen Faktoren verschränkte – Komponente bei der kognitiven Aneignung der Schrift, [...]“*

Wie bereits in Kapitel II.2 dargestellt wurde, werden Schulleistungen von einem multifaktoriellen Bedingungsgefüge beeinflusst. Die vorliegende Untersuchung zeigt, dass für den Bereich Lese- und Rechtschreibleistung ebenfalls mehrere Faktoren relevant sind.

Die Entwicklung des Lesens und Schreibens unterliegt zusätzlich zu den Vorläuferfertigkeiten einem starken Einfluss vieler weiterer Faktoren. Die Korrelationen zwischen der phonologischen Informationsverarbeitung (wobei diese bzgl. der Rechtschreibung höher waren als bzgl. des Lesens) und den späteren Lese-Rechtschreibleistungen sind bspw. zusätzlich bedingt durch den Erstunterricht, durch die Methodik, durch die Unterstützung im Elternhaus oder durch Konzentration und Verhalten des Kindes im Anfangsunterricht. Es ist keine perfekte Korrelation abzusehen. Die in der hier vorliegenden Untersuchung dargestellte Kombination der Variablen Intelligenz und phonologische Informationsverarbeitung konnte tatsächlich bessere prognostische Ergebnisse als die reine Vorhersage der Leistung durch die phonologische Informationsverarbeitung erzielen.

Die hier dargestellten Ergebnisse sind zudem ein Zeichen dafür, dass die nonverbale Intelligenz mit großer Wahrscheinlichkeit einen beträchtlichen Anteil zur Früherkennung von Kindern mit Schwierigkeiten im Rechtschreiben beiträgt, jedoch weniger von Kindern mit Schwierigkeiten im Lesen. Die Diskussion über die Diskrepanzdefinition zwischen kognitiven Leistungen und Lese-Rechtschreibniveau hatte gezeigt, dass diese offenbar keine Relevanz besitzt. Diese Erkenntnis wurde vor allem auf der Basis dessen erbracht, dass die diskrepanten und nicht-diskrepanten Kinder keiner unterschiedlichen Förderung bedürfen. Die Intelligenz leistet aber einen maßgeblichen Anteil zur Vorhersage. Es ist möglich, dass sie nur einen Teil der Vorhersage bestimmt, wie es auch in dieser Studie gezeigt werden konnte. Schneider und Näslund (1993, S. 95) bestätigten dies bereits in der LOGIC-Studie, in der signifikante Korrelationen zwischen den kognitiven Fähigkeiten und der sprachlichen Informationsverarbeitung auf die Rechtschreibung erhoben werden konnten. Gleichfalls stellten die Autoren des BISC (vgl. Kapitel I.2.2.1, S. 71) hohe Korrelationen zwischen den kognitiven Fähigkeiten und der phonologischen Informationsverarbeitung (r = . 60 zum ersten Messzeitpunkt und r = .52 zum zweiten Messzeitpunkt) und zu den schriftsprachlichen Leistungen (r =. 60 zum ersten Messzeitpunkt und r = .62 zum zweiten Messzeitpunkt) fest. Dabei ist ferner auf den sogenannten Matthew-Effekt hinzuweisen, der besagt, dass die Leseleistung die Entwicklung von Intelligenzleistungen beeinflusst und sich dies in IQ-Tests messbar bemerkbar macht. Weber (2003, S. 404f.) konnte zeigen, dass die kognitiven Leistungen von schwachen Lesern mit zunehmendem Alter im Vergleich zu Kindern mit guten Leseleistungen sinken. Somit nimmt, je älter ein Kind wird, der Unterschied zwischen Intelligenzleistung und Lese-Rechtschreibleistung zu. Es ist also möglich, dass der unspezifische Prädiktor der Intelligenz vor dem Schriftspracherwerb von höherer Relevanz ist, da die Leistungen größtenteils noch nicht von den Lesefertigkeiten der Schüler beeinflusst sind. In dieser Studie ist das nur teilweise der Fall, da die Verbesserung der

Prognose durch Hinzunahme von kognitiven Leistungen nur für einige Modelle bestätigt werden konnte. Zu beachten ist außerdem, dass die Intelligenz im Alter von 5 bis 9 Jahren veränderbar ist. Bis zu einem Alter von 14 Jahren verändert sich die Intelligenzleistung der Kinder unregelmäßig und ist damit schwer prognostizierbar (vgl. Weiss & Osterland 1997, S. 25). Die Ergebnisse der vorliegenden Untersuchung stehen also keinesfalls im Widerspruch zur Forschungslage, sondern bilden einen spezifischen Teilbereich ab, in dem die kognitiven Leistungen von Kindern einen erheblichen Anteil ausmachen. Mit den in dieser Studie gefundenen Ergebnissen bestätigen sich auch die Erkenntnisse von Muter, Hulme, Snowling und Taylor (1997, S. 360ff.). Sie konnten herausfinden, dass die Intelligenz das Lesen indirekt über die phonologische Bewusstheit beeinflusst. Küspert (1998, S. 152ff.) konnte bereits durch eine schrittweise multiple Regressionsanalyse nachweisen, dass nonverbale Intelligenz und frühe schriftsprachliche Fähigkeiten die stärksten Prädiktoren für die Leseleistung sind.

Die mögliche Vorhersage schriftsprachlicher Leistungen durch kognitive Fähigkeiten ist nicht automatisch auf eine folgende Intervention zu schließen, konstatiert Walter (2014b, S. 37f). Bereichsunspezifische Trainings wie bspw. des Arbeitsgedächtnisses scheinen kaum geeignet zu sein (vgl. Melby-Levåg & Hulme 2013, zit. nach Walter 2014b, S. 38). Mähler und Schuchardt (2014, S. 84) formulieren, dass „die Frage noch ungeklärt (sei), ob ein Training die Arbeitsgedächtnisfunktionen von Kindern bedeutsam steigern und auch den Schriftspracherwerb positiv beeinflussen könnte." Hofmann (2006, S. 105) dagegen bringt die Veränderbarkeit von Hirnstrukturen und deren Kompensationsleistung mit ein. Das kindliche Gehirn sei dazu in der Lage Defizite und Schädigungen mit entsprechendem Training auszugleichen. Gleichfalls unterstreichen diese hirnstrukturellen Erkenntnisse zur Plastizität des menschlichen Gehirns die vorliegenden Ergebnisse. Marx und Keller[37] (2010, S. 139ff.) konnten in einer Untersuchung zeigen, dass ein Vorschulprogramm zum induktiven Denken sich auf die Entwicklung der Intelligenz- und der Sprachleistung von Kindern im Vorschulalter auswirkt und außerdem bei Schulkindern substantielle Effekte mit sich bringt. Dies könnte ein Indiz dafür sein, dass eine Früherkennung durch nonverbale kognitive Leistungen sowie die phonologische Informationsverarbeitung, wie es in dieser Studie der Fall ist, die Lese- und Rechtschreibleistungen besser vorhersagen und - wenn man noch einen Schritt weitergeht -, deren Förderung einen Effekt auf diese ausüben könnte. Den Trainingseffekt des induktiven Denkens konnten Klauer und Phye (2008) bereits in einer

37 Marx und Keller (2010, S. 139ff.) untersuchten mit 153 Erstklässlern und 133 Kindergartenkindern inwieweit sich ein Vorschulprogramm zum induktiven Denken auf die Entwicklung der Intelligenz- und der Sprachleistung auswirkt. Die Vorschulkinder konnten eine hohe bis sehr hohe Steigerung ihrer kognitiven nonverbalen Leistung und ihrer Sprachleistung verzeichnen. Bei Schulkindern waren die Effekte etwas geringer, aber ebenfalls aussagekräftig.

Metaanalyse mit 74 experimentellen Studien (N = 3595 Probanden) nachweisen, die acht Jahre später (vgl. Klauer 2014b, S. 10ff.) durch weitere 35 Studien ergänzt wurden. Die Befunde konnten positive Ergebnisse auf die fluide Intelligenz (d = .60), auf das Lernen in Unterrichtslektionen (d = .68) sowie auf weitere kognitive Variablen (d = .65) erzielen. Dies spricht für einen Effekt des Trainings des induktiven Denkens auf schulische Leistungen. Welche Teilbereiche jedoch Effekte nach einem Training verursachen, und welche Bereiche durch ein Training keine wirksamen Effekte hervorbringen, sei durch andere Studien zu untersuchen. Dennoch bleibt zu erwähnen, dass sich „unspezifische Förderansätze für den Bereich des Schriftspracherwerbs bisher nicht bewährt haben (vgl. z.B. zusammenfassend Zimmermann 1996; Walter 2001; Suchodoletz, 2006).“ (vgl. Walter & Schulz 2013, S. 81). Es ist zu vermuten, dass die Intelligenz eher als Risiko- bzw. Schutzfaktor einzuordnen ist und damit als eine Art „Resilienzfaktor“ (vgl. a.a.O., S. 37), der dazu in der Lage ist, gewisse Schwierigkeiten im Schriftspracherwerb zu kompensieren (vgl. auch Bird, Bishop & Freeman 1995, S. 446ff.). Die Förderung des induktiven Denkens in Bezug auf den vorschulischen Bereich für bessere Chancen im Schriftspracherwerb im ersten Schuljahr muss noch untersucht werden.

Zusammenfassend steht die Befundlage der vorliegenden Untersuchung nicht gänzlich in Übereinstimmung mit bisherigen Befunden. In vorigen Studien ließen sich korrelative Zusammenhänge zwischen den Messungen zur phonologischen Informationsverarbeitung und den Lese- und Rechtschreibleistungen wiederfinden (vgl. bspw. Vellutino & Scanlon 1987, S. 321ff.; Mann 1984 S. 117ff.; 1986 S. 69ff.; 1991 S. 194ff.; Schneider & Näslund 1992, S. 256ff; Steinbrink, Schwanda & Vogt 2008, S. 649; Jansen, Mannhaupt, Marx & Skowronek 2002, S. 48ff.). Hierbei ist aber noch zu untersuchen, inwieweit die phonologische Bewusstheit tatsächlich eine wichtige Vorläuferfertigkeit darstellt, oder aber sie während des Schriftspracherwerbs erworben wird und sich weiterentwickelt (vgl. Valtin 2010, S. 6ff. oder Rothe 2008[38]) und dadurch ggf. zuverlässige Prognosen ohne Berücksichtigung des Erstunterrichts unzuverlässig sind. Außerdem wurde auch der Zusammenhang zwischen der Intelligenz und den Lese- und Schreibleistungen repliziert (vgl. bspw. Pfeiffer & Zielinski 1975, S. 1ff.; Deimel, Ziegler & Schulte-Körne 2005, S. 6). Zusätzlich konnten die Befunde zeigen, dass Phonolino ein reliables und valides Messverfahren ist, das aber in der prognostischen Validität Schwächen aufweist. Die Kombination aus curriculumbasierter Messung mit Inhalten der phonologischen Bewusstheit wurde in diesem Zusammenhang erstmalig im deutschen Sprachraum überprüft. Andere Studien zur CBM hatten jedoch vermuten lassen, dass es sich hierbei um ein reliables und valides

38 Die Untersuchung von Rothe 2008 weist methodische Schwierigkeiten auf, daher ist nicht eindeutig zu belegen, ob die gemessenen Effekte ggf. Messfehlern unterliegen.

Messverfahren handelt (vgl. bspw. Deno 1985, 2003; Fuchs & Fuchs 1993; Madelaine & Wheldall 1999; VanDerHeyden, Witt, Naquin & Noell 2001; Ardoin et al. 2004, zusammenfassend Waymann et al. 2007). Dies wurde in dieser Studie repliziert, wobei die Werte geringer ausfallen als die vorliegenden CBM im Schulalter vermuten ließen. Die Prognose von Schwierigkeiten im Lese- und Rechtschreiben in der Schule konnte hauptsächlich in der Kombination mit der Intelligenzleistung der Kinder zu einer angemessenen Vorhersageleistung führen, die für die praktische Anwendbarkeit kaum Relevanz besitzt. Außerdem konnten erste Indizien gesammelt werden, dass es sich (nach Stufe 2 von L.S. Fuchs) um ein Messverfahren handelt, das (mit den Gesamtwerten der einzelnen Messungen) dazu in der Lage ist, Veränderungen sensibel zu erfassen. Diese Erkenntnisse müssen dringend erneut geprüft und durch weitere Aspekte genauer beleuchtet werden.

Im Vergleich zu bisherigen Vorschul-Screenings konnte die CBM Phonolino dagegen teils bessere Ergebnisse erzielen. Die Studien zu Screenings zur Früherkennung von Lese- und Rechtschreibschwierigkeiten konnten unterschiedlich hohe RATZ-Werte aufweisen (vgl. Tab. 84).

Die Einzelmessungen von Phonolino erzielten mit den höchsten RATZ-Werten 50 im Lesen (Phono 6) und 48 % im Rechtschreiben (Phono 4). In Kombination mit dem CFT ließen sich noch höhere RATZ-Werte (70 % für „Phono 4 + CFT“ im Rechtschreiben und 36 % für Phono 9 + CFT“ im Lesen) erzielen. Damit können die Einzel-Werte der vorliegenden Untersuchung sowohl den RATZ-Wert des Bielefelder Screenings (33 % bzw. 25 % und 16%) als auch den ungefähren RATZ-Wert von unter 25% der Differenzierungsprobe sowie den RATZ-Wert zur Lesegeschwindigkeit vom Rundgang durch Hörhausen (RATZ = 25 %) überschreiten. Nur der RATZ-Wert des Rundgangs durch Hörhausen ist für das Kriterium „Leseverständnis“ mit RATZ = 77% deutlich höher. Die RATZ-Werte zum Rechtschreiben sind in anderen Studien deutlich niedriger als in gewissen Einzelmessungen der hier vorliegenden (vgl. Tab. 84). Bei einem direkten Vergleich sollte dennoch Vorsicht geboten sein. Die Kennwerte sind anhand unterschiedlicher Stichproben zu verschiedenen Zeiten und mit anderen Kriteriumsmaßen entstanden. Sie bieten nur eine grobe Einordnung der hier gewonnenen Ergebnisse. Zudem wurde bei den jeweiligen Untersuchungen keine explizite nonverbale Kognition als Komponente einberechnet, dies könnte die Ergebnisse der vorhandenen Studien aufgrund der Relevanz der kognitiven Leistungen zur Vorhersage von Lese- und Rechtschreibfähigkeiten stark beeinflussen.

Tab. 84: Vergleich der RATZ-Werte anderer Studien zur Früherkennung von Lese- und Rechtschreibschwierigkeiten am Ende der 1. Klasse im Vergleich zu den Ergebnissen der vorliegenden Studie (angelehnt an Marx & Lenhard 2010, S. 12)

Verfahren	Studie	N	RATZ (Lesen)	RATZ (Rechtschreiben)
BISC[39]	Marx & Weber (2006)	176	33 % / 25 %*	16 %
DP	Steinbrink, Schwanda, Klatte & Lachmann (2010)	664	6 – 28 % (je nach Kriterium und Cut-Off)	12 – 18 % (je nach Kriterium und Cut-Off)
PB-LRS	Gütekriterien von Marx & Lenhard (2010, S. 12) berechnet auf der Basis von Barth & Gomm, 2004, Tab. 18	450	-	55 %
HASE	Treutlein, Roos & Schöler (2007, S. 18ff.)	1 562	58 %	49 – 59 % (je nach Kriterium)
Rundgang durch Hörhausen	Martschinke, Kirschhock & Frank, 2005, S. 38 f.)	375	25 % / 77 %*	37 %
Phonolino[40] (Überblick)	vorliegende Studie	162 - 182	-2 bis 50 % (Phono 2 – Phono 9)	19 – 48 % (Phono 3 – Phono 9)
Phonolino[41] (mit CFT)	vorliegende Studie	152 - 171	30 % / 60% (Phono 5 + CFT / Phono 9 + CFT)	48 – 75 % (Phono 2 + CFT – Phono 9 + CFT)
BISC 1	vorliegende Studie	179/180**	37 %	47 %
BISC 2	vorliegende Studie	172/173**	37 %	45 %

Anmerkung: *erster Wert ist die Lesegeschwindigkeit, zweiter Wert das Leseverständnis; ** erster Wert ist die Anzahl der Probanden bzgl. des Lesens, zweiter Wert ist die Anzahl der Probanden bzgl. des Rechtschreibens

39 Die Ergebnisse des BISC zur prognostischen Validität fallen in der vorliegenden Studie deutlich höher aus (vgl. letzte Zeile der Tabelle).
40 Es wurden nur die Ergebnisse mit signifikantem Regressionskoeffizienten ausgewählt. Die Signifikanz der Konstante wurde vernachlässigt.
41 Es wurden nur die Ergebnisse mit signifikantem Regressionskoeffizienten ausgewählt. Die Signifikanz der Konstante wurde vernachlässigt.

Die gesamte Untersuchung erbrachte wissenschaftliche Erkenntnisse, die sich nur bedingt in einer praktischen Erkenntnis niederschlagen. Phonolino ließ sich nicht vollständig, wie erhofft, als eigenständiges Messverfahren zur schnellen Diagnose und Prognose durch Erzieherinnen im Kindergarten validieren (bzw. nur für einzelne Messzeitpunkte), sondern bedarf für eine verbesserte Vorhersage einer Ergänzung durch die Daten einer Intelligenzdiagnostik, was den Einsatz in einer Kindertagesstätte aufgrund fehlender Praktikabilität ausschließt. Indizien für die Veränderungssensibilität der curriculumbasierten Messung liegen vor, sollten jedoch zwingend repliziert und erweitert werden.

Die Untersuchung konnte einen Zusammenhang zwischen zukünftigen Lese-Rechtschreibschwierigkeiten von Vorschulkindern sowie der Vorschulmessung „Phonolino“ aufzeigen, jedoch nicht für alle Messungen zufriedenstellend beantworten. Insbesondere die Befunde der logistischen Regression und die darauf folgenden klassifikatorischen Analysen dieser Studie sind vielversprechend: Durch die Messungen von Phonolino konnten für die Lesefähigkeit gute, aber unspezifische RATZ-Werte sowohl von Phonolino-Einzelmessungen (Phono 6 mit 50 %, Phono 9 mit 35 %), als auch in Kombination mit der kognitiven Leistung (Phono 9 + CFT mit 36 %) erreicht werden. Für die Rechtschreibung konnte mit Phono 2 + CFT (75 %), Phono 5 + CFT (70 %), Phono 9 + CFT (66,4 %) sogar ein sehr guter RATZ-Wert identifiziert werden.

3 Ausblick

Die eingangs erläuterten und nun anhand der Untersuchungsergebnisse überprüften Forschungsfragen lassen verschiedene Schlüsse zu, die Anlass geben, einen Ausblick für weitere Forschungsgegenstände in Praxis und Theorie darzulegen. Im Folgenden werden die Einsatzmöglichkeiten, Chancen und Grenzen von curriculumbasierten Messungen, speziell bezogen auf die phonologische Informationsverarbeitung diskutiert.

Der Ausgangspunkt der vorliegenden Arbeit war die defizitäre Prognostik vorhandener Früherkennungserfahrungen für schriftsprachliche Leistungen im Vorschulalter. Es wurden drei mögliche Ursachenkomplexe vermutet (vgl. Abb. 66), die zu den schlechten Ergebnissen in den Studien und damit zu einer unsicheren Klassifikationsgenauigkeit von Risiko- und Nicht-Risikokindern geführt haben könnten. Dabei wurde außer Acht gelassen, dass die Studien den Forschungsstandards für empirische Studien hinsichtlich einer systematisch wissenschaftlichen Vorgehensweise nicht genügen.

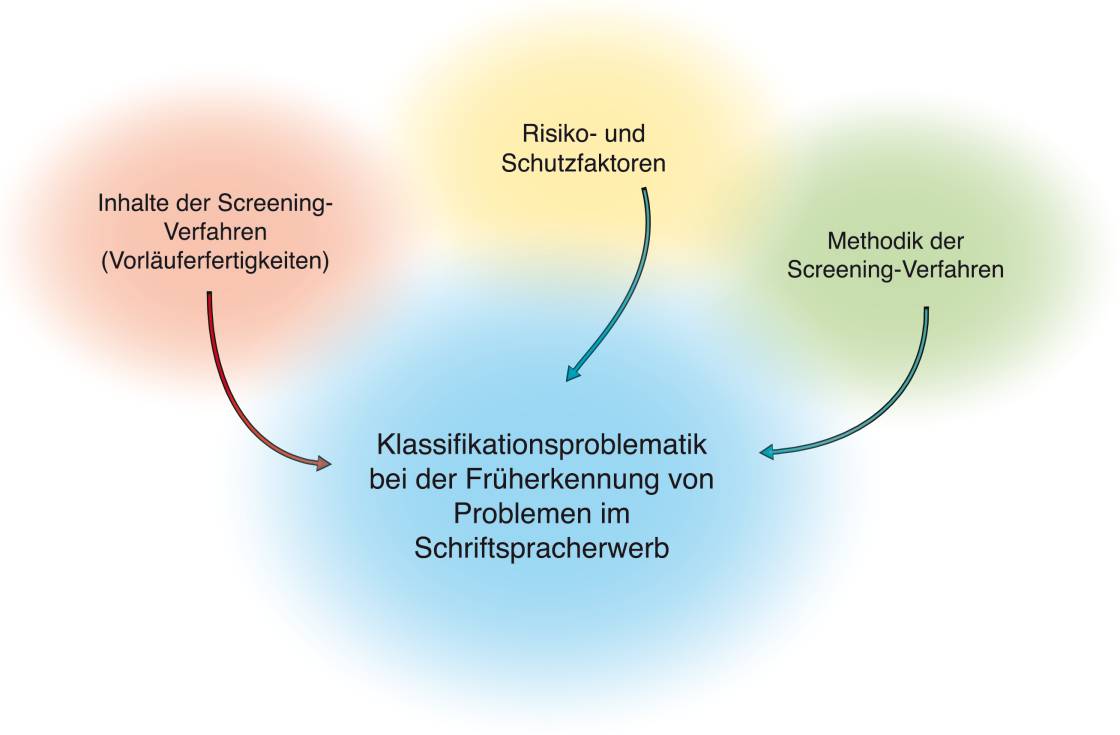


Abb. 66: Ursachenfaktoren für die Klassifikationsproblematik der Screening-Verfahren zur Prognose von schriftsprachlichen Leistungen in der Schule

Die drei Einflussfaktoren auf die Prognosegenauigkeit der Screening-Verfahren waren einerseits die Inhalte der Tests (und damit mögliche Prädiktoren, die vielleicht nicht ausreichend prädiktiv wirken oder durch weitere ergänzt werden müssen) und andererseits endogene und exogenen Faktoren, die vulnerabel oder resilient auf die Leistungen der Schüler bis zum Erhebungszeitpunkt der schriftsprachlichen Leistungen wirken können. Als dritte mögliche Ursache für die unzulängliche Klassifikation der Verfahren wurde die Methodik genannt, die in der vorliegenden Studie verändert und mit einer curriculumbasierten Messung mit dennoch ähnlichen Inhalten wie in den vorhandenen Tests ersetzt worden ist. Hierbei wurde vermutet, dass mehrere Messzeitpunkte die Leistungen der Kinder besser bestimmen könnten und aufgrund der genaueren Ergebnisse eine präzisere Prognose getroffen werden kann. Dies konnte in der vorliegenden Studie noch nicht ausreichend bestätigt werden. Es liegt die Vermutung nahe, dass die anderen beiden Ursachenkomplexe, die hier nicht Teil der empirischen Studie waren, die Ergebnisse so stark verändern, dass sie in Folgestudien näher betrachtet werden sollten. Ein Indiz ist in dieser Studie der Prädiktor der nonverbalen kognitiven Leistung der Kinder im letzten Kindergartenjahr, bei dem noch festzustellen ist, ob er ausschließlich als Risiko- und Schutzfaktor wirkt oder als weiterer Inhalt der Screening-Verfahren aufgenommen werden muss, da er sich als trainierbar herausgestellt hat. Die Ergebnisse konnten zeigen, dass die kognitiven Fähigkeiten einen bedeutsamen Anteil der Vorhersageleistung für Lese- und Rechtschreibschwierigkeiten im ersten Schuljahr einnehmen.

In der vorgestellten Studie wurde die Methodik der Screening-Verfahren variiert, die damit gleichzeitig eine andere methodisch-didaktische Einsatzmöglichkeit über das RTI-Paradigma mit sich bringt. Da jedoch die Prognoseleistung der erstellten curriculumbasierten Messung Phonolino nicht vollständig zufriedenstellend ist, bleiben zusätzlich noch die anderen beiden Bereiche zu erforschen oder aber die Methodik muss mit anderen Inhalten oder einer anderen Art der Durchführung sowie einer anderen Stichprobe erneut evaluiert werden. Zudem ist es von Nöten, die Veränderungssensibilität in Bezug auf die individuellen Leistungen und ggf. auch in Bezug auf ein darauf folgendes Treatment und der Möglichkeit der Verbesserung genauer zu untersuchen. Es handelt sich in diesem Fall um eine Erststudie, die dringend eine Replikation erfordert.

Weiterführende Forschung zu den Inhalten der Früherkennungsverfahren

Höse, Wyschkon, Moraske, Eggeling, Quandte, Kohn, Poltz, von Aster und Esser (2016) konnten nachweisen, dass ein Training der phonologischen Bewusstheit im Vorschulalter nur kurzfristige Erfolge brachte. Die Trainingsgruppe konnte keine Überlegenheit in Bezug auf die phonologische Bewusstheit oder dem Lesen und Schreiben im ersten Schuljahr zeigen. Nur in Bezug auf den Anteil der leseschwachen Erstklässler konnte ein Unterschied zur Kontrollgruppe verzeichnet werden. Die Ergebnisse der vorliegenden Studie zeigen auf, dass besonders die Inhalte der existierenden Screening-Verfahren zu überdenken sind, wobei der Faktor der kognitiven Voraussetzungen der Kinder ins Zentrum der Betrachtung gerückt ist. Die Befunde zum Training im induktiven Denken von Sonntag (2010, S. 5ff.) zeigen, dass dieses zu höheren Leistungen im Intelligenztest (CFT 20) sowie in Variablen der Gedächtnisleistung führt. Diese Befunde zeigen, dass kognitive Leistungen tendenziell trainierbar und dadurch potenziell als Prädiktoren geeignet sind, da aus ihnen eine praktische Relevanz abzuleiten ist. Inwieweit diese kognitiven Faktoren die Vorhersage der schriftsprachlichen Leistungen in Folgeschuljahren verbessern und ein dahingehendes Training die Erfolgschancen für den Schriftspracherwerb erhöht, sollte in anderen Studien ergründet werden. Simanowski-Schulz (2014, S. 161ff.) konnte in ihrer Längsschnittstudie aufzeigen, dass exekutive Funktionen wie Updating[42], Shifting[43] und Inhibition[44] spezifische sensitive Prädiktoren für kombinierte Schwächen in Lesen, Schreiben und Rechnen am Ender der ersten Klasse sind. Es bleibt des Weiteren zu überlegen, ob weitere Faktoren existieren, die prädiktiv für die Lese-Rechtschreib-Leistungen der Kinder wirken. Hier sei beispielsweise auf die linguistische Kompetenz der Kinder verwiesen, deren prädiktive Kraft

42 Updating: Überwachung und Kodierung neu eingehender Informationen
43 Shifting: Fähigkeit irrelevant gewordene Aufgaben zu beenden und sich relevanten Aufgaben zu widmen
44 Inhibition: Fähigkeit zur bewussten oder absichtsvollen Unterdrückung einer vorherrschenden Reaktion

auf die Vorläuferfertigkeiten und auf spätere schriftsprachliche Fähigkeiten kürzlich durch Ennemoser et al. (2012, S. 65/66) nachgewiesen werden konnte:

> *„Die linguistische Kompetenz leistet darüber hinausgehend jedoch zusätzliche Beiträge zur Varianzaufklärung im Leseverständnis und bietet somit einen weiteren (und hierfür spezifischeren) Ansatzpunkt für vorschulische Präventionsmaßnahmen."*

Schon Stanovich und Siegel (1994, S. 24) konnten ausfindig machen, dass die sprachlichen Leistungen eine bessere Vorhersage der Leseleistung als die kognitiven Fähigkeiten treffen konnten. Auch Goldammer, Mähler, Bockmann und Hasselhorn (2010, S. 49) vermuten, dass bei der Früherkennung ein zu geringerer Anteil an sprachbasierten Fähigkeiten als Prädiktoren eingebracht werden. Zusätzlich zu den phonologischen Fähigkeiten zählen bspw. Walter und Schulz (2013, S. 65) weitere nicht-phonologische Variablen wie morphosyntaktische und semantische Kompetenzen, Wortschatz, Sprachverständnis oder auch das Hörverstehen auf. Dies muss gesondert mit den bisher nicht nachgewiesenen Trainingseffekten bei allgemeinsprachlichen Leistungen, die einen Transfer auf die Schriftsprache zuließen, untersucht werden (vgl. a.a.O.). Ein interessantes Untersuchungsergebnis konnte Walter (2014, S. 8) aufzeigen, der ebenfalls einen Einfluss der phonologischen Bewusstheit auf mathematische Fähigkeiten verzeichnete. Diese globale Vorhersage von Schulleistung über die phonologische Informationsverarbeitung muss in weiteren Studien erforscht werden und es ist zu überdenken, ob die phonologische Informationsverarbeitung tatsächlich einen spezifischen Prädiktor für die schriftsprachlichen Leistungen darstellt.

Valtin (2010, S. 6) empfiehlt als Vorbereitung für das schulische Lernen bei Kindern im Vorschulbereich eine umfassende Förderung in den folgenden Bereichen anzubieten:

- Sprachförderung
- Förderung metasprachlichen Verhaltens
- Anleitung zur Vergegenständlichung von Sprache durch spielerische Übungen
- Begegnungen mit konzeptioneller Schriftlichkeit durch Vorlesen und den Umgang mit Büchern
- Malen, Zeichnen und Schreiben, Umgang mit Schreibmaterialien

Inwieweit diese Bereiche eine Verbesserung der Prognose zulassen und Effekte in Bezug auf die Förderung zur Vorbereitung des Schriftspracherwerbs erzielen, muss des Weiteren erforscht werden.

Weiterführende Forschung zu den Risiko- und Schutzfaktoren

Marx und Lenhard (2010, S. 15) beschreiben außerdem, dass die Prognose immer auch abhängig von äußeren Einflüssen oder weiteren Faktoren ist:

> *„Die Prognose von Lese-Rechtschreibschwierigkeiten muss auch deshalb schwerfallen, da ein komplexes Zusammenwirken vielfältiger individueller Faktoren und unterschiedlicher Einflüsse der Umwelt zu diesen Schwierigkeiten führt, wobei ein Screening-Verfahren nur einen Ausschnitt der Faktoren erfassen kann."*

Bei der Entwicklung von Screening-Verfahren existieren damit zwei weitere mögliche Fehlerquellen, die sich (vgl. Hartke und Koch 2008, S. 41) bei der Zuweisung der Kinder zu Risikogruppen auf die Ergebnisse auswirken können:

1) die unzureichende Berücksichtigung von Schutzfaktoren:

   Bei Identifikation eines Risikowertes bei einem Kind kann nicht automatisch von einer logischen Verknüpfung ausgegangen werden. Wenn ein Kind einen besonders hohen Risikowert aufweist, kann es trotzdem sein, dass positive Faktoren (Schutzfaktoren) existieren, die die Risikobelastung im Bereich des Prädiktors aufheben können.

2) die Überbewertung von Risikofaktoren:

   Ein Beispiel aus der Praxis beschreibt die Gefahr der Fehlinterpretation von Risikofaktoren: 90% der Kinder mit Schulschwierigkeiten kommen aus bildungsfernen Haushalten. Somit könnte der Faktor „Bildungsferne der Eltern" als Prädiktor für Schulversagen gewertet werden. Es wurde übersehen, dass 90 % der Kinder, die aus bildungsfernen Familien stammen, mindestens einen Schulabschluss der Hauptschule erreichen (vgl. a. a. O., S. 42).

Hartke und Koch (a. a. O., S. 41f.) vermuten, dass es bei Einzelfallentscheidungen eine relativ hohe Irrtumswahrscheinlichkeit gibt, die schwer einzuschätzen ist. Screenings mangelt es meistens an prognostischer Qualität, die für das einzelne Kind für die Indikation von Präventionsprogrammen auf einer unsicheren Vorhersage beruhen. Somit ist es aufgrund des multifaktoriellen Bedingungsgefüges fragwürdig, ob eine Früherkennung auf diesem Niveau sinnvoll ist. Insbesondere die Indikation von Präventionsmaßnahmen bei Kindern, die als falsch-positiv identifiziert worden sind, wäre zum einen Verschwendung von Ressourcen, könnte aber zum anderen auch negative Folgen wie die emotionale

Verunsicherung der Eltern oder Stigmatisierung des Kindes mit sich bringen (vgl. a. a. O., S. 42). Hierfür bieten die Gleichungen der logistischen Regression einen Vorteil, da diese dazu in der Lage sind, individuell das Risiko für jedes Kind zu berechnen. Doch auch hier ist wie in dieser Untersuchung zu sehen, keine vollkommene Sicherheit zu erwarten.
Zukünftig bleibt zu untersuchen, welche Risiko- und Schutzfaktoren (vgl. auch Kap. II.2 - Endogene und exogene Einflüsse auf die Schülerleistungen) auf die Leistungen der Kinder einwirken und wie diese möglicherweise in ein diagnostisches Modell integriert werden können, sodass die Prognose über die Vorläuferfertigkeiten durch diese Fakten unterstützt werden kann.

Weiterführende Forschung zur Methodik von Früherkennungs-Verfahren

In den USA ist man bereits zu differenzierten Erkenntnissen bei der Entwicklungen von curriculumbasierten Messungen mit einer breiten Basis an Studien im Bereich Lesen gekommen und konnte die curriculumbasierten Messungen als reliables und valides Messverfahren etablieren (vgl. bspw. Ardoin et al. 2004; Colón & Kranzler, 2006; Deno 1985, 2003; Fuchs, Deno & Mirkin, 1984; Fuchs et al., 1989b; Fuchs, Fuchs, Hamlett & Germann 1993; Jones & Krause, 1988; Shapiro et al., 2006; Wayman et al., 2007; Wiley & Deno, 2005; Yeo, 2010). Die Möglichkeit des Einsatzes einer curriculumbasierten Diagnostik im Kindergarten sollte weiterhin beleuchtet werden. Es hat sich herausgestellt, dass die CBM in Kombination mit den Inhalten der phonologischen Informationsverarbeitung im Kindergarten nur partiell die gewünschten Prognose-Ergebnisse erzielen konnte. Die CBM war dennoch dazu in der Lage Veränderungen darzustellen, was eine wichtige Voraussetzung für die Nutzung einer curriculumbasierten Messung in der Praxis ist. Für die pädagogische Praxis wird es zukünftig essentiell sein, die Früherkennungs- und Interventionsmöglichkeiten im vorschulischen Bereich und deren Auswirkung auf schulische Leistungen zu erforschen, Studiengänge für Erzieher zu etablieren, entsprechende Forschungsprofile zu implementieren und damit nachhaltig zur Qualitätsverbesserung der pädagogischen Diagnostik und Förderung im Kindergarten beizutragen (vgl. Thole, Fölling-Albers & Roßbach 2008, S. 29). Der Fortschritt in der Frühdiagnostik und damit einhergehend in der Prävention kann eine Chance darstellen, Kinder durch frühzeitiges Erkennen einer Schwierigkeit vor der Etablierung und der Resistenz dieser Störung zu bewahren. Daher ist sowohl die theoretische als auch die empirische Weiterentwicklung in der Frühpädagogik von großer Bedeutung. Durch eine Lernfortschrittsdiagnostik kann der häufig als problematisch betrachtete Übergang zwischen Kindergarten und Grundschule (vgl. Faust 2008, S. 225; Knauf & Schubert 2008, S. 1ff.) in seinen Folgen abgemildert werden. Diese Art der Dokumentation von Leistungen kann eine Möglichkeit sein, den Lehrkräften einen Datensatz inklusive der Hinweise zu möglichen

Interventionsmaßnahmen zu liefern, um mit der Förderung in der ersten Klasse nahtlos anzuschließen und damit möglichst gute Effekte zu garantieren. Grundsätzlich würde der Einsatz von curriculumbasierten Messungen im letzten Kindergartenjahr zur Früherkennung von Schwierigkeiten im Lesen und Rechtschreiben in der Schule viele Vorteile gegenüber bisherigen Verfahren mit sich bringen, die hier in Kürze nochmals zusammengefasst werden:

- Risikokinder erhalten die Chance, durch die frühzeitige Erkennung nicht ausreichend entwickelter Vorläuferfertigkeiten eine angemessene Förderung zu erhalten. Dadurch kann möglichen Problemen im Schriftspracherwerb entgegengewirkt werden.
- Durch die Dokumentation der Leistungen über ein Jahr hinweg ist der Lernprozess für das pädagogische Fachpersonal in Kindertagesstätten nachvollziehbar und kann als Lernfortschrittsdiagnostik für eine prozessbegleitende Beobachtung bzw. systematisch-formative Evaluation von Fördermaßnahmen (vgl. auch Walter 2010b, S. 163ff.) genutzt werden, um eine individuelle Förderung zu gestalten.
- Vor dem Hintergrund einer stabilen Datenbasis für jedes Kind kann der Übergang zwischen Kindertagesstätte und Schule durchgängig gestaltet werden, da die CBM eine Grundlage für die Lehrkräfte der Schüler bietet, an deren Kenntnisstand anzuknüpfen.
- Individuelle Fördermaßnahmen können durch eine CBM zeitnah evaluiert und auf ihre Wirksamkeit hin geprüft werden. Wenn die Leistungen des Kindes nicht steigen, kann das Interventionsverfahren ggf. auch angepasst werden.
- Die Messfehler, die bei Einmalmessungen entstehen können, werden bei der CBM durch die häufigen Messungen besser kontrolliert.
- Die Daten der CBM ermöglichen aufgrund einer klaren Datenbasis zudem eine Diskussionsgrundlage für Elterngespräche.

Der Anspruch an ein neues Messverfahren sollte sein, dass es eine zeitökonomische und kindgerechte Alternative zu herkömmlichen Testverfahren darstellt, die von geschultem und dem Kind vertrautem Fachpersonal direkt in der Kindertagesstätte durchgeführt werden kann. Dies konnte mit dem vorliegenden Messverfahren (noch) nicht gewährleistet werden, da Phonolino allein nicht mit allen Messungen aussagekräftigen Klassifikationen zuließ.

Die hier vorliegende Studie lässt einige Möglichkeiten einer Replikation offen (vgl. Kap. IX.2 - Kritische Diskussion der Untersuchung). Es ist bspw. vorstellbar, durch die

Erhöhung der Messrate (statt monatlicher Messungen beispielsweise wöchentliche Messungen) eine verbesserte Änderungssensibilität zu erzielen.

Veränderung der Inhalte und der Methodik der Früherkennungs-Verfahren

Die weitere Erforschung der curriculumbasierten Messung erscheint als sinnvoll, da dieses Messverfahren bezüglich seiner hohen Praxisrelevanz über die veränderte Form der Intervention im vorschulischen Bereich besonders gut einsetzbar wäre (vgl. Kap. III.4.2 - Der RTI-Ansatz („Response-To-Intervention")). Die Ergebnisse der Studie in Kombination mit bisherigen Forschungsergebnissen zeigten jedoch insgesamt, dass eine reine Replikation der Studie die Befunde voraussichtlich nicht ausschlaggebend ändern würde, sondern dass explizit eine weitere Variable als wichtiger Prädiktor hinzugezogen werden muss. Somit bestätigen die vorliegenden Ergebnisse auch die Befunde und Schlussfolgerungen von Stock (2009, S. 93), dass neben der Betrachtung phonologischer Merkmale noch andere Komponenten betrachtet werden müssen, jedoch „[...] der phonologischen Bewusstheit i. e. S. vor allem für das Erlernen der Rechtschreibung eine substanzielle Bedeutung beikommt, [...]". Da der Bedarf für ein Verfahren wie die Lernfortschrittsmessung in der Praxis vorhanden ist, wäre zu überlegen, ob in zukünftigen Studien eine curriculumbasierte Messung erstellt wird, die zusätzlich zum Bereich der phonologischen Informationsverarbeitung auch kognitive Komponenten berücksichtigt. So müsste bei der Klassifikation der Kinder kein umfangreicher IQ-Test durchgeführt werden, der die Ökonomie des Verfahrens stark einschränkt. Des Weiteren sollte untersucht werden, welche weiteren Faktoren prädiktiv auf die Schriftsprache in der Schule wirken, sodass diese ebenfalls in der CBM Berücksichtigung finden könnten. Es ist gleichfalls zu bedenken, ob für die Prävention von Schwierigkeiten im Schriftspracherwerb eine jüngere Zielgruppe gewählt werden sollte. Aufgrund der Heterogenität des Entwicklungsstands können einige der Kinder zum Zeitpunkt der Messungen schon lesen und schreiben. Es wäre möglich, dass eine frühere Diagnostik und damit eine frühere Förderung im Kindergarten sich nachhaltiger auf den Schriftspracherwerb in der Schule auswirken.

Besonders sinnvoll wäre zudem die Konzeption einer CBM, die sich sowohl im Kindergarten als auch in der Schule anwenden ließe. Dieser bisher i. d. R. sehr stufenhaft verlaufende Übergang könnte dadurch flexibler gestaltet werden, zumal das Leistungsniveau der Schüler bei Schuleintritt sehr heterogen ist und die „Stunde Null des Schriftspracherwerbs" nicht existiert. Mit einer CBM, die den Übergang beschreibt, könnten erfolgreiche Förderprogramme von der Kindertagesstätte an bis in die Schule angewendet und gleichzeitig zeitimmanent evaluiert werden. Im Zuge dessen wäre die Entwicklung eines Konzeptes für ein adäquates, flexibles Förderangebot im Sinne des RTI-Ansatzes zu

empfehlen. Vermutlich ist es für die Praktikabilität des Verfahrens zudem sinnvoll, dies computerbasiert anzulegen (vgl. Walter 2011a, S. 106ff.; Fuchs 2004, S. 191; Fuchs, Fuchs, Hamlett & Hasselbring 1987, S. 14ff.; Fuchs, Fuchs & Hamlett 1993, S. 49ff.).

Chancen und mögliche Einsatzgebiete

Eine grundsätzliche Chance beim Einsatz von CBM besteht in der pädagogischen Nutzbarkeit der Messungen. Es ist zu begrüßen, wenn in Deutschland ebenfalls wie in den USA eine evidenzbasierte Förderung im Rahmen des RTI-Ansatzes mit Hilfe curriculumbasierter Messungen umgesetzt und damit eine fundierte Alternative zu summativen Verfahren von Schulleistungen geboten werden kann (vgl. auch Voß & Hartke 2014, S. 95). Dies stellt zusätzlich Anforderungen an die Messverfahren, um auf deren Grundlage pädagogische Entscheidungen treffen zu können (vgl. a. a. O., S. 96):

- Die Veränderungssensibilität sollte für verschiedene Leistungsgruppen erhoben werden, wöchentliche Leistungsunterschiede müssen messbar sein.
- Die CBM sollte Normwerte zu verschiedenen Messzeitpunkten vorhalten, um Pädagogen die Möglichkeit zu geben, Zieldefinitionen für die Kinder festzulegen.
- Die prognostische Validität der ersten Messzeitpunkte muss geprüft werden, um einen Trend ermitteln zu können.
- Es müssen Regeln zur Datenanalyse vorliegen, die mit konkreten Handlungsempfehlungen verbunden sind.
- Für die Effizienz des Einsatzes von CBM sollten Daten computerbasiert gesammelt werden können.

Diese Forderungen für die Schule sind gleichzeitig auf den Kindergartenbereich zu übertragen und insbesondere für den Übergang zwischen Kindergarten und Schule anzustreben. Die in vielen Bundesländern neu eingeführte Schuleingangsphase, in der die Schüler die ersten zwei Schuljahre in einem, zwei oder drei Jahren durchlaufen können, hat einen Bedarf an klassifikatorischen Tests geweckt, die beim Übergang der Kinder vom Kindergarten in die erste Klasse bereits Anschlussmöglichkeiten für eine erste Einschätzung bieten und angebahnte und anzubahnende Förderungen liefern, an denen die Lehrkräfte in der Schule wieder ansetzen können. Da in den meisten Bundesländern mit der Einführung der Schuleingangsphase die Zurückstellung für ein Jahr in den Kindergarten aufgehoben wurde, werden sich noch nicht „schulreife“ Kinder in den ersten Klassen befinden, die einen erhöhten Bedarf an Wiederholung der Übungen zur Vorbereitung auf den Schriftspracherwerb haben. Es ist sinnvoll das auf der Grundlage ihrer phonologischen Fähigkeiten zugeschnittene Förderprogramm aus dem Kindergarten an dieser Stelle fortzusetzen beziehungsweise ihre Fortschritte durch ein curriculumbasiertes Messverfahren

zu verfolgen und die Interventionsmaßnahme ggf. anzupassen. Somit ist die curriculumbasierte Messung auch ein sensibles Diagnostikverfahren zur Etablierung der Inklusion in Regelklassen.

Die Ergebnisse heben nicht die Notwendigkeit einer frühen Identifikation und einer frühzeitigen Intervention auf (vgl. auch Kap. I.1 - Notwendigkeit früher Diagnostik und Prävention). Marx (1992a, S. 251) zeigt zusammenfassend auf, dass eine späte Identifikation und damit eine späte Förderung von Kindern mit Lese- Rechtschreibschwäche fatale Folgen mit sich bringen kann. Schulische Interventionsmaßnahmen reichen häufig nicht aus, Fördermaßnahmen weisen bei älteren Kindern oft keinen dauerhaften Erfolg auf und bei schulischen Interventionsmaßnahmen sind teilweise sogar Verschlechterungen zu befürchten. Zudem verschlechtern sich die Schüler oftmals zusätzlich in anderen Schulfächern und sind verhaltens- oder emotional auffällig (vgl. ebd.). Wissenschaftlich fundierte Interventions- und Diagnostikmethoden bedürfen somit ebenfalls in Zukunft einer beständigen Revidierung anhand neuester Erkenntnisse. Eine Intervention kann nur auf dem gegenwärtigen Stand der Wissenschaft basieren und ist nicht dauerhaft gültig. Die Verbesserung der Förderung und Therapie von Kindern mit Lese-Rechtschreibschwäche benötigt somit ein Konzept, das sowohl die verschiedenen Lebenslagen des Kindes und seine individuellen psychosozialen Voraussetzungen einbezieht, als auch eine differenzierte Diagnostik und Therapie sowohl in der Schule als auch außerschulisch und insbesondere vorschulisch berücksichtigt. Anhand einer ausführlichen Diagnostik wird ein Förderkonzept erstellt, das durch weitere Testzeitpunkte immer wieder revidiert und überprüft werden sollte und gerade in späteren Schuljahren nicht immer zum Erfolg führt. Dieser großangelegten und komplizierten Fördermaßnahme können jedoch präventive Maßnahmen in Kindergärten durch eine ausdifferenzierte, den Lernprozess begleitende Diagnostik und anschließende Förderung entgegengestellt werden. Dies könnte eine CBM gewährleisten, deren Inhalte prädiktiv die Lese- und Rechtschreibentwicklung vorhersagen. Hierzu seien auch Tramontana, Hooper und Selzer (1988, S. 89ff.) genannt, die auf eine möglichst frühe Förderung verweisen, um durch fördernde Maßnahmen die Schwierigkeiten zu beheben, bevor sie durch sekundäre Problematiken überlagert werden.

Die Frühdiagnostik stellt zudem die erste Eskalationsstufe des RTI-Modells (vgl. Kap. III.4.2) dar, anhand dessen Kinder sowohl im Kindergarten als auch in der Schule in ihren Lernprozessen begleitet werden können. Je mehr der inklusive Ansatz in deutschen Schulen und Kindergärten umgesetzt wird, desto mehr wird eine Lernfortschrittsdiagnostik ein wichtiges Mittel zur Planung, Durchführung und Evaluation von Fördermaßnahmen darstellen (vgl. Liebers & Seifert 2012, S. 1ff). Die curriculumbasierte Messung besitzt das

Potenzial, dass das gemeinsame Lernen aller Schüler im Sinne der Inklusion kultiviert wird und gleichzeitig die Lehrkräfte in ihrer Arbeit durch ein zuverlässiges Instrument unterstützt werden können. CBM kann ein Mittel sein, eine inklusive Pädagogik zu unterstützen und hierarchische Diagnostik- und Förderstrukturen aufzubrechen und so den Fokus von einer selektierenden Diagnostik auf eine Diagnostik des Lernprozesses und einer Reflexion der Fördermethoden sowie der Potenziale, Möglichkeiten und Chancen des Kindes zu lenken. Es ist nicht geprägt von einer Einteilung der Kinder in sonderpädagogische Kategorien oder von einem Labeling anhand diagnostizierter Defizite, Störungen oder Mängel, sondern bietet die Möglichkeit das Lernen der Kinder zu begleiten, zu strukturieren und gegebenenfalls Lernprozesse und Fördermaßnahmen zu optimieren.

Die o. g. Verfahren stellen eine wichtige Strategie dar, die Qualität gemeinsamen Unterrichts im Sinne einer inklusiven Pädagogik zu verbessern. Sie entsprechen und unterstützen Bemühungen um die Verminderung von pädagogischer und diagnostischer Aggressivität sowie hierarchischer Diagnose- und Fördervorstellungen, die Personen zum Objekt eigener Planungen machen. Sie enthalten implizit eine Absage an die Vorstellung linearer Entwicklung des Menschen und seiner Segmentierbarkeit etwa in Teilleistungen. Schließlich unterstützen sie stark die Tendenz, den Fokus abzuwenden von der distanzierten Feststellung von Problemen, Defiziten, Mängeln, Defekten, Verzögerungen, Störungen usw. eines Menschen und sich hinzuwenden zur gemeinsamen Reflexion seiner Erfahrungen, Stärken, Bedürfnisse, Vorlieben und Begabungen. Sie öffnen sich den Möglichkeiten, den Chancen und den Erfordernissen, die für die individuelle Entwicklung nützlich sind. Diese Verfahren führen zu keinem Label, keiner Kategorie, keinem Status, keinem Etikett, keinem sonderpädagogischen Förderbedarf; sie führen zu einer größeren Nähe zu Lebens- und Lernsituationen eines Menschen und zu gemeinsamen Bemühungen um seine Perspektiven innerhalb der Gesellschaft.

Die curriculumbasierte Messung als änderungssensibles Messverfahren könnte eine angemessene Frühdiagnostik mit parallel stattfindender Intervention zur Prävention einer Lese-Rechtschreibproblematik in der Schule durch die Lernfortschrittsdiagnostik und einer dadurch rechtzeitigen Anpassung der Fördermaßnahme darstellen. Es wird zukünftig zu erforschen sein, welche Prädiktoren sich in einer curriculumbasierten Messung operationalisieren lassen, die den Schriftspracherwerb zuverlässig vorhersagen und zudem den klassischen Gütekriterien sowie auch der Handhabbarkeit und Nützlichkeit genügen, und zugleich einen robusten dynamischen Indikator darstellen, wie es bereits für die Leseleistung durch Walter (2010a, 2010b, 2011a, 2011b) im deutschsprachigen Raum gezeigt werden konnte.

Da dieses Messverfahren zur Prozessdiagnostik noch nicht existiert und die entsprechenden möglichen Prädiktoren dazu noch erforscht werden müssen, bleibt aufgrund der begrenzten Vorhersagekraft der Vorschulscreenings und der mangelhaften Ökonomie der Verfahren zu empfehlen, alle Kinder universell präventiv im Kindergarten mit einer vorschulischen Maßnahme zu unterstützen und eines der auf dem Markt existierenden evaluierten Förderprogramme einzusetzen, um einem schulischen Versagen im Schriftspracherwerb vorzubeugen und die Kinder adäquat vorzubereiten. Damit wird verhindert, dass durch ein Screening-Verfahren identifizierte falsch-negative Kinder nicht entsprechend gefördert werden (vgl. auch Lenhard & Schneider 2010, S. 20). Dies ist in den meisten Kindergärten bereits etabliert und wird oftmals durch ein phonologisch gestütztes Erstlesewerk in der ersten Klasse weitergeführt (vgl. auch die Forderungen von Walter 2005b, S. 63ff.). Ergebnisse des Trainings der phonologischen Bewusstheit konnten bisher gute Erfolge erzielen, trotzdem sollte, wie Walter (a. a. O., S. 48) konstatiert und es auch die Ergebnisse dieser Untersuchung zeigen, auf eine „Glorifizierung“ der phonologischen Bewusstheit als „Allheilmittel“ verzichtet werden. Aufgrund aktueller Ergebnisse zum induktiven Denken (vgl. Marx & Keller 2010, S. 139ff.; Sonntag 2010, S. 5ff.) bleibt gleichwohl zu erforschen, inwieweit ein Training des induktiven Denkens in Kombination mit der phonologischen Bewusstheit von Vorteil ist. Es bedarf dafür einer Prozessbegleitung für die Vorläuferfertigkeiten des Lese- und Rechtschreibprozesses, deren wissenschaftliche Erforschung noch aussteht.

Die vorliegende empirische Untersuchung dokumentiert, dass die phonologische Informationsverarbeitung als schriftsprachliche Vorläuferfertigkeit nicht ausschließlich für eine stabile Prognose geeignet ist. Zur künftigen Prävention und Reduktion von Schwächen im Bereich des Schriftspracherwerbs ist es notwendig weitere wissenschaftliche Untersuchungen zur curriculumbasierten Messung im Vorschulbereich zu forcieren.

X. Literatur

Adams, M. J. (1994). *Beginning to read: Thinking and learning about print.* Cambridge, M. A. (new edition): MIT Press.

Allor, J. H., Gansle, K. A. & Denny, R. K. (2006). The Stop and Go Phonemic Awareness Game: Providing modeling, practice, and feedback. *Preventing School Failure, 50 (4*), 23 – 30.

Alloway, T. P., Gathercole, S. E., Adams, A.-M., Willis, C. Eaglen, R. & Lamont, E. (2005). Working Memory and phonological awareness as predictors towards early learning goals at school entry. *British Journal of Developmental Psychology, 23,* 417-426.

Alloway, T. P., Gathercole, S. E., Willis, C. & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology, 87,* 85 – 106.

Amelang, M. & Zielinski, W. (1994). *Psychologische Diagnostik und Intervention.* Berlin-Heidelberg: Springer-Verlag.

American Psychiatric Association (Ed.). (2000). *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders* (4. ed.). Arlington, VA: American Psychiatric Association.

Angermaier, M. (1974). *Begabungsmängel bei Legasthenikern*. Fernstudienlehrgang Legasthenie, Studienbegleitheft 2, 27-46.

Ardoin, S. P. & Christ, T. J. (2008). Evaluating Curriculum-Based Measurement Slope estimates using Data from Triannual Universal Screenings. *School Psychology Review, 37 (1)*, 109 – 125.

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., Slider, N. J. & Williams, K. L. (2004). Examining the Incremental Benefits of Administering a Maze and Three Versus One Curriculum-Based Measurement Reading Probes When Conducting Universal Screening. *School Psychology Review, 33 (2),* 218-233.

Arnold, K.-H. & Hartig, J. (2008). Testtheoretische Grundlagen der Förderdiagnostik. In K.-H. Arnold, O. Graumann & A. Rakhkochkine (Hrsg.): *Handbuch Förderung. Beratung, Diagnostik, Förderbereiche, Förderplanung, Kompetenzentwicklung*. Weinheim & Basel: Beltz.

Aro, M. & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics, 24*, 621 – 635.

Aktinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Ed.): *The psychology of learning and motivation* (Vol. 2, p. 89-155). New York: Academic Press.

Atteslander, P. (2000). *Methoden der empirischen Sozialforschung*. Berlin (9., neu bearb. und erw. Aufl.): de Gruyter.

Aye, A. v. (1982). Statistische und methodologische Problemstellungen psychologischer Präventionsforschung. In J. Brandtstädter & A. v. Eye (Hrsg.): *Psychologische Prävention. Grundlagen, Programme, Methoden* (S. 305-440). Bern: Huber.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2006). *Multivariate Analysemethoden* (11. Aufl.). Berlin: Springer.

Baddeley, A. (1986). *Working Memory.* Oxford Psychology Series No.11. Oxford (u.a.) (4. ed.): Oxford University Press.

Baddeley, A. D. (1990). *Human memory: Theory and practice.* Needham Heights, MA: Allyn & Bacon.

Baddeley, A. D. (1992). Working Memory. *Science, 255* (5044), 556-559.

Baddeley, A. (2003): Working memory and language: An overview. *Journal of Communication Disorder, 36,* 189 – 208.

Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G. Bowers (Ed.): *The psychology of learning and motivation* (p. 47-90). New York: Academic Press

Barth, K. (1997). Was Hänschen nicht lernt… . Lernstörungen schon im Kindergarten erkennen. *Kindergarten heute, 27* (3), 6-13.

Barth, K. (2006). Früherkennung und Prävention von Lernstörungen. Möglichkeiten und Grenzen diagnostischer und präventiver Ansätze in der Früherkennung von Lernstörungen. *Frühförderung interdisziplinär, 25*, 169-186.

Barth, K. & Gomm, B. (2004). *Teststatistik.* URL: http://www.reinhardt-verlag.de/pdf/material01716.pdf (Stand: 29.01.2013). München: Reinhardt.

Barth, K. & Gomm, B. (2005). *Gruppentest zur Früherkennung von Lese- und Rechtschreibschwierigkeiten. Phonologische Bewusstheit bei Kindergartenkindern und Schulanfängern* (PB_LRS). Manual. München: Reinhardt.

Barth, K. & Gomm, B. (2008a). *Gruppentest zur Früherkennung von Lese- und Rechtschreibschwierigkeiten. Phonologische Bewusstheit bei Kindergartenkindern und Schulanfängern* (PB_LRS). Manual. München (2. durchges. Aufl.): Reinhardt.

Barth, K. & Gomm, B. (2008b). Gruppentest zur Früherkennung von Lese- und Rechtschreibschwierigkeiten. Phonologische Bewusstheit bei Kindergartenkindern und Schulanfängern (PB-LRS). In W. Schneider, H. Marx & M. Hasselhorn (Hrsg.): *Diagnostik von Rechtschreibleistungen und -kompetenz*. Tests und Trends N. F. Band 6. Göttingen (u.a.): Hogrefe.

Bäuerlein, K. Niklas, F. & Schneider, W. (2014). Fähigkeitsindikatoren Primarschule (FIPS) – Überprüfung des Lernerfolgs in der ersten Klasse. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik* (Tests & Trends, NF Bd. 12, S. 127-144). Göttingen: Hogrefe.

Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg., 2001). PISA 2000: *Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich.* Opladen: Leske + Budrich.

Baurmann, J. (1996). Aspekte der Aneignung von Schriftlichkeit und deren Reflektion. In Günther, H. & Ludwig, O. (Hrsg.): *Schrift und Schriftlichkeit – Writing and its use.* (S. 1118-1129) Berlin, New York: de Gruyter.

Becker, J. (2004). Computergestütztes Adaptives Testen (CAT) von Angst entwickelt auf der Grundlage der Item Response Theorie (IRT). URL: http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000001495 (Stand: 05.08.2016).

Bee-Götsche, P. H. (1993). Effekte einer Förderung des Kurzzeitgedächtnisses auf die Entwicklung phonemischer Bewußtheit im Kindergartenalter. *Psychologie in Erziehung und Unterricht, 40*, 182-190.

Bender, R. (2001). Interpretation von Effizienzmaßen der Vierfeldertafel für Diagnostik und Behandlung. *Medizinische Klinik, 96(2),* 116-121.

Bergk, M. (1980). *Leselernprozeß und Erstlesewerke. Analyse des Schriftspracherwerbs und seiner Behinderungen mit Kategorien der Aneignungstherapie*. Bochum: Kamp.

Best, H. & Wolf, C. (2010). Logistische Regression. In Wolf, C. & Best, H.: *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 827-854). Wiesbaden: VR Verlag für Sozialwissenschaften.

Betz, D. & Breuninger, H. (1987). *Teufelskreis Lernstörungen: Theoretische Grundlegung und Standardprogramm* (Materialien für die psychosoziale Praxis). Weinheim (2. überarb. Aufl.): Psychologie-Verlags-Union.

Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *American Speech-Language-Hearing Association, 38*,446-462.

Birkel, P. (2007a). *Weingartener Grundwortschatz Rechtschreibtest für erste und zweite Klassen (WRT1+).* 2. Auflage, Göttingen: Hogrefe.

Birkel, P. (2007b). *Weingartener Grundwortschatz Rechtschreibtest für zweite und dritte Klassen (WRT2+)*. 2. Auflage, Göttingen: Hogrefe.

Birkel, P. (2007c). *Weingartener Grundwortschatz Rechtschreibtest für dritte und vierte Klassen (WRT3+). 2*. Auflage, Göttingen: Hogrefe.

Birkel, P. (2007d). *Weingartener Grundwortschatz Rechtschreibtest für vierte und fünfte Klassen (WRT4+).* 2. Auflage, Göttingen: Hogrefe.

Birkel, P. & Stammet, C. (2008). Die Entwicklung der Rechtschreibfähigkeit von Schülern der Grund- und Hauptschule aus der Sicht einer Neueichung des WRT. In W. Schneider, H. Marx & M. Hasselhorn (Hrsg.): *Diagnostik von Rechtschreibleistungen und -kompetenz. Tests und Trends* N. F. (S. 61-93), Band 6. Göttingen (u.a.): Hogrefe.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.): *Statistical theories of mental test scores*. Reading. MA: Addison-Wesley.

Bishop, D. V. M. & Edmundson, A. (1987a). Language-impaired 4-year-olds: distinguishing transient form persistent impairment. *Journal of Speech and Hearing Disorders, 52*, 156-173.

Bishop, D. V. M. & Edmundson, A. (1987b). Specific language impairment as a maturational lag: Evidence from longitudinal data on language and motor development. *Developmental Medicine and Child Neurology, 29*, 442-459.

Bjorklund, D. F. & Schneider, W. (2006). Ursprung, Veränderung und Stabilität der Intelligenz im Kindesalter: Entwicklungspsychologische Perspektiven. In W. Schneider (Hrsg.): *Enzyklopädie der Psychologie: Themenbereich C, Theorie und Forschung*. Serie V, Entwicklungspsychologie 2. Kognitive Entwicklung (S. 769-821). Göttingen: Hogrefe.

Blaser, R., Preuss, U. & Felder, W. (2010). Evaluation einer vorschulischen Förderung der phonologischen Bewusstheit und der Buchstaben-Laut Korrespondenz. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 38* (3), 181-188.

Blaser, R., Preuss, U., Groner, M., Groner, R. & Felder, W. (2007). Kurz-, mittel- und längerfristige Effekte eines Trainings in phonologischer Bewusstheit und in Buchstaben-Laut-Korrespondenz auf die phonologische Bewusstheit und die Lese- und Rechtschreibleistung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 35 (4)*, S. 273-280.

Bloom, B. S., Hastings, J. T. & Madaus, G. F. (Ed.) (1971). *Handbook of formative and summative evaluation of student learning.* New York, NY: McGraw-Hill.

Blumenstock, L. (1979). *Prophylaxe der Lese- und Rechtschreibschwäche. Erstellung und Erprobung eines sprachlich-akustischen Trainingsverfahrens zur Verbesserung der grundlegenden Lese- und Rechtschreibfähigkeit der Schüler im 1. Schuljahr*. Mannheim: Beltz.

Blumenthal, Y., Kuhlmann, K. & Hartke, B. (2014). Diagnostik und Prävention von Lernschwierigkeiten im Aptitude Treatment Interaction – (ATI-) und Response to Intervention (RTI-) Ansatz. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik (Tests und Trends N.F. Band 12)* (S. 61-82). Göttingen: Hogrefe.

Börnert, M. (2014). *Lernverlaufsdiagnostik. Definition, Einsatzbereiche und Perspektiven für die pädagogische Praxis.* Potsdamer Zentrum für empirische Inklusionsforschung (ZEIF), 2. URL: http://www.uni-potsdam.de/fileadmin01/projects/inklusion/PDFs/ZEIF-Blog/B%C3%B6rnert_2014_Lernverlauf.pdf (Stand: 08.08.2016).

Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Heidelberg: Springer (4. Aufl.).

Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer (7. Aufl.).

Bos, W., Lankes, E. M., Prenzel, M., Schwippert, K., Valtin, R. & Walther, G. (2003): Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Zusammenfassung ausgewählter Ergebnisse. In W. Bos, E. M., Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther: *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann. URL: http://www.bmbf.de/pubRD/erste_ergebnisse_aus_iglu-zusammenfassung.pdf (Stand: 01.01.2013).

Bradley, L. & Bryant P. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature, 271*, 746-747.

Bradley, L. & Bryant P. (1983). Categorising sounds and learning to read – a causal connection. *Nature, 301*, 419-421.

Bradley, L. & Bryant P. (1985). *Rhyme and reason in reading and spelling*. Michigan: The University of Michigan Press.

Bradley, L. & Bryant P. (1991). Phonological skills before and after learning to read. In S. A. Brady & D. Shankweiler (Ed.): *Phonological processes in literacy. A tribute to Isabelle Liberman* (p. 37-45). Hillsdale, NJ: Erlbaum.

Breitenbach, E. & Lenhard, W. (2001). Aktuelle Forschung auf der Suche nach neurobiologischen Korrelaten der Lese-Rechtschreib-Störung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 29 (3),* 167-177.

Breuer, H. (1998). Früherkennung und Frühförderung von Schreib-Lese-Lernvoraussetzungen. *Sprache, Stimme, Gehör, 22*, 34-38.

Breuer, H. & Weuffen, M. (1993). *Lernschwierigkeiten am Schulanfang. Schuleingangsdiagnostik zur Früherkennung und Frühförderung*. Weinheim: Beltz.

Breuer, H. & Weuffen, M. (2006). *Lernschwierigkeiten am Schulanfang*. Weinheim (7. akt. Aufl.): Beltz.

Brosius, F. (2013). *SPSS 21*. Zwickau: Westermann.

Bruck, M. (1988). The word recognition and spelling of dyslexic children. *Reading Research Quarterly, XXIII/I,* 51-69.

Brügelmann, H. (2003a). *Vorschulische Prädiktoren des Misserfolgs beim Schriftspracherwerb in der Schule*. Abschlussbericht des Projekts LOGIK-R an das MWF/Düsseldorf. Vervielf. Ms. Arbeitsgruppe Primarstufe/ FB2 der Universität: Siegen. URL: www.grundschulverband.de (Stand: 15.05.2009).

Brügelmann, H. (2003b). Fördern: nur das Reparieren einer unzulänglichen Technik? Zur Prävention von Lese-Rechtschreib-Schwierigkeiten durch phonologische Trainings. *Grundschule, 10*, 54-57.

Brügelmann, H. (2005). Das Prognoserisiko von Risikoprognosen – eine Chance für „Risikokinder"? In Hofmann, B. & Sasse, A. (Hrsg.): Übergänge. *Kinder und Schrift zwischen Kindergarten und Schule* (S. 146-172). Berlin.

Brügelmann, H. (1984). Lesen- und Schreibenlernen als Denkentwicklung. Voraussetzungen eines erfolgreichen Schrifterwerbs. *Zeitschrift für Pädagogik, 30*, 69-91.

Brügelmann, H. & Heymann, H. W. (2002). *PISA 2000: Befunde, Deutungen, Folgerungen.* Langfassung des Artikels aus Pädagogik 54, Heft 3, 40-43. URL: http://www2.agprim.uni-siegen.de/printbrue/brue-pisa.02r.pdf (Stand: 21.02.2013).

Brunner, M. & Schöler, H. (2008). HASE. *Heidelberger Auditives Screening in der Einschulungsuntersuchung.* (2. überarb. und erw. Auflage). Binswangen: WESTRA.

Bryant, P. & Goswami, U. (1987). Phonological awareness and learning to read. In J. Beech & A. Colley (Ed.). *Cognitive approaches to reading* (p. 213-243). Chichester: Wiley.

Bryant, P., Maclean, Crossland (1989). Nursery rhymes, phonological skills and reading. *Journal of Child Language, 16,* 407-428.

Bühl, A. (2008): *SPSS Version 16. Einführung in die moderne Datenanalyse*. München (11. aktualisierte Aufl.): Pearson Studium.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.

Bundschuh, K., Heimlich, U. & Krawitz, R. (2002). *Wörterbuch der Heilpädagogik*. Bad Heilbrunn (2. durchges. Aufl.): Klinkhardt.

Bus, A. G. & Ijzendoorn, M. H. van (1999). Phonological awareness and early reading: a meta-analysis of experimental training studies. *Journal of Educational Psychology, 91* (3), 403-414.

Byrne, B. & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology, 81*, 313-321.

Byrne, B. & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83*, 451-455.

Byrne, B. & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*, 104-111

Byrne, B. & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*, 488-503.

Byrne, B., Fielding-Barnsley, R. & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate of response. *Journal of Educational Psychology, 92*, 659-667.

Carroll, J. M., Snowling, M. J.; Hulme, C. & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39*, 913-923.

Castles, A. & Coltheart, M. (2004): Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*, 77 – 111.

Castles, A. McLean, G. M. T. & McArthur, G. (2010). Dyslexia (neuropsychological). *Cognitive Science, 1*, 426-432.

Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S. & Ingvar (1998). The illiterate brain. Learning to read and write during childhood influences the functional organization of the adult brain. *Brain, 121*, 1053-1063).

Catell, Weiß & Osterland (1997). *Grundintelligenztest Skala 1. CFT 1.* Göttingen (5. Rev. Auflage): Hogrefe.

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36(5*), 948-958.

Catts, H. W. & Hogan, T. P. (2003). Language basis of reading disabilities and implications for early identification and remediation. *Reading Psychology, 24*, 223-246.

Catts, H. (2006). *School-wide screening. Presentation at the National SEA Conference on Responsiveness to Intervention: Integrating RTI within the SLD Determination Process*, Kansas City, MO. URL: http:// www.nrcld.org/sea/presentations_worksheets/ screening/Catts_screening.pdf (Stand: 13.01.2013)

Caughy, M.O. (1996). Health and environmental effects on the academic readiness of schoolage children. *Developmental Psychology, 32*, 515-522.

Caylak, E. (2007). A Review of Association and Linkage Studies for Genetical Analyses of Learning Disorders. *American Journal of Medical Genetics Part B, 144B*, 923-943.

Christ, T. J. (2006). Short-Term Estimates of Growth Using Curriculum-Based Measurement of Oral Reading Fluency: Estimating Standard Error of the Slope to Construct Confidence Intervals. *School Psychology Review, 35 (1)*, 128-133.

Christiansen, C. (2007). *Wuppis Abenteuerreise. Übungsprogramm zur Förderung der phonologischen Bewusstheit.* Oberursel: Finken.

Clauß & Ebner (1968). *Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen*. Berlin: Volk und Wissen Volkseigener Verlag.

Clauß, G. & Ebner (1982). *Statistik. Für Soziologen, Pädagogen, Psychologen und Mediziner*. Band 1 Grundlagen. Frankfurt: Deutsch.

Clauß, G., Finze, F.-R. & Partzsch, L. (2004). *Statistik Für Soziologen, Pädagogen, Psychologen und Mediziner.* Frankfurt am Main: Verlag Harri Deutsch.

Clay, Marie M. (1985). *The Early Detection of Reading Difficulties.* Portsmouth, NH (third edition): Heinemann.

Clemens, N. H., Hagan-Burke, S., Luo, W., Cerda, C., Blakely, A., Frosch, J., Gamez-Patience, B. & Jones, M. (2015). The Predictive Validity of a Computer-Adaptive Assessment of Kindergarten and First-Grade Reading Skills. *School Psychology Review, 44 (1),* 76-97.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 122(1), 155-159.

Colòn, E. P. & Kranzler, J. H. (2006). Effect of Instructions on Curriculum-Based Measurement of Reading. *Journal of Psychoeducational Assessment, 24 (4),* 318-328.

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of reading aloud: Dual-route and parallel distributed-processing approaches. *Psychological Review, 100*, 589-608.

Coltheart, M. & Rastle, K. (1994). Serial Processing in Reading Aloud: Evidence for Dual-Route Models of Reading. *Journal of Experimental Psychology: Human Perception and Performance, 20 (6*), 1197 – 1211.

Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology 78 (1), 98-104.

Coulmas, F. (1982). *Über Schrift*. Frankfurt am Main: Suhrkamp.

Cratzius, M. (2003). *Die Einbindung des Absatz- und Produktionsbereichs in Innovationsprozesse.* Wiesbaden: Deutscher Universitätsverlag.

Crämer, C. & Schumann, G. (2002). Schriftsprache. In S. Baumgartner & I. Füssenich (Hrsg.): *Sprachtherapie mit Kindern* (S. 256-336). München (5. Aufl.): Ernst Reinhardt.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Cutting, L. E. & Denckla, M. B. (2001): The relationship of rapid serial naming and wordreading in normally developing readers: An exploratory model. *Reading and Writing: An Interdisciplinary Journal, 14,* 673 – 705.

Danielson, L., Doolittle, J. & Bradley, R. (2007). Professional development, capacity building, and research needs: Critical Issues for Response to Intervention Implementation. *School Psychology Review, 36 (4*), 632 – 637.

Dehn, M. (1994). *Schlüsselszenen zum Schriftspracherwerb. Arbeitsbuch zum Lese- und Schreibunterricht in der Grundschule.* Weinheim, Basel: Beltz.

Deimel, W., Ziegler, A. & Schulte-Körne, G. (2005). *Modell Schriftsprach-Moderatoren (MSM). Zwischenbericht der wissenschaftlichen Begleitung nach zwei Jahren* – Kurzfassung. URL: http://www.kjp.med.uni-muenchen.de/download/MSM-ZwischenberichtKurzfassung2.pdf (Stand: 05.10.2010).

Denckla, M. B. & Rudel, R. (1974). Rapid ‚automized' naming of pictured objects, colors, letters and numbers by normal children. *Cortex, 10*, 168-202.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52 (3),* 219-232.

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.). *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York: Guilford Press.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*, 184-192.

Deno, S. & Fuchs, D. (1987). Developing curriculum-based measurement systems for data based special education problem solving. *Focus on Exceptional Children, 18 (8),* 1-16.

Deno, S. L., Fuchs, L. S., Marston, D. & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*, 507–526.

Deno, S., Lembke, E., & Reschly-Anderson, A. (2007). *Progress Monitoring: Student Group Content Module.* URL: http://www.nysrti.org/docs/Progress%20Monitoring%20Study%20Group%20Content%20Module.pdf (Stand: 30.01.2013).

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36–45.

Deno, S. L*.,* Reschly-Anderson, A*.,* Lembke, E*.,* Zorka, H. *&* Callender, S. *(*2002*). A model for school wide implementation: A case example.* Paper presented at the annual meeting of the National Association of School Psychology*,* Chicago, IL*.*

Diaz-Bone, R. (2013). *Statistik für Soziologen.* Konstanz: UTB GmbH (2. Auflage).

Diehl, K. (2014). Lesekompetenzen formativ evaluieren mit dem IEL-1 – Inventar zur Erfassung der Lesekompetenzen von Erstklässlern. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik* (Tests & Trends, NF Bd. 12, S. 83-99). Göttingen: Hogrefe.

Diehl, K. & Hartke, B. (2007). Curriculumnahe Lernfortschrittsmessungen. *Sonderpädagogik, 37 (4),* 195-211.

Diehl, K. & Hartke, B. (2011). Zur Reliabilität und Validität des formativen Bewertungssystems IEL-1. Inventar zur Erfassung der Lesekompetenz von Erstklässlern. Empirische Sonderpädagogik, 3, 121-146.

Diehl, K. & Hartke, B. (2012). IEL-1. Inventar zur Erfassung der Lesekompetenz im 1. Schuljahr. Göttingen: Hogrefe.

Diekmann, A. (2010). Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. Reinbek (vollst. überarb. 20. Aufl.): Rowohlt.

Dilling, H. Mombour, W. & Schmidt, M. H. (1994). *Internationale Klassifikation psychischer Störungen ICD-10* Kapitel V (F). Forschungskriterien. Bern: Huber.

Döbert, M. & Hubertus, P. (2000). *Ihr Kreuz ist die Schrift. Analphabetismus und Analphabetisierung in Deutschland*. Münster/Stuttgart: Bundesverband Alphabetisierung e. V. (Hrsg.). URL: www.alphabetisierung.de/fileadmin/files/Dateien/Downloads_Texte/Ihr Kreuz-gesamt.pdf (Stand: 25.10.2010).

Dummer, L. (1989). Späte Früherfassung – eine sinnvolle Alternative. In H. Skowronek (Hrsg.): *Schwierigkeiten beim Lesen- und Schreibenlernen* (S. 27-39). Universität Bielefeld: SFB 227, Preprint. Nr. 21.

Eichler, W. (1986). Zu Uta Frith's Dreiphasenmodell des Lesens und (Schreiben)lernens. Oder: Lassen sich verschiedene Modelle des Schriftspracherwerbs aufeinander beziehen und weiterentwickeln? In G. Augst (Hrsg.): *New Trends in Graphemics and Orthography* (S. 234-247). Berlin: de Gruyter.

Ehri, L. C. (1979). Linguistic insight: Threshold of reading acquisition. In T. G. Waller & G. E. MacKinnon (Ed.). *Reading Research: Advances in Theory and Practice* (Band 1, p. 63-114). New York: Academic Press.

Ehri, L.C. (1986). Sources of difficulty in learning to spell and read. In M. Wolraich & D.K. Routh (Eds.): *Advances in developmental and behavioral pediatrics, Vol. 7* (p. 121-195). Greenwich: JA/Press Inc.

Ehri, L. C. (1987). Learning to read and spell words. *Journal of Reading Behavior, 19*, 5-31.

Ehri, L.C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18,* 116-125.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z. & Shanahan, T. (2001). Phoemic awareness instruction helps children learn to read: Evidence from the National Reading panels's meta-analysis. *Reading Research Quarterly, 36*, 250-287.

Einsiedler, W. (2003). *Diagnose und Training phonologischer Bewusstheit*. Grundschule 35(9), S. 54.

Einsiedler, W. & Kirschhock, E.-M. (2003). Forschungsergebnisse zur phonologischen Bewusstheit. *Grundschule, 9,* 55-57.

Einsiedler, W., Frank, A., Kirschhock, E.-M., Martschinke, S. & Treinies, G. (2002). Der Einfluss verschiedener Unterrichtsmethoden auf die phonologische Bewusstheit sowie auf Lese- und Rechtschreibleistungen im 1. Schuljahr. *Psychologie in Erziehung und Unterricht, 49*, 194-209.

Elbro, C. (1998). Reading-listening discrepancy definition of dyslexia. In P. Reitsma & L. Verhoeven (Eds.): *Problems and Interventions in Literacy Development* (p. 129-146). Dordrecht: Kluwer.

Elliott, J., Lee, S.W. & Tollefson, N. (2001). A reliability and validity study oft he dynamix indicators of basic early literacy skills – modified. *School Psychology Review, 30 (1*), 33-49.

Ennemoser, M., Marx, P., Weber, J. & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens. Evidenz aus zwei Längsschnittstudien vom Kindergarten bis zur 4. Klasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 44* (2), 53-67.

Enzmann, D. (2012). *Multiple Regression.* Universität Hamburg, Fachbereich Sozialwissenschaften. URL: http://www2.jura.uni-hamburg.de/instkrim/kriminologie/Mitarbeiter/Enzmann/Lehre/StatIKrim/ (Stand: 23.12.2012)

Esser, G. & Schmidt, M. (1993). Die langfristige Entwicklung von Kindern mit Lese-Rechtschreibschwäche. *Zeitschrift für Klinische Psychologie, 22*, 100-116.

Esser, G., Wyschkon, A,& Schmidt, M.H. (2002). Was wird aus Achtjährigen mit einer Lese- und Rechtschreibstörung? Ergebnisse im Alter von 25 Jahren. *Zeitschrift für Klinische Psychologie und Psychotherapie, 31*, (4), 235-242.

Evans, B.J.C. (2004). Visual factors in dyslexia. In M. Turner & J. Rack (Ed.): *The study of dyslexia* (p. 1-22). New York: Kluwer academic.

Falkai, P. & Wittchen, H.-U., Döpfner, M., Gaebel, W., Maier, W., Rief, W., Saß, H. & Zaudig, M. (Hrsg.) (2015). Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5 ®. Göttingen: Hogrefe.

Faust, G. (2008). Übergänge gestalten – Übergänge bewältigen. In W. Thole, H.-G. Roßbach, M. Fölling-Albers & R. Tippelt (Hrsg.), *Bildung und Kindheit. Pädagogik der Frühen Kindheit in Wissenschaft und Lehre* (S. 225-240). Opladen & Farmington Hills: Barbara Budrich.

Fay, E. & Stumpf, H. (1999). Leistungsdaten. In Jäger, R.S. & Petermann, F. (Hrsg.): *Psychologische Diagnostik: Ein Lehrbuch* (S. 380 – 396). Weinheim (4. Aufl.): Beltz.

Felton, R. H. & Wood, F. R. (1992). A reading level match study of nonword reading skills in poor readers with varying IQs. *Journal of Learning Disabilities, 25*, 318-326.

Feilke, H. (2001). Über sprachdidaktische Grenzen: Von „Erfindern", „Entdeckern" und „Mentoren". *Didaktik Deutsch, 10*, 4-25.

Ferdinand, W. (1972). Über die Erfolge des ganzheitlichen und synthetischen Lese- (Schreib-) Unterrichts in der Grundschule. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 4* , 105-112.

Fischer, M. Y. & Pfost, M. (2015). Wie effektiv sind Maßnahmen zur Förderung phonologischen Bewusstheit? Eine meta-analytische Untersuchung der Auswirkungen deutschsprachiger Trainingsprogramme auf den Schriftspracherwerb. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 47 (1),* 35-51.

Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe (3. überarb. und erw. Aufl.).

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., Francis, D. J., Fowler, A. E. & Shaywitz, B. A. (1994). Cognitive profiles of reading disability. Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*, 6-23.

Foegen, A., Jiban, C. & Deno, S. (2007). Progress Monitoring Measures in Mathematics. A Review of the Literature. *The Journal of Special Education, 41 (2),* 121-139.

Forster, M. & Martschinke, S. (2008). *Leichter lesen und schreiben mit der Hexe Susi*. Donauwörth (Neuauflage): Auer Verlag.

Foster, G., Schmidt, C. & Sabatino, D. (1976). Teachers' expectancies and the label „learning disabilities". *Journal of Learning Disabilities, 9*, 58-61.

Fowler, A. & Scarborough, H. (1993). *Should reading-disabled adults be distinguished from other adults seeking literacy instruction?* A review of theory and research. NCAL Publication TR 93-7. Philadelphia, PA: National Center on Adult Literacy.

Fox, B., Routh, D. K. (1975). Analyzing spoken language into words, syllables, and phonemes: A developmental study. *Journal of Psycholinguistic Research, 4*, 331-342.

Frank, A., Kirschhock, E.-M. & Martschinke, S. (2003). Der Rundgang durch Hörhausen. *Grundschule, 9*, 58-59.

Franzkowiak, T. (2008). *Vom BLISS-Symbol zur alphabetischen Schrift. Entwicklung und Erprobung eines vorschulischen Förderansatzes zur Prävention von Lernschwierigkeiten vom Schriftspracherwerb*. Dissertation in FB 2. Universität: Siegen. URL: http://dokumentix.ub.uni-siegen.de/opus/volltexte/2008/351/

Frith, U. (1980). *Cognitive processes in spelling*. London: Academic Press.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall & M. Coultheart (Ed.). *Surface dyslexia* (p. 301-330). London: Erlbaum.

Frith, U. (1999). Paradoxes in the definition of dyslexia. *Dyslexia, 5,* 192-214.

Fröhlich, L. P., Metz, D. & Petermann, F. (2009). Kindergartenbasierte Förderung der phonologischen Bewusstheit „Lobo vom Globo". *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 18 (4*), 204-212.

Frostig, M. (1964). The Marianne Frostig Developmental Perception. *Perceputal and Motor Skills, 19,* 464-499.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33 (2),* 188-192.

Fuchs, D. & Deno, S.L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58*, 232-243.

Fuchs, L. S. & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61*, 15–24.

Fuchs, L. S., Deno, S. L. & Mirkin, P. K. (1984). The Effects of Frequent Curriculum-Based Measurement and Evaluation on Pedagogy, Student Achievement, and Student Awareness of Learning. *American Educational Research Journal, 21 (2)*, 449-460.

Fuchs, L. S., & Fuchs, D. (1986). The effects of systematic formative evaluation: a meta-analysis. *Exceptional Children, 53,* 199-208.

Fuchs, L. S. & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 58*, 45–58.

Fuchs, L. S. & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 13*, 204–219.

Fuchs, L. S. & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28,* 659–671.

Fuchs, L. S. & Fuchs, D. (2000). Curriculum-based measurement and performance assessment. In E. S. Shapiro & T. R. Kratochwill (Hrsg.): *Behavioral assessment in schools: Theory, research, and clinical foundations* (S. 168–201). New York (2. Aufl.): Guilford.

Fuchs, L. S. & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders. *Peabody Journal of Education, 77 (2),* 64–84.

Fuchs, L. S. & Fuchs, D. (2004). Determining Adequate Yearly Progress from Kindergarten Through Grade 6 with Curriculum-Based Measurement. *Assessment for Effective Intervention, 29 (4),* 25-37.

Fuchs, L. S., Fuchs, D. & Hamlett, C. L. (1989). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10*, 43–52.

Fuchs, L. S., Fuchs, D. & Hamlett, C. L. (1993). Technological advances linking the assessment of students' academic proficiency to instructional planning. *Journal of Special Education Technology, 12*, 49–62.

Fuchs, L. S., Fuchs, D., Hamlett, C. L. & Hasselbring, T. S. (1987). Using computers with curriculum-based monitoring: Effects on teacher efficiency and satisfaction. *Journal of Special Education Technology, 8 (4),* 14-27.

Fuchs L. S., Fuchs, D., Hamlett, C. L., Walz, L. & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*, 27-48.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Dutka, S. & Katzaroff, M. (2000). The importance of providing background information on the structure and scoring of performance assessments. *Applied Measurement in Education, 13*, 83–121.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9(2),* 20-29.

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disability Quarterly, 25*, 33-45.

Fuchs, D., Fuchs, L., Thompson, A., Al Otaiba, S., Yen, L., & Yang, N. (2005).*Respective contributions of phonological awareness and decoding to reading development in Title I and non–Title I schools*. Unpublished manuscript.

Gasteiger-Klicpera, B. & Klicpera, C. (2004). Lese-Rechtschreibschwäche. In G.W. Lauth, M. Grünke & J.C. Brunstein (Hrsg.): *Intervention bei Lernstörungen. Förderung, Training und Therapie in der Praxis* (S. 46-54). Göttingen (u.a.): Hogrefe.

Gathercole, S. & Baddeley, A.D. (1993a). *Working Memory and Language*. Hove, Sussex: Lawrence Erlbaum Associates.

Gathercole, S. & Baddeley, A. D. (1993b). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education, 8*, 259-272.

Georgiewa, P., Grünling, C., Ligges, M., Filz, C., Möller, U. & Blanz, B. (2004). Lebensalterspezifische Veränderungen phonologischer Defizite bei Lese-Rechtschreibstörung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 33 (4)*, 281-289.

Girden, E. R. (1992). *ANOVA: Repeated measures*. Sage university papers. Quantitative applications in the social sciences (07-084). Newbury Park, Calif.: Sage Publications.

Goldammer, A. von, Mähler, C., Bockmann, A.-K. & Hasselhorn, M. (2010). Vorhersage früher Schriftsprachleistungen aus vorschulischen Kompetenzen der Sprache und der phonologischen Informationsverarbeitung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42 (1),* 48-56.

Good, R. H., Gruba, J., & Kaminski, R. A. (2001). Best Practices in Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an Outcomes-Driven Model. In A. Thomas & J. Grimes (Eds.): *Best Practices in School* Psychology IV (pp. 679-700). Washington, DC: National Association of School Psychologists.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. URL: http://dibels.uoregon.edu/. (Stand: 19.01.2013).

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance of decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high stakes outcomes. *Scientific Studies of Reading, 5(3)*, 257-288.

Goswami, U. (1988). Orthographic analogies and reading development. *Quarterly Journal of Experimental Psychology, 40A*, 239-268.

Goswami, U. (1993). Toward an interactive model of reading development: Decoding vowel graphemes in beginning reading. *Journal of Experimental Child Psychology, 56*, 443-475.

Goswami, U. (2001). The ‚Phonological Representations' hypothesis in dyslexia. In G. Schulte-Körne (Hrsg.). *Legasthenie: erkennen, verstehen, fördern* (S. 66-74). Bochum: Winkler.

Gough, P. B., Hoover, W. A. & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Hrsg.): *Reading comprehension difficulties: Processes and intervention* (S. 1-13). Mahwah, NJ: Erlbaum.

Gough, P. B. & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Graf, E. (1994). *Lese-Rechtschreib-Schwäche. Ein prozessanalytischer Ansatz*. Bern: Lang.

Graney, S. B. & Shinn, M. R. (2005). Effects of Reading Curriculum-Based Measurement (R-CBM) Teacher Feedback in General Education Classrooms. *School Psychology Review, 34 (2),* 184-201.

Grigorenko, E.L. (2001). Developmental dyslexia: An update on genes, brains and environments. *Journal of Child Psychology and Psychiatry, 42*, 91-125.

Grimm, H. (1995). Gestörter Sprachlernprozess: Ursachen und schulische Folgen. In W. Niemeyer (Hrsg.): *Kommunikation und Lese-Rechtschreibschwäche* (S. 53-70). Bochum: Winkler.

Grissemann, H. (1995). Legasthenie heute: Zur Revision des Legastheniekonzepts unter dem Aspekt der Förderdiagnostik. In W. Niemeyer (Hrsg.). *Kommunikation und Lese-Rechtschreibschwäche* (S. 25-52). Bochum: Winkler.

Grissemann, H. & Baumberger, W. (2000). *Zürcher Leseverständnistest für das 4. bis 6. Schuljahr(ZLVT 4-6).* Göttingen, Hogrefe.

Grotlüschen, A. & Riekmann, W. (2011). *leo. - Level-One Studie*. Presseheft. Universität Hamburg, Hamburg. URL: http://blogs.epb.uni-hamburg.de/leo/ (Stand: 21.01.2013).

Grund, M., Haug, G. & Naumann, C. L. (2003a). *Diagnostischer Rechtschreibtest für 4. Klassen (DRT4*). 2. Auflage, Weinheim: Beltz.

Grund, M., Haug, G. & Naumann, C. L. (2003b). *Diagnostischer Rechtschreibtest für 5. Klassen (DRT5)*. 2. Auflage, Weinheim: Beltz.

Günther, K. B. (1986). Ein Stufenmodell der Entwicklung kindlicher Lese- und Schreibstrategien. In H. Brügelmann (Hrsg.): *ABC der Schriftsprache. Rätsel für Kinder, Eltern und Forscher* (S. 32-54). Konstanz: Faude.

Günther, H. (1995). Die Schrift als Modell der Lautsprache. *Osnabrücker Beiträge zur Sprachtheorie, 51*, 15-32.

Hacker, D. (2002). Phonologie. In S. Baumgartner & I. Füssenich (Hrsg.): *Sprachtherapie mit Kindern* (S. 13-62). München: Reinhardt Verlag.

Haffner, J.; Zerahn-Hartung, C.; Pfüller, U.; Parzer, P.; Strehlow, U.; Resch, F. (1998): Auswirkungen und Bedeutung spezifischer Rechtschreibprobleme bei jungen Erwachsenen – empirische Befunde in einer epidemiologischen Stichprobe. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 26,* 124-135.

Hagvet, B. E. (1997). Phonological and lingustic-cognitive precursors of reading abilities. *Dyslexia, 3*, 163-177.

Harrison-Palmer, R. (2012). *The simple view of reading.* URL: https://portfolio.pebblepad.co.uk/cumbria/viewasset.aspx?oid=506154&type=webfolio&pageoid=506166 (Stand: 07.08.2016)

Hartke, B. & Koch, K. (2008). Qualitätsstandards von Prävention und Präventionsforschung. In J. Borchert, B. Hartke & P. Jogschies (Hrsg.): *Frühe Förderung entwicklungsauffälliger Kinder und Jugendlicher*. Stuttgart: Kohlhammer.

Hartmann, E. (2007). Erfolg versprechende Prävention von Leseschwierigkeiten in Kindergarten und Primarschule: ein Überblick. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete, 76*, 114-127.

Hasselhorn, M., Grube, D. & Mähler, C. (2000). Theoretisches Rahmenmodell für ein Diagnostikum zur differentiellen Funktionsanalyse des phonologischen Arbeitsgedächtnisses. In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), *Diagnostik von Lese- und Rechtschreibschwierigkeiten* (S. 167-181). Göttingen: Hogrefe.

Hasselhorn, M., Marx, H. & Schneider, W. (2008). In W. Schneider, H. Marx & M. Hasselhorn (Hrsg.). *Diagnostik von Rechtschreibleistungen und -kompetenz. Tests und Trends* N. F. (S. 1-7) Band 6. Göttingen (u.a.): Hogrefe.

Hasselhorn, M., Schneider, W. & Marx, H. (2000). Früh-, Förder- und Differenzialdiagnostik von Lese-Rechtschreibschwierigkeiten: Eine Einführung. In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.). *Diagnostik von Lese- und Rechtschreibschwierigkeiten. Tests und Trends N. F*. Band 1. Göttingen (u.a.): Hogrefe

Hasselhorn, M., Schumann-Hengsteler, R. Gronauer, J. Grube, D., Mähler, C. Schmid, I., Seitz-Stein, K. & Zoelch, C. (2012). *Arbeitsgedächtnistestbatterie für Kinder von 5 bis 12 Jahren* (AGTB 5-12). Göttingen: Hogrefe.

Hasselhorn, M., Seidler-Brandler, U. & Körner, K. (2000). Ist das „Nachsprechen von Kunstwörtern" für die Entwicklungsdiagnostik des phonologischen Arbeitsgedächtnisses geeignet? In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.). *Diagnostik von Lese- und Rechtschreibschwierigkeiten. Tests und Tr*ends N. F. Band 1. Göttingen (u.a.): Hogrefe.

Hasselhorn, M., Tiffin-Richards, M. C., Woerner, W., Banaschewski, T. & Rothenberger, A. (2000). Spielt der phonetische Speicher des Arbeitsgedächtnisses eine bedeutsame Rolle für die Differentialdiagnose von Lese-Rechtschreib-Schwierigkeiten? In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.). *Diagnostik von Lese- und Rechtschreibschwierigkeiten. Tests und Trends N. F.* Band 1. Göttingen (u.a.): Hogrefe.

Hatcher, P. J., Hulme, C. & Ellis, A. W. (1994): Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65*, 41 – 57.

Hattie, J. (2003). *Teachers make the difference. What is the research evidence?* Auckland: University of Auckland. URL: www.decd.sa.gov.au (Stand: 27.01.2013).

Hatz, H. (2015*). Phonologische Bewusstheit und Schriftspracherwerb. Auswirkungen eines Trainings phonologischer Bewusstheit und eines um Rechtschreibinhalte erweiterten Trainings im ersten Schuljahr auf den Erwerb des Lesens und Rechtschreibens bei Schülerinnen und Schülern mit gering ausgebildeten schriftsprachspezifischen Vorläuferfertigkeiten*. URL: file:///Users/leaschulz/Downloads/Dissertation_Hubertus_Hatz_Februar_2015_Veroeffentlichung_online.pdf (Stand: 20.08.2016).

Hayou-Thomas, M. E., Harlaar, N. Dale, P. S. & Plomin, R. (2006). Genetic and environmental mediation of the prediction from preschool language and nonverbal ability to 7-year reading. *Journal of Research in Reading, 29 (1)*, 50-74.

Hemmerling, A. (2007). *Der Kindergarten als Bildungsinstitution. Hintergründe und Perspektiven*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Hemminger, U., Roth, E., Schneck, S., Jans, T. & Warnke, A. (2000). Testdiagnostische Verfahren zur Überprüfung der Fertigkeiten im Lesen, Rechtschreiben und Rechnen. Eine kritische Übersicht. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 28 (3),* 188-201.

Hesse, I. & Latzko, B. (2009). *Diagnostik für Lehrkräfte*. Opladen & Farmington Hills: Barbara Budrich.

Hinshelwood, (1904). A case of congenital word blindness. *British Medical Journal, 2*, 1303-1304.

Hintze, J.M. & Shapiro, E.S (1997). Curriculum-based measurement and literature-based reading: Is curriculumbased measurement meeting the needs of changing reading curricula? *Journal of School Psychology*, 35, 351-375.

Hippmann, K. (2008). *Prädiktoren des Schriftspracherwerbs im Deutschen*. Dissertationsschrift. Rheinisch-Westfälische Technische Hochschule Aachen. URL: http://darwin.bth.rwth-aachen.de/…/2008/Hippmann_Kathrin.pdf (Stand: 23.01.2013).

Hippmann, K. Jambor-Fahlen, S. & Becker-Mrotzek, M. (2016). Lesen macht stark. Ein Diagnose- und Förderinstrument für die Grundschule. Erste Ergebnisse der Pilotierung für den Bereich der Lesefähigkeit. *Lernen und Lernstörungen, 5 (1),* 58-68.

Höse, A., Wyschkon, A., Moraske, S., Eggeling, M., Quandte, S., Kohn, J., Poltz, N., von Aster, M. und Esser, G. (2016). Prävention von Lese- und Rechtschreibstörungen: Kurz- und mittelfristige Effekt einer Förderung der phonologischen Bewusstheit und der Buchstaben-Laut-Verknüpfung bei Risikokindern im Vorschulalter. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 29,* 1-15.

Hofmann, B. (2006). Gehirnforschung und Lese-Rechtschreibschwierigkeit. In B. Hofmann & A. Sasse (Hrsg.): *Legasthenie. Lese-Rechtschreibstörungen oder Lese-Rechtschreibschwierigkeiten? Theoretische Konzepte und praktische Erfahrungen mit Förderprogrammen* (S.100-118). Frankfurt: Deutsche Gesellschaft für Lesen und Schreiben.

Hofmann, B. & Sasse, A. (2006a). Eine Legasthenie ist doch behebbar! Eine Befragung ehemaliger Legastheniker. In B. Hofmann & A. Sasse (Hrsg.): *Legasthenie. Lese-Rechtschreibstörungen oder Lese-Rechtschreibschwierigkeiten? Theoretische Konzepte und praktische Erfahrungen mit Förderprogrammen* (S. 78-97). Frankfurt: Deutsche Gesellschaft für Lesen und Schreiben.

Hofmann B. & Sasse A. (Hrsg., 2006b): *Stellungnahme der DGLS zum Problem von LRS/Legasthenie. In Legasthenie. Lese-Rechtschreibstörungen oder Lese-Rechtschreibschwierigkeiten? Theoretische Konzepte und praktische Erfahrungen mit Förderprogrammen* (S. 13-20). Frankfurt: Deutsche Gesellschaft für Lesen und Schreiben.

Hohn, W. E. & Ehri, L. (1983). Do alphabet letters help prereaders acquire phonemic segmentation skill? *Journal of Educational Psychology, 75*, 752-762.

Holling, H. & Gediga, G. (2016). *Statistik – Testverfahren*. Göttingen: Hogrefe.

Hoover, W. & Gough, P. (2000). *The Reading Acquisition Framework – An overview by Wesley A. Hoover and Philip B. Gough.* URL: http://www.sedl.org/reading/framework/overview.html (Stand: 03.08.2016).

Horn, W. F. & Packard, T. (1985). Early identification of learning problems: A meta-analysis. *Journal of Educational Psychology, 77*, 597-607.

Hosp, M. K. & Hosp, J. (2003). Curriculum-based measurement for reading, math, and spelling: How to do it and why. *Preventing School Failure, 48 (1),* 10–17.

Hosp, M. K., Hosp, J. L. & Howell, K. W. (2006). *The ABC's of CBM. A Practical Guide to Curriculum-Based Measurement.* New York: Guilford Press.

Howell, K. W. & Nolet, V. (2001). *Curriculum-Based Evaluation: Teaching and Decision Making*. Wadeworth: Belmont CA.

Huber, C. & Grosche, M. (2012). Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik, 8*, 312-322.

IBM (2011): *SPSS Missing Values 20*. URL: http://www.math.uni-leipzig.de/pool/tuts/SPSS/IBM%20SPSS%20Missing%20Values.pdf (Stand: 27.07.2016).

Ingenkamp, K. & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik*. Weinheim und Basel (6. neu ausgest. Aufl.): Beltz.

Irvin, P. S., Alonzo, J., Nese, J. F. T., & Tindal, G. (2013). *Learning to read: kindergarten readiness growth in reading skills.* Retreived from the National Center on Assessment and Accountability for Special Education (NCAASE) URL: http://ncaase.com/ publications/in-briefs (Stand: 18.08.2016).

Jansen, H. (1992). *Untersuchungen zur Entwicklung lautsynthetischer Verarbeitungsprozesse im Vorschul- und frühen Grundschulalter.* Egelsbach: Verlag Hänsel-Hohenhausen.

Jansen, H. (1994). Fördert der Besuch des Schulkindergartens die Entwicklung schriftsprachlicher Fertigkeiten? *Zeitschrift für Pädagogische Psychologie, 8 (3/4),* 165 – 175.

Jansen, H. (2008). Rechtschreibdiagnostik in pädagogische-psychologischen Interventionen: Leistungen und Grenzen heutiger Rechtschreitests in der praktischen Anwendung. In W. Schneider, H. Marx & M. Hasselhorn (Hrsg.): *Diagnostik von Rechtschreibleistungen und -kompetenz. Tests und Trends N. F.* (S. 159-189) Band 6. Göttingen (u.a.): Hogrefe.

Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H. (1999). *Das Bielefelder Screening (BISC)*. Göttingen: Hogrefe.

Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese- und Rechtschreibschwierigkeiten (BISC)*. Göttingen (2. Auflage): Hogrefe.

Jansen, H. & Marx, H. (1999). Phonologische Bewußtheit und ihre Bedeutung für den Schriftspracherwerb. *Forum Logopädie, 2,* 7-16.

Jansen, H. & Thomé, G. (1998). Entwicklung der /ε/ - Phonemanalyse im ersten Schuljahr und ihr Zusammenhang mit der Wortschreibung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 30 (1)*, 20 – 25.

Jenkins, J. R. (2002): Early identification and intervention for young children with reading/learning disabilities. In: R. Bradley, L. Danielson & D. Hallahan (Eds.): *Identification of learning disabilites* (p. 99 – 149). Hillsdale, NJ: Erlbaum.

Jeske, W. (1981), Schule als ‚soziale Dirigierungsstelle'? In W. Twellmann: *Handbuch Schule und Unterricht, Bd. 2* (S. 225- 246). Düsseldorf: Schwann.

Johnston, R.S. (1982). Phonological coding in dyslexic readers. *British Journal of Psychology, 73,* 455-460.

Johnson, M. H. (1998). The neural basis of cognitive development. In D. Kuhn & R. S. Siegler (Hrsg.). *Handbook of child psychology* (p. 1-49), Vol 2: Cognition, perception, and language (5. Ed.). New York: Wiley.

Jones, E. & Krause, J. (1988). The effectiveness of data-based instruction by student teachers in classrooms for pupils with mild learning handicaps. *Teacher Education and Special Education, 11*, 9–19.

Jöreskog, K. G. & Sörbom, D. (1984). *LISREL VI – analysis of linear structural relationships by maximum likelihood and least squares methods*, research report (81-88), University of Uppsala 1984.

Jung, U. O. H. (1981). Linguistische Aspekte der Legasthenieforschung. In Valtin, R. Jung, U. O. H. & Scheerer-Neumann, G. (Hrsg.): *Legasthenie in Wissenschaft und Forschung*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Kalb, G., Rabenstein, R., Rost, D. H. (1979). *LUV. Lesen und Verstehen, Diagnose und Training*. Göttingen (u. a.): Hogrefe.

Kaminski, R., Cummings, K.D., Powell-Smith, K.A., & Good, R.H.III (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills for formative assessment and evaluation. In A. Thomas and J. Grimes (Eds.): *Best practices in school psychology* V. Bethesda, MD: National Association of School Psychologists.

Kaminski, R. A. & Good, R. H. Toward a Technology for Assessing Basic Early Literacy Skills. *School Psychology Review, 25(2)*, 215-227.

Katz, R.B., Shankweiler, D. & Liberman, I.Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32*, 474-484.

Kelava, A. (2010). SPSS-Beispiel zum Kapitel 4: Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. Aus: Moosbrugger, H. & Kelava, A. (2007). Testtheorie und Fragebogenkonstruktion. Heidelberg: Springer.

Kim, Y.-S. & Pallante, D. (2010). *Predictors of reading skills for kindergartners and first grade students in Spanish: a longitudinal study. Reading and Writing*, Online-Erscheinung. URL: http://www.springerlink.com/content/w701755864115724/ (Stand: 20.09.2010).

Kirkbride, S. & Wright, B. C. (2002). The role of analogy use in improving early spelling performance. *Educational and Child Psychology, 19 (4),* 91-102.

Kirschhock, E.-M. (2003). *Die Entwicklung schriftsprachlicher Kompetenzen im ersten Schuljahr*. URL: www.opus.ub.uni-erlangen.de/opus/volltexte/2004/55/ (Stand: 20.10.2010).

Klauer, K. J. (2001). Forschungsmethoden der Pädagogischen Psychologie. In A. Krapp & B. Weidenmann (Hrsg.): *Pädagogische Psychologie* (S. 75-97). Weinheim (4. vollst. überarb. Aufl.): Beltz.

Klauer, K. J. (2006). Erfassung des Lernfortschritts durch curriculumbasierte Messung. *Heilpädagogische Forschung, 32 (1)*, 16-26.

Klauer, K. J. (2011). Lernverlaufsdiagnostik – Konzept, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik, 3 (3),* 207-224.

Klauer, K. J. (2014a). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), Lernverlaufsdiagnostik (Tests und Trends N.F. Band 12) (S. 1-18). Göttingen: Hogrefe.

Klauer, K. J. (2014b). Training des induktiven Denkens – Fortschreibung der Metaanalyse von 2008. *Zeitschrift für Pädagogische Psychologie, 28 (1-2),* 5-19.

Klauer, K. J. & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research, 78,* 85-123.

Klein, G. (2007). Frühe Kindheit und Vorschulalter. In J. Walter & F. B. Wember (Hrsg.): *Sonderpädagogik des Lernens. Handbuch Sonderpädagogik*, Band 2 (S. 220-244). Göttingen (u. a.): Hogrefe.

Kleiter, E. F.(1988). *Lehrbuch der Statistik* Band 2/2. Weinheim: Deutscher Studien Verlag.

Klicpera, C. & Gasteiger-Klicpera, B. (1993). *Lesen und Schreiben. Entwicklung und Schwierigkeiten. Die Wiener Längsschnittuntersuchungen über die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit*. Bern: Huber.

Klicpera, C. & Gasteiger-Klicpera, B. (1995). *Psychologie der Lese- und Schreibschwierigkeiten. Entwicklung, Ursachen und Förderung.* Weinheim: Psychologie Verlags Union.

Klicpera, C. & Gasteiger Klicpera, B. (2001). Macht Intelligenz einen Unterschied? Rechtschreiben und phonologische Fertigkeiten bei diskrepanten und nichtdiskrepanten Lese/Rechtschreibschwierigkeiten. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 29 (1),* 37-49.

Klicpera, C., Gasteiger-Klicpera, B. & Schabmann, A. (1994). Wieweit unterschieden sich durchschnittliche Leser mit Rechtschreibschwierigkeiten von Kindern mit Lese- und Rechtschreibschwierigkeiten? Verlauf, Art der Rechtschreibfehler und Lernvoraussetzungen. *Zeitschrift für Kinder- und Jugendpsychiatrie, 22*, 87-96.

Klicpera, C., Schabmann, A. & Gasteiger-Klicpera, B. (2010). *Legasthenie - LRS. Modelle, Diagnose, Therapie und Förderung.* München (3. akt. Aufl.): Ernst Reinhardt.

Klieme, E., Jude, N., Baumert, J. & Prenzel, M. (2010). *PISA 2000-2009: Bilanz der Veränderungen im Schulsystem.* In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.): PISA 2009. Bilanz nach einem Jahrzehnt (S. 277-297). Münster: Waxmann.

Klopp, E. (2013). Explorative Faktorenanalyse. URL: psydoc.sulb.uni-saarland.de/volltexte/2013/4823/pdf/Exploartive_Faktorenanalyse_final.pdf (Stand: 01.12.2015).

KMK (2010). *Pädagogische und rechtliche Aspekte der Umsetzung des Übereinkommens der Vereinten Nationen vom 13. Dezember 2006 über die Rechte von Menschen mit Behinderungen* (Behindertenrechtskonvention - VN-BRK) in der schulischen Bildung (Beschluss der Kultusministerkonferenz vom 18.11.2010). URL: http://www.kmk.org/fileadmin/ veroeffentlichungen_beschluesse/2010/2010_11_18-Behindertenrechtkonvention.pdf (Stand: 17.01.2013).

KMK (2011). *Inklusive Bildung von Kindern und Jugendlichen mit Behinderungen an Schulen*. Beschluss der Kultusministerkonferenz vom 20.10.2011. URL: http://www.kmk.org/fileadmin/ veroeffentlichungen_beschluesse/2011/2011_10_20-Inklusive-Bildung.pdf (Stand: 17.01.2013).

Knauf, T. & Schubert, E. (2008). Der Übergang vom Kindergarten in die Grundschule. Grundlagen, Lösungsansätze und Strategien für eine systematische Neustrukturierung des Schulanfangs. In M. R. Textor: *Kindergartenpädagogik,* Online-Handbuch. URL: www.Kindergartenpädagogik.de (Stand: 05.05. 2009).

Knopp, E. & Hartke, B. (2010). Das Inventar Rechenfische – Anwendung, Reliabilität und Validität eines Verfahrens zur Erfassung des Leistungsstandes von Erstklässlern in Mathematik. *Empirische Sonderpädagogik, 3*, 5-25.

Krajewski, K. (2003). *Vorhersage von Rechenschwäche in der Grundschule*. Hamburg: Verlag Kovac.

Krajewski, K., Liehm, S. & Schneider, W. (2002). *Deutscher Mathematiktest für zweite Klassen (DEMAT 2+).* Göttingen: Hogrefe.

Kretschmann, R. (2002). Störungen beim Schriftspracherwerb – Ursachen und Prävention aus systemischer und entwicklungsökologischer Sicht. URL: http://test.alphabetisierung.de /fileadmin/files/Dateien/Downloads_Texte/Kretschmann_Storungen_systemisch.pdf (Stand: 21.01.2013).

Krüger, H.H. & Rauschenbach, T. (1997). *Einführungskurs Erziehungswissenschaft. 4. Einführung in die Arbeitsfelder der Erziehungswissenschaft* (S. 96-99). Frankfurt am Main (2., durchges. Aufl.): Deutsches Institut für Internationale Pädagogische Forschung.

Kubinger, K. D. & Proyer, R. (2004). Gütekriterien. In K. Westhoff, L. J. Hellfritsch, L. F. Hornke, K. D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel & G. Reimann (Hrsg.): *Testkuratorium der Föderation Deutscher Psychologenvereinigungen. Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (S. 186-194). Lengerich: Pabst.

Kuhlmann, K. (2013). *Entwicklung und Evaluation eines Verfahrens zur Lernverlaufsdokumentation im Lernbereich Rechtschreiben der 2. Klassenstufe.* URL: http://rosdok.uni-rostock.de/file/rosdok_disshab_0000001127/rosdok_derivate_0000005238/Dissertation_Kuhlmann_2014.pdf (Stand: 05.08.2016).

Lang, H. Moosbrugger, A. Pöschel, G. Reimann (Hrsg.): *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (S. 186-194). Lengerich: Pabst Science Publishers.

Küspert, P. (1998). *Phonologische Bewusstheit und Schriftspracherwerb. Zu den Effekten vorschulischer Förderung der Phonologischen Bewusstheit auf den Erwerb des Lesens und Rechtschrei*bens. Frankfurt am Main: Lang.

Küspert, P. & Schneider, W. (1998). *Würzburger Leise Leseprobe (WLLP).* Göttingen: Hogrefe.

Küspert, P. & Schneider, W. (2006). *Hören, lauschen, lernen – Sprachspiele für Vorschulkinder*, Göttingen (5. überarb. Auflage): Vandenhoeck & Ruprecht.

LaBuda, M. & DeFries, J. C. (1989). Differential prognosis of reading-disabled children as a function of gender, social-economic status, IQ and severity: A longitudinal study. Reading and Writing: *An Interdisciplinary Journal, 1,* 25-36.

Landerl, K. (1996). *Legasthenie in Deutsch und Englisch.* Frankfurt: Peter Lang.

Landerl, K. (2001). Beeinträchtigungen der phonologischen Verarbeitung – ein wesentliches Handicap für das Lesenlernen. In G. Schulte-Körne (Hrsg.). *Legasthenie: erkennen, verstehen, fördern* (S. 61-66). Bochum: Winkler.

Landerl, K., Linortner, R. & Wimmer, H. (1992). Phonologische Bewußtheit und Schriftspracherwerb im Deutschen. *Zeitschrift für Pädagogische Psychologie, 6,* 17-32.

Landerl, K. & Wimmer, H. (1994). Phonologische Bewußtheit als Prädiktor für Lese- und Schreibfertigkeiten in der Grundschule. *Zeitschrift für Pädagogische Psychologie, 8 (3/4)*, 153-164.

Landerl, K. & Wimmer, H. (1998). Lesenlernen bei deutsch- und englischsprachigen Kindern. Normaler und gestörter Leseerwerb. In R. Weingarten & H. Günther (Hrsg.): *Schriftspracherwerb* (S. 82-97). Baltmannsweiler: Schneider Verlag Hohengehren.

Landerl, K. & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: an 8-year follow-up. *Journal of Educational Psychology, 100 (1),* 150 – 161.

Landerl, K., Wimmer, H. & Moser, E. (1997). *Salzburger Lese- und Rechtschreibtest (SLRT).* Bern: Huber.

Larsen, P. J., H□ien, T., Lundberg, I. & □degaard, H. (1990). MRI evaluation of the size and symmetry of the planum temporale in adolescents with developmental dyslexia. *Brain and Language, 39*, 289-301.

Leach, D. J. (1983). Early screening procedures. *School Psychology International, 4*, 47-55.

Lenel, A. (2005). *Schrifterwerb im Vorschulalter. Eine entwicklungspsychologische Längsschnittstudie*. Weinheim: Beltz PVU.

Lenhard, W. & Marx, P. (2010). *Auswertungshilfe zur Berechnung von Testkennwerten von Screeningverfahren*. Universität Würzburg: Lehrstuhl Psychologie IV. URL: http://www.i4.psychologie.uni-wuerzburg.de/mitarbeiter/dr_wolfgang_lenhard/materialien (Stand: 26.7.2016).

Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.

Liberman, I. Y., Mann, V. A., Shankweiler, D. & Werfelmann, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex, 18*, 367-365.

Liberman, I. Y., Shankweiler, D. Fischer, W. F. & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Lichtenstein, R. & Ireton, H. (1984). *Preschool screening: Identifying young children with developmental an educational problems*. Orlando: Grune & Stratton.

Liebers, K., & Seifert, C. (2012). Assessmentkonzepte für die inklusive Schule – eine Bestandsaufnahme. Zeitschrift für Inklusion, 0(3). URL: http://www.inklusion-online.net/index.php/inklusion/article /view/166/156 (Stand: 18.01.2013).

Liegle, L. (2008). Erziehung als Aufforderung zur Bildung. In W. Thole, H.-G. Roßbach, M. Fölling-Albers & R. Tippelt (Hrsg.), *Bildung und Kindheit. Pädagogik der Frühen Kindheit in Wissenschaft und Lehre*. (S. 85-113), Opladen & Farmington Hills: Barbara Budrich.

Lienert, G. A. (1969). *Testaufbau und Testanalyse*. Weinheim: Verlag Beltz.

Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim (6. Aufl.): Psychologie Verlags Union.

Ligges, C. & Blanz, B. (2007). Übersicht über Bildgebungsbefunde zum phonologischen Defizit der Lese- Rechtschreibstörung bei Kinder und Erwachsenen: Grundlegende Defizite oder Anzeichen von Kompensation? *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 35 (2),* 107-117.

Linder, M. (1951). Über Legasthenie (spezielle Leseschwäche) 50 Fälle, ihr Erscheinungsbild und Möglichkeiten der Behandlung. *Zeitschrift für Kinderpsychiatrie, 18*, 100.

Linder, M. (1962). *Lesestörungen bei normalbegabten Kindern*. Zürich.

Linder, M. & Grissemann, H. (1996). *Zürcher Lesetest (ZLT).* Bern (5. Aufl.): Huber.

Lotz, M., Schoreit, E. & Kempte, I. (2013). Die Entwicklung des Leseverständnisses und der Rechtschreibfähigkeit. In F. Lipowsky, G. Faust & C. Kastens (Hrsg.): *Persönlichkeits- und Lernentwicklung an staatlichen und privaten Grundschulen. Ergebnisse der PERLE-Studie zu den ersten beiden Schuljahren* (S. 77-98). Münster: Waxmann.

Lukesch, H. (1998). *Einführung in die pädagogisch-psychologische Diagnostik*. Regensburg (2., vollst. überarb. Aufl.): Roderer.

Lundberg, I. (1995). The computer as a tool of remediation in the education of students with reading disabilities – a theory-based approach. *Learning Disability Quarterly, 18*, 89-99.

Lundberg, I. & H□ien, T. (1989). Phonemic Deficits: A Core Symptom of Developmental Dyslexia?. *The Irish Journal of Psychology, 10 (4)*, 579 – 592.

Lundberg, I. & H□ien, T. (1990). Patterns of information processing skills and word recognition strategies in developmental dyslexia. *Scandinavian Journal of Educational Research, 34*, 231-240.

Lundberg, I., Frost, J. & Peterson, O.-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23 (3),* 265-284.

Lundberg, I., Olofsson A. & Wall, S. (1980). Reading and spelling skills in the first school year predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Lysakowski, R. S. & Walberg, H. J. (1982). Instructional effects of cues, participation, and corrective

feedback: A quantitative synthesis. *American Educational Research Journal, 19*, 559-578.

Lyon, G.R. et al (2001). Rethinking learning disabilities. In C. E. Finn et al. (Eds.): *Rethinking special education for a new century* (p. 259-287). Washington: Fordham Foundation.

Lyytinen, H., Aro, M., Eklund, K., Erskine, J., Guttorm, T., Laakso, M.-L., Leppänen, P. H. T., Lyytinen, P., Poikkeus, A.-M., Richardson, U. & Torppa, M. (2004). The development of children at familial risk for dyslexia: birth to early school age. *Annals of Dyslexia, 54 (2),* 184-220.

Lyytinen, H., Ahonen, T., Eklund, K., Guttorm, T., Kulju, P., Laakso, M.-L., Leiwo, M., Leppänen, P., Lyytinen, P., Poikkeus, A.-M., Richardson, U., Torppa, M. & Viholainen, H. (2004). Early development of children at familial risk for dyslexia – follow-up from birth to school age. *Dyslexia, 10*, 146-178.

MacDonald, G. W. & Cornwall, A. (1995). The relationship between phonological awareness and reading and selling achievement eleven years later. *Journal of Learning Disabilities, 28,* 523-527.

MacLean, M., Bryant, P. E. & Bradley L. (1987). *Rhymes, nursery rhymes, and reading in early ch*ildhood. Merrill-Palmer Quarterly, 33, 255-281.

Machemer (1972). Entwicklung eines Übungsprogramms für Eltern zur Behandlung von Legasthenikern nach verhaltenstheoretischem Modell. *Schule und Psychologie, 19*, 336-346.

Madelaine, A. & Wheldall, K. (1999). Curriculum-based Measurement of Reading: a critical review. International *Journal of Disability, Development & Education, 46 (1*), 71-85.

Madelaine, A., & Wheldall, K. (2004). Curriculum-based measurement of reading: Recent Advances. *International Journal of Disability, Development and Education, 51 (1),* 57-82.

Mähler, C. & Schuchardt, K. (2014). Die Bedeutung des Arbeitsgedächtnisses für den Schriftspracherwerb. In G. Schulte-Körne (Hrsg.): *Legasthenie und Dyskalkulie. Neue Methoden zu Diagnostik und Förderung (S. 77-85).* Bochum: Winkler.

Mahlau, K. (2013). Vergleich zwischen inklusiven und separierenden Unterrichtskonzepten unter besonderer Berücksichtigung von Kindern mit spezifischer Sprachentwicklungsstörung: Lernausgangslage und erste Ergebnisse. *Forschung Sprache, 1,* 4-22.

Mahlau, K., Blumenthal, Y., Diehl, K., Schöning, A., Sikora, S., Voß, S. & Hartke, B. (2014). Das Rügener Inklusionsmodell (RIM) – RTI in der Praxis. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik* (Tests & Trends, NF Bd. 12, S. 101-125). Göttingen: Hogrefe.

Malmquist, E. (1958). *Factors related to reading disabilities in the first grade of the elementary school.* Stockholm: Acta Universitatis Stockholmiensis.

Mann, V. (1984). Longitudinal prediction and prevention of reading difficulty. *Annals of Dyslexia, 34*, 117-134.

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition, 24*, 65-92.

Mann, V. A. (1991). Phonological awareness and early reading ability: One perspective. In D. J. Sawyer & B. J. Fox (Ed.). *Phonological awarenss in reading – The evolution of current perspectives* (p. 191-215). New York: Springer.

Mann, V. A. (1993). Phoneme awareness and future reading ability. *Journal of Learning Disabilities, 26,* 259-269.

Mann, V. A. & Libermann, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17*, 592-599.

Mannhaupt, G. (2003). Früherkennung und Prävention von Problemen im Schriftspracherwerb. *Grundschule, 9*, 45-49.

Mannhaupt, G. (2006). Ergebnisse von Therapiestudien. In W. von Suchodoletz (Hrsg.): *Therapie der Lese-Rechtschreib-Störung (LRS). Traditionelle und alternative Behandlungsmethoden im Überblick* (S. 93-110). Stuttgart (2. überarb. und erw. Aufl.): Kohlhammer.

Mannhaupt, G. (2008). Prävention von Lese-Rechtschreibschwierigkeiten im Kindergarten. In J.Borchert, B. Hartke & P. Jogschies (Hrsg.): *Frühe Förderung entwicklungsauffälliger Kinder und Jugendlicher*. Stuttgart: Kohlhammer.

Mannhaupt, G. & H. Jansen (1989). Phonologische Bewußtheit: Aufgabenentwicklung und Leistungen im Vorschulalter. *Heilpädagogische Forschung, 15 (1*), 50 – 56.

Marsh, G., Friedman, M., Welch, V. & Desberg, P. (1980). The development of strategies in spelling. In U. Frith (Ed.): *Cognitive processes in spelling* (p. 339 – 353). Academic Press: London.

Marston, D. (1988). The effectiveness of special education: A time-series analysis of reading performance in regular and special education settings. *The Journal of Special Education, 21,* 13–26.

Marston, D., Diment, K., Allen, D., & Allen, L. (1992). Monitoring pupil progress in reading. *Preventing School Failure, 36(2)*, 21–25.

Martinéz, R. S., Nellis, L. M. & Prendergast, K. A. (2006). Closing the Achievement Gap Series: Part II Response to intervention (RTI) – Basic elements, practical applications, and policy recommendations. *Center for evaluation & education policy, 4 (8),* 1 – 7.

Martschinke, S., Kirschhock, E.-M., Frank, A. (2001). *Diagnose und Förderung im Schriftspracherwerb. Der Rundgang durch Hörhausen. Erhebungsverfahren zur phonlogischen Bewusstheit*. Band 1. Donauwörth: Auer.

Martschinke, S., Kirschhock, E.-M., Frank, A. (2005). *Diagnose und Förderung im Schriftspracherwerb. Der Rundgang durch Hörhausen. Erhebungsverfahren zur phonlogischen Bewusstheit*. Band 1. Donauwörth (4. Aufl.): Auer.

Martschinke, S., Kirschhock, E.-M., Frank, A. (2011). *Diagnose und Förderung im Schriftspracherwerb. Der Rundgang durch Hörhausen. Erhebungsverfahren zur phonlogischen Bewusstheit*. Band 1. Donauwörth (7. überarb. Aufl): Auer.

Marx, H. (1992a). Methodische und inhaltliche Argumente für und wider eine frühe Identifikation und Prädiktion von Lese- und Rechtschreibschwierigkeiten. *Diagnostica, 38 (3)*, 249-268.

Marx, H. (1992b). *Vorhersage von Lese- Rechtschreibschwierigkeiten in Theorie und Anwend*ung. Unveröffentlichte Habilitationsschrift. Universität Bielefeld.

Marx, H. (1992c). Frühe Identifikation und Prädiktion von Lese-Rechtschreibschwierigkeiten: Bestandsaufnahme bisheriger Bewertungsgesichtspunkte von Längsschnittstudien. *Zeitschrift für Pädagogische Psychologie, 6*, S. 35-48.

Marx, H. (1997). Erwerb des Lesens und Rechtschreibens: Literaturüberblick. In F. E. Weinert (Hrsg.): *Entwicklung im Grundschulalter* (S. 85-111). Weinheim: Beltz PVU.

Marx, H. (1998). *Knuspels Leseaufgaben (KNUSPEL-L*). Göttingen: Hogrefe.

Marx, E. (2006). Profitiert das kindliche Sprachsystem von anderen kognitiven Entwicklungsbereichen? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38,* 139-145.

Marx, H. (2007). Theorien und Determinanten des Erwerbs der Schriftsprache. In H. Schöler & A. Welling: Sonderpädagogik der Sprache (S. 92-148). Göttingen: Hogrefe.

Marx, H. (2009). Does fostering inductive reasoning promote children's language acquisition? *Educational & Child Psychology, 26*, 40-58.

Marx, H. & Jansen, H. (1999). Möglichkeiten und Grenzen der Früherkennung und Vorhersage von Lese-Rechtschreibschwierigkeiten. *Forum Logopädie, 6,* 7-16.

Marx, H., Jansen, H. & Skowronek, H. (2000). Prognostische, differentielle und konkurrente Validität des Bielefelder Screenings zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC). In M. Hasselhorn, W. Schneider & H.

Marx, H. & Jungmann, T. (2000). Abhängigkeit der Entwicklung des Leseverstehens von Hörverstehen und grundlegenden Lesefertigkeiten im Grundschulalter: Eine Prüfung des Simple View of Reading-Ansatzes. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 32* (2), 81-93.

Marx, E. & Keller, K. (2010). Effekte eines induktiven Denktrainings auf die Denk- und Sprachentwicklung bei Vorschulkindern und Erstklässlern in benachteiligten Stadtteilen. *Zeitschrift für Pädagogische Psychologie, 24(2)*, 139-146.

Marx, E., Keller, K. & Beuing, R. (2011). Die Erzieherin als Trainerin. Effekte kombinierter Denk- und Sprachförderung in Kindertagesstätten. *Psychologie in Erziehung und Unterricht, 58*, 41-51.

Marx, P. & Lenhard, W. (2010). Diagnostische Merkmale von Screeningverfahren. In M. Hasselhorn & W. Schneider (Hrsg.): *Frühprognose schulischer Kompetenzen.* Göttingen: Hogrefe. URL: http://www.i4.psychologie.uni-wuerzburg.de/fileadmin/06020400/user_upload/Lenhard/TestsundTrends9_Marx_und_Lenhard_Endversion.pdf (Stand: 03.08.2016).

Marx, P. & Weber, J. (2004). Wie gut lassen sich Lese-Rechtschreibschwierigkeiten bereits im Kindergarten vorhersagen? Neue Ergebnisse zum Bielefelder Screening. In A. Möckel, E. Breitenbach, W. Drave & H. Ebert (Hrsg.): *Lese-Schreibschwäche – Vorbeugen, Erkennen, Helfen* (S. 194-208). Würzburg: Bentheim Bildung.

Marx, P. & Weber, J. (2006). Vorschulische Vorhersage von Lese- und Rechtschreibschwierigkeiten. Neue Befunde zur prognostischen Validität des Bielefelder Screenings (BISC). *Zeitschrift für Pädagogische Psychologie, 20 (4)*, 251- 259.

Marx, P., Weber, J. & Schneider, W. (2001). Legasthenie versus allgemeine Lese-Rechtschreibschwäche. Ein Vergleich der Leistungen in der phonologischen und visuellen Informationsverarbeitung. *Zeitschrift für Pädagogische Psychologie, 15 (2*), 85-98.

May, P. (2012). HSP: Die Hamburger Schreib-Probe. Handbuch für alle Stufen: Diagnose orthographischer Kompetenz 1.-9. Schuljahr, Hamburg (6. neustandardisierte Aufl.): Verlag für pädagogische Medien im Ernst Klett Verlag.

May, P. & Okwumo, S. (1999). *Effekte vorschulischer Trainings zur Schriftanbahnung auf das Rechtschreiblernen* im ersten Schuljahr. Forschungsbericht – psychologisches Institut II der Universität Hamburg. URL: http://www1.uni-hamburg.de/psycholo/frames/projekte /PLUS/May_Okwumo_99_Eff_VS_Train.pdf (Stand: 05.10.2010).

Mayringer, H. & Wimmer, H. (2003). *SLS 1-4. Salzburger Lese-Screening für die Klassenstufen 1-4*. Bern: Huber.

Mayringer, H., Wimmer, H. & Landerl, K. (1998). Die Vorhersage früher Lese- und Rechtschreibschwierigkeiten: Phonologische Schwächen als Prädiktoren. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 30 (2)*, 57-69.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M. & Mengler E. D. (2000). On the "specifics" of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry, 41(7),* 869-874.

McConnell, S. R., Priest, J. S., Davis, S. D., & McEvoy, M. A. (2002). Best practices in measuring growth and development for preschool children. In A. Thomas and J. Grimes (Eds.): *Best practices in school psychology* (pp. 1231–1246). Washington DC (4th ed.): National Association of School Psychologists.

Mehrens, W. A. & Clarizio, H. F. (1993). Curriculum-based measurement: Conceptual and psychometric considerations. *Psychology in the Schools, 30 (3),* 241-254.

Meiers, K. (1998). Schriftspracherwerb. Anfangsunterricht im Lesen und Schreiben. In G. Lange, K. Neumann & W. Ziesenis (Hrsg.): *Taschenbuch des Deutschunterrichts*, Band 1 (S.144-171). Baltmannsweiler (6., vollst. überarb. Aufl.): Schneider.

Metz, U., Marx, P. Weber, J. & Schneider, W. (2003). Overachievement im Lesen und Rechtschreiben: Folgerungen für die Diskrepanzdefinition der Legasthenie. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie, 35 (3),* 127-134.

Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein (2004). *Bildungspläne der Bundesländer für die frühe Bildung in Kindertageseinrichtungen*. URL: http://www.bildungsserver.de/Bildungsplaene-der-Bundeslaender-fuer-die-fruehe-Bildung-in-Kindertageseinrichtungen-2027.html (Stand: 30.01.2013).

Ministerium für Bildung und Weiterbildung NRW (2014). *Schriftlicher Bericht zum Thema „Response to Intervention" zur Sitzung des Ausschusses am 12.2.2014.* URL: https://www.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMV16-1622.pdf (Stand: 07.08.2014).

Missall, K. N. & McConnell, S. R. (2004). *Technical Report. Psychometric characteristics of individual growth & development in dicators: Picture naming, rhyming, and alliteration*. URL: http://ggg.umn.edu (Stand: 21.01.2013).

Missall, K. N., McConnell, S. R. & Cadigan, K. (2006). Early literacy development: Skill growth and relations between classroom variables for preschool children. *Journal of Early Intervention, 29,* 1–21.

Missall, K., Reschly, A., Betts, J., McConnell, S., Heistad, D., Pickart, M., Sheran, C., & Marston, D. (2007). Examination of the predictive validity of preschool early literacy skills. *School Psychology Review, 36,* 433-452.

Moll, K., Fussenegger, B., Willburger, E. & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. *Scientific Studies Of Reading, 13 (1),* 1 – 25.

Moll, K., Landerl, K. & Kain, W. (2008). In W. Schneider, H. Marx & M. Hasselhorn (Hrsg.): *Diagnostik von Rechtschreibleistungen und -kompetenz. Tests und Trends N. F.* (S. 129-144) Band 6. Göttingen (u.a.): Hogrefe.

Moosbrugger, H. & Rauch, W. (2004). Konstruktionsgrundlagen von Verfahren der Eignungsbeurteilung. In K. Westhoff, L. J. Hellfritsch, L. F. Hornke, K. D. Kubinger, F. Lang, H. Moos- Brugger, A. Püschel, G. Reimann (Hrsg.): Testkuratorium der Föderation Deutscher Psychologenvereinigungen, *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (S. 195-200). Lengerich: Pabst Science Publishers.

Morais, J. (1991). Constraints on the development of phonemic awareness. In S. Brady & D. P. Shankweiler (Eds.): *Phonological processes in literacy – a tribute to Isabelle Y. Liberman* (pp. 5-27). Hillsdale, NJ: Erlbaum.

Morais, J., Bertelson, P., Cary, L. & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24,* 45-64.

Morais, J., Cary, L., Alegria, J. & Bertelson, P. (1985). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323-331.

Morgan, W. P. (1896). A case of congenital wordblindness. *British Medical Journal, 7*, 21-28.

Mummendey, H. D. (1995). *Die Fragebogen-Methode. Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung*. Göttingen (2. korr. Aufl.): Hogrefe.

Müller, M. M. & Hartmann, E. (2009). Lernfortschritte im Unterricht erheben – Möglichkeiten und Grenzen des curriculumbasierten Messens. *Schweizerische Zeitschrift für Heilpädagogik, 15 (10),* 36-42.

Müller, R. (1967). Aktuelle und dispositionelle Faktoren beim Lesen und Schreiben. *Der Schulpsychologe, 3/*4, 37-49.

Müller, R. (1980a). *Leseschwäche, Leseversagen, Legasthenie*. Band 1. Weinheim (2. Aufl.): Beltz.

Müller, R. (1980b). *Leseschwäche, Leseversagen, Legasthenie*. Band 2. Weinheim (2. Aufl.): Beltz.

Müller, R. (2003a). *Diagnostischer Rechtschreibtest für 1. Klassen (DRT1).* Weinheim (2. akt. Auflage): Beltz.

Müller, R. (1983). *Diagnostischer Rechtschreibtest für 2. Klassen (DRT 2).* Weinheim (2. akt. Auflage): Beltz.

Müller, R. (2003b). *Diagnostischer Rechtschreibtest für 2. Klassen (DRT 2).* Göttingen (4. akt. Auflage): Beltz.

Müller, R. (2003c). *Diagnostischer Rechtschreibtest für 3. Klassen (DRT 3).* Göttingen (4.akt. Auflage): Beltz.

Müller, C. M. & Hartmann, E. (2013). Schulweite Prävention von Lernproblemen im RTI-Modell. *Schweizerische Zeitschrift für Heilpädagogik, 15 (9),* S. 25-33.

Müller, C. M. & Hartmann, E. (2014). *Lernfortschrittsdiagnostik: Grundrechenarten. 12 Drei-Minuten-Tests für den inklusiven Mathematikunterricht – ZR bis 100*. Hamburg: Persen-Verlag.

Muter, V., Hulme, C., Snowling, M. & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65*, 370-396.

Muter, V. & Snowling, M. (1998). Concurrent and longitudinal predictors of reading: the role of metalinguistic and short- term- memory skills. *Reading Research Quarterly, 33*, 320- 337.

Naegele, I. M. (1993). Wie hilfreich sind die LRS-Erlasse und Richtlinien der Bundesländer? In: M. Naegele & R. Valtin (Hrsg.): *LRS in den Klassen 1-10* (S. 21-30). Weinheim (u.a.), (3. unveränd. Aufl.): Beltz.

Näslund, J. C. & Schneider, W. (1991). Longitudinal effects of verbal ability, memory capacity and phonological awareness on reading performance. *European Journal of Psychology of Education, 6*, 375-392.

Näslund, J. C. & Schneider, W. (1993). Emerging literacy form kindergarten to second grade: Evidence from the Munich Longitudinal Study on the Genesis of Individual competencies. In H. Grimm & H. Skowronek (Ed.): *Language Acquisition Problems and Reading Disorders: Aspects of Diagnosis and Intervention* (p. 295-318). Berlin: De Gruyter.

Näslund, J. C. & Schneider, W. (1996). Kindergarten letter knowledge, phonological skills and memory processes: Relative effects on early literacy. *Journal of Experimental Child Psychology, 62*, 30-59.

Noack, P. & Brändel, S. (2007). Förderung der phonologischen Bewusstheit bei Kindergartenkindern. Trainingseffekte und Prüfung möglicher Moderatoren. *Diskurs Kindheits- und Jugendforschung, 2 (3),* 311-322.

Nation, K. & Hulme, C. (1997). Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills. *Reading Research Quarterly, 32*, 154-167.

Nation, K. & Hulme, C. (1998). The role of analogy in early spelling development. In C. Hulme & R. M. Joshi (Ed.): *Reading and spelling development and disorders* (p. 433-446). London: Erlbaum.

Niemeyer, W. (1974). *Legasthenie und Milieu*. Hannover: Schroedel.

Organisation for Economic Cooperation and Development - OECD (2016). *Ländernotiz. Programme for international student assessment (PISA). PISA 2015 Ergebnisse, Deutschland*. URL: http://www.oecd.org/pisa/pisa-2015-Germany-DEU.pdf (Stand: 06.08.2016)

Oerter, R. (2000). Implizites Lernen beim Sprechen, Lesen und Schreiben. *Unterrichtswissenschaft, 28*, 239-257.

Orden, G. C., van, Pennington, B. F. & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 9*7, 488-522.

Olsen, A. V. (1966). Relation of achievement test scores and specific reading abilities to the Forstig Developmental Test of Visual Perception. *Perceptual and Motor Skills, 22,* 179-184.

Olson, R. K. (2004). Genetic and environmental causes of reading disabilities: results from the Colorado Learning Disabilities Research Center. In M. Turner & J. Rack (Ed.): *The study of dyslexia* (p. 23-33). New York: Kluwer academic.

Padget, S. Y. (1998). Lessons from research on dyslexia: Implications for a classification system for learning disabilities. *Learning Disability Quarterly, 21*, 167-178.

Pennington, B. F., Gilger, J. Pauls, D. Smith, S. S. A., Smith, S. D. & DeFries, J. C. (1991). Evidence for a major gene transmission of developmental dyslexia. *Journal of American Medical Association, 18*, 1527-1534.

Perfetti, C. A., Georgi, M. C. & Beck, I. (1993). Implications of the Pittsburgh Study for issues of risk. In H. Grimm & H. Skowronek (Eds.): *Language acquisition problems and reading disorders aspects of diagnosis and intervention* (S. 193-211). New York: de Gruyter.

Petermann, F. (2003). Legasthenie und Rechenstörung – Einführung in den Themenschwerpunkt. Kindheit und *Entwicklung, 12 (4),* 193-196.

Petermann, F. (2006). Veränderungsmessung. In D. H. Rost (Hrsg.): *Handwörterbuch Pädagogische Psychologie* (S. 821- 827). Weinheim (3. überarb. Aufl.): Beltz.

Petermann, F., Daseking, M. (2012). ZLT – II Zürcher Lesetest – II Weiterentwicklung des Zürcher Lesetests (ZLT) von Maria Linder und Hans Grissemann. Göttingen, Hogrefe.

Pfeiffer, G. & Zielinski, W. (1975). Über den Zusammenhang zwischen Rechtschreibung und Intelligenzleistung. Psychologie in Erziehung und Unterricht, 2, 1-8.

Poskiparta, E., Niemi, P. & Vauras, M. (1999). Who benefits from training in linguistic awareness in the first grade, and what components show training effects? *Journal of Learning Disabilities, 32*, 437-446, 456.

Priest, J. S., McConnell, S. R., Walker, D., Carta, J., Kaminski, R. A., McEvoy, M. A., Good, R. H., Greenwood C. R. & Shinn M. R. (2001). General Growth Outcomes for Young Children: Developing a Foundation for Continuous Progress Measurement. *Journal o Early Intervention, 24*, 163-180.

Priest, J. S., Silberglitt, B., Hall, S. & Estrem, T. L. (2000). *Progress on preschool IGDIs for early literacy. Presentation at Heartland Area Education Association*, Des Moines, IA.

Probst, H. & Euker, N. (2012). Pädagogische Testdiagnostik in der inklusiven Schule. In V. Moser (Hrsg.): *Die inklusive Schule. Standards für die Umsetzung* (S. 195-206). Stuttgart: Kohlhammer.

Ptok, M. & Dunkelmann, A. (2011). Basale phonologische Verarbeitung und Intelligenzpotenzial bei Vorschulkindern. *HNO, 60*,427-431.

Punt, M., Jong, M. de, Groot, E. de & Hadders-Algra, M. (2010). Minor neurological dysfunction in children with dyslexia. *Developmental medicine & child neurology, 10*, 1- 6.

Puolakanaho, A., Poikkeus, A.-M., Ahonen, T., Tolvanen, A. & Lyytinen, H. (2004). Emerging phonological awareness differentiates children with and without familial risk for dyslexia after controlling for general language skills. *Annals of Dyslexia, 54 (2),* 221-243.

Quaiser-Pohl, C. (2010). Wie teste ich Kinder? Zur Gestaltung der diagnostischen Situation. In C. Quaiser-Pohl & Rindermann, H. (Hrsg.): *Entwicklungsdiagnostik* (S. 57-68). München: Ernst Reinhardt.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Ranschburg, P. (1916). *Die Leseschwäche und die Rechtschreibschwäche der Kinder*. Berlin.

Richter, S. (1992). Die Rechtschreibentwicklung im Anfangsunterricht und Möglichkeiten der Vorhersage ihrer Störungen. Hamburg: Dr. Kovač.

Richter, S. & Brügelmann, H. (1994). Der Schulanfang ist keine Stunde Null. In S. Richter & H. Brügelmann (Hrsg.): *Wie wir recht schreiben lernen* (S. 62-77). Lengwil: Libelle.

Rietz, C., Koll, S.v., Franke, S. (2013). RTI in der Methodenausbildung von Sonderpädagoginnen und Sonderpädagogen. Erste Implemtentierungsschritte und Evaluationsergebnisse. *Empirische Sonderpädagogik, 5 (4*), 374-384.

Rispens, J. & van Yperen, T.A. (1990). The identification of specific reading disorders: Measuring a severe discrepancy. In G. T. Pavlidis (Ed.): *Perspectives of Dyslexia. Vol. 2: Cognition, Language and Treatment*. Chichester: John Wiley & Sons.

Roebers, C. M. & Zoelch, C. (2005). Erfassung und Struktur des phonologischen und visuell-räumlichen Arbeitsgedächtnisses bei 4-jährigen Kindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 37 (3)*, 113-121.

Roos, J. & Schöler, H. (2007). *Zur Wirkung des Trainings der phonologischen Bewusstheit im Vorschulalter auf den Schriftspracherwerb.* Abschlussbericht des Projektes EVES.

Röhr, H. (1978). *Voraussetzungen zum Erlernen des Lesen und Rechtschreibens*. Münster: Dissertation.

Roseth, C. J., Missall, K. N. & McConnell S. R. (2012). Early literacy individual growth and development indicators (EL-IGDIs): Growth trajectories using a large internet-based sample. *Journal of School Psychology, 50*, 483-501.

Rost, D. H. & Hartmann, A. (1992). Lesen, Hören und Verstehen. *Zeitschrift für Psychologie, 200*, 345-361.

Roth, E. (1999). *Prävention von Lese- und Rechtschreibschwierigkeiten. Evaluation einer vorschulischen Förderung der phonologischen Bewußtheit und der Buchstabenkenntnis*. Frankfurt a. M. (u. a.).: Lang.

Roth, E. & Schneider, W. (2002). Langzeiteffekte einer Förderung der phonologischen Bewusstheit und der Buchstabenkenntnis auf den Schriftspracherwerb. *Zeitschrift für Pädagogische Psychologie, 16 (2),* 99-107.

Roth, F. P., Speece, D. L. & Cooper, D.H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research, 95(5),* 259-272.

Rothe, E. (2008). Effekte vorschulischen und schulischen Trainings der phonologischen Bewusstheit auf den Schriftspracherwerb. *Wortspiegel: Fachzeitschrift der LOS, 02/08*, 20 – 23.

Runte, M. (o.J.). *Missing Values: Konzepte und statistische Literatur*. URL: http://www.runte.de/matthias/publications/missingvalues.pdf (Stand: 27.07.2016).

Scarborough, H. S. (1991). Early syntactic development of dyslexic children. *Annals of Dyslexia, 41*, 207-220.

Schabmann, A., Schmidt, B. M., Klicpera, C., Gasteiger-Klicpera, B. & Klingebiel, K. (2009). Does systematic reading instruction impede prediction of reading a shallow orthography? *Psychology Science Quarterly, 51 (3),* 315-338.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D. & Foorman, B. R. (2004). Kindergarten prediction of reading skills: a longitudinal comparative analysis. *Journal of Educational Psychology, 96 (2),* 265-282.

Scheerer-Neumann, G. (1979). *Intervention bei Lese-Rechtschreibschwäche*. Bochum: Kamp.

Scheerer-Neumann, G. (1981a). Prozeßanalyse der Leseschwäche. In R. Valtin, U. Jung & G. Scheerer-Neumann (Hrsg.): *Legasthenie in Wissenschaft und Unterricht* (S. 183-209), Darmstadt: Wissenschafltiche Buchgesellschaft.

Scheerer-Neumann, G. (1981b). The utilization of intraword structure in poor readers: Experimenta evidence an a training program. *Psychological Research, 43*, 155-178.

Scheerer-Neumann, G. (1988). *Rechtschreibtraining mit rechtschreibschwachen Hauptschülern auf kognitionspsychologischer Grundlage*. Eine empirische Untersuchung. Opladen: Westdeutscher Verlag.

Scheerer-Neumann, G. (1989a). Entwicklungsprozesse beim Lesenlernen: Eine Fallstudie. In M. Beck (Hrsg.): *Schriftspracherwerb - Lese-Rechtschreibschwäche* (S. 15-38). Tübingen: Deutsche Gesellschaft für Verhaltenstherapie.

Scheerer-Neumann, G. (1989b). Lese-Rechtschreibschwäche im Kontext der Entwicklung. In I. Naegele & R. Valtin (Hrsg.): *LRS in den Klassen 1 - 10* (S. 58-77). Weinheim: Beltz.

Scheerer-Neumann, G. (1993). LRS und Legasthenie. Begriffliche Klärungen und inhaltliche Implikationen. In M. Naegele & R. Valtin (Hrsg.): *LRS in den Klassen 1-10* (S. 21-30). Weinheim (u.a.), (3. unveränd. Aufl.): Beltz.

Scheerer-Neumann, G. (1997a). Rechtschreibschwäche im Kontext der Entwicklung. In M. Naegele & R. Valtin (Hrsg.): *LRS in den Klassen 1-10* (S. 58-77). Weinheim (u.a.), (3. unveränd. Aufl.): Beltz.

Scheerer-Neumann, G. (1997b). Lesen und Leseschwierigkeiten. In F. E. Weinert (Hrsg.): *Psychologie des Unterrichts und der Schule* (S. 279-325). Göttingen: Hogrefe.

Scheerer-Neumann, G. (1998a). Schriftspracherwerb: „The State of the Art" aus psychologischer Sicht. In L. Huber, G. Kegel, A. Speck-Hamdan (Hrsg.): *Einblicke in den Schriftspracherwerb*. Braunschweig: Westermann.

Scheerer-Neumann, Gerheid (1998b), Stufenmodelle des Schriftspracherwerbs – Wo stehen wir heute? In H. Balhorn (Hrsg.) et al.: *Schatzkiste Sprache 1, von den Wegen der Kinder in die Schrift*, Beiträge zur Reform der Grundschule Bd. 104 (S. 54-62). Frankfurt:

Scheerer-Neumann, G. (2006). Das Lesen lernen (wollen). Wie das Eintauchen in die Welt der Buchstaben gelingt. In: *Grundschulunterricht, 5, 16-20.*

Scheerer-Neumann, G. (2004). Lese-Rechtschreib-Schwäche. Wo stehen wir heute? In G. Thomé (Hrsg.): *Lese-Rechtschreib-Schwierigkeiten (LRS) und Legasthenie (S. 22-40).* Weinheim und Basel, (2. Aufl.).: Beltz.

Scheerer-Neumann, G. (2008). Frühe Rechtschreibförderung zur Vorbeugung von Rechtschreibschwäche. In J.Borchert, B. Hartke & P. Jogschies (Hrsg.): *Frühe Förderung entwicklungsauffälliger Kinder und Jugendlicher*. Stuttgart: Kohlhammer.

Schenk-Danzinger, L. (1961). Probleme der Legasthenie. *Schweizer Zeitschrift für Psychologie, 20*, 29-48.

Schenk-Danzinger, L. (1975). *Handbuch der Legasthenie im Kindesalter*. Weinheim: Beltz.

Schlee, J. (1976). *Legasthenieforschung am Ende?* München: Urban und Schwarzenberg.

Schlee, J. (2008). 30 Jahre ‚Förderdiagnostik' – eine kritische Bilanz. *Zeitschrift für Heilpädagogik, 4*, 122-131.

Schmid-Barnkow, I. (2003). Das Unbewusste der phonologischen Bewusstheit. *Grundschule, 9*, 38-40.

Schneider, W. (1980). *Bedingungsanalysen des Recht-Schreibens*. Weinheim: Beltz.

Schneider, W. (1989). Möglichkeiten der frühen Vorhersage von Leseleistungen im Grundschulalter. *Zeitschrift für Pädagogische Psychologie, 3 (2),* 157-168.

Schneider, W. (1993). Introduction: The early prediction of reading and spelling. *European Journal of Psychology of Education, 3*, 199-203.

Schneider, W. (1997). Rechtschreiben und Rechtschreibschwierigkeiten. In F. E. Weinert (Hrsg.): *Enzyklopädie der Psychologie. Serie Pädagogische Psychologie, Bd. 3: Psychologie des Unterrichts und der Schule* (S. 327-363). Göttingen: Hogrefe.

Schneider, W. (2001a). Die Würzburger Längsschnittstudien zur frühen Prävention von Lese-Rechtschreibschwäche. In G. Schulte-Körne (Hrsg.). *Legasthenie: erkennen, verstehen, fördern* (S. 214-219). Bochum: Winkler.

Schneider, W. (2001b). Das Konzept der phonologischen Bewusstheit und seine Bedeutung für den Schriftspracherwerb. In Akademie für Lehrerfortbildung und Personalführung (Hrsg.): *Lese-Rechtschreib-Schwierigkeiten: Diagnose – Förderung – Materialien* (S. 81 – 90), Donauwörth (2.erw.Aufl): Auer.

Schneider, W. (2008). Entwicklung und Erfassung der Rechtschreibkompetenz im Jugend- und Erwachsenenalter. In W. Schneider, H. Marx & M. Hasselhorn (Hrsg.): *Diagnostik von Rechtschreibleistungen und -kompetenz. Tests und Trends N. F*. Band 6. (S. 145-157). Göttingen (u.a.): Hogrefe.

Schneider, W., Blanke, I, Faust V. & Küspert, P. (2011). *WLLP-R Würzburger Leise Leseprobe – Revision*. Ein Gruppentest für die Grundschule. Göttingen: Hogrefe.

Schneider, W., Brügelmann, H. & Kochan, B. (1990). Lesen- und Schreibenlernen in einer neuen Sicht: Vier Perspektiven auf den Stand der Forschung. In H. Brügelmann & H. Ballhorn (Hrsg.): *Das Gehirn, sein Alfabet und andere Geschichten* (S. 220-244). Konstanz: Faude

.

Schneider, A., Esser, G. & Sommerfeld, E. (2003). EEG-Kohärenzanalyse zur Untersuchung eines Automatisierungsdefizits bei Lese-Rechtschreibstörung – eine Pilotstudie. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 31 (4)*, 255-266.

Schneider, W. & Küspert, P. (2003). Frühe Prävention der Lese-Rechtschreib-Störung. In W. von Suchodoletz (Hrsg.): *Therapie der Lese-Rechtschreib-Störung (LRS). Traditionelle und alternative Behandlungsmethoden im Überblick* (S. 108-128). Stuttgart: Kohlhammer.

Schneider, W. & Küspert, P. (2004). Förderung von phonologischer Bewusstheit. In G.W. Lauth, M. Grünke & J.C. Brunstein (Hrsg.): *Intervention bei Lernstörungen. Förderung, Training und Therapie in der Praxis* (S. 46-54). Göttingen: Hogrefe.

Schneider, W., Marx, P. & Weber, J. (2002). Auffälligkeiten in der Sprachentwicklung. Risikofaktoren für Lese- und Rechtschreibschwierigkeiten. *Kinderärztliche Praxis, 3*, 186 – 194.

Schneider, W. & Näslund, J. C. (1992). Cognitive prerequisites of reading and spelling. A longitudinal approach. In A. Demetriou, M. Shayer & A. Efklides (Hrsg.): *Neo-piagetian theories of cognitive development*. London, New York: Routledge.

Schneider, W. & Näslund, J. C. (1993). The impact of early metalinguistic competencies and memory capacity on reading and spelling in elementary school: Results of the Munich Longitudinal Study on the Genesis of Individual competencies (LOGIC). *European Journal of Psychology of Education, 8*, 273-288.

Schneider, W. & Näslund, J. C. (1997). The early prediction of reading and spelling: Evidence form the Munich longitudinal study on the genesis of individual competencies. In C. K. Leong & R. M. Joshi (Eds.): *Cross-language studies of learning to read and spell* (p. 139-159). Dordrecht: Kluwer.

Schneider, W. & Näslund, J. C. (1999). The impact of early phonological processing skills on reading and spelling in school. In F. E. Weinert & W. Schneider (Eds.): *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (p. 126-147). Cambridge: Cambridge University Press.

Schneider, W., Näslund, J. C. & Weinert, F. E. (1989). Prerequisites of school achievement. In F. E. Weinert & W. Schneider: *The Munich Longitudinal Study on Genesis of Individual Competencies (LOGIS).* Report No. 5: Results of wave three. (p. 155-182). München: Max-Planck Institut für Psychologische Forschung.

Schneider, W., Marx, P. & Weber, J. (2002). Auffälligkeiten in der Sprachentwicklung. Risikofaktoren für Lese-Rechtschreib-Schwierigkeiten. *Kinderärztliche Praxis, 3*, 186-194.

Schneider, W., Roth, E. & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: a comparison of three kindergarten intervention programs. *Journal of Educational Psychology, 92 (2)*, 284-295.

Schneider, W., Roth, E., Küspert, P. & Ennemoser, M. (1998). Kurz- und langfristige Effekte eines Trainings der sprachlichen (phonologischen) Bewußtheit bei unterschiedlichen Leistungsgruppen: Befunde einer Sekundäranalyse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 30 (1),* 26-39.

Schneider, W., Roth, E. & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children-at-risk for dyslexia: A comparisons of three kindergarten intervention programs. *Journal of Educational Psychology, 92,* 284-295.

Schneider, W. & Stefanek, J. (2007). Entwicklung der Rechtschreibleistung vom frühen Schul- bis zum frühen Erwachsenenalter. Längsschnittliche Befunde der Münchner LOGIK-Studie. *Zeitschrift für Pädagogische Psychologie, 21 (1)*, 77-82.

Schneider, W., Stefanek, J. & Dotzler, H. (1997). Erwerb des Lesens und des Rechtschreibens: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert (Hrsg.): *Entwicklung im Grundschulalter* (S. 113-129). Weinheim: Beltz.

Schneider, W., Visé, M., Reimers, P. & Blaesser, B. (1994). Auswirkungen eines Trainings der sprachlichen Bewußtheit auf den Schriftspracherwerb in der Schule. *Zeitschrift für Pädagogische Psychologie, 8 (3/4),* 177-188.

Schnitzler, C. (2008). *Phonologische Bewusstheit und Schriftspracherwerb*. Stuttgart, New York: Thieme.

Schöler, H. & Brunner, M. (2008). *HASE. Heidelberger Auditives Screening in der Einschulungsuntersuchung*. Wertingen (2. überarb. und erw. Aufl.): Westra.

Schründer-Lenzen, A. (2009). *Schriftsprache und Unterricht. Bausteine professionellen Handlungswissens.* Wiesbaden (3. Aufl.): Vs Verlag.

Schuchardt, K., Kunze, J., Grube, D. & Hasselhorn, M. (2006). Arbeitsgedächtnisdefizite bei Kindern mit schwachen Rechen- und Schriftsprachleistungen. *Zeitschrift für Pädagogische Psychologie, 20 (4),* 261-268.

Schuck, K.-D., Eggert, D & Raatz, U. (1975). *Columbia Mental Maturity Scale CMM 1*-3. Weinheim: Beltz Test Gesellschaft.

Schulte-Körne, G. (2004). Lese-Rechtschreib-Störung. Symptomatik, Diagnostik, Verlauf, Ursachen und Förderung. In G. Thomé (Hrsg.): *Lese-Rechtschreib-Schwierigkeiten (LRS) und Legasthenie* (S. 64-86). Weinheim und Basel: (2. Aufl.)Beltz.

Schulte-Körne, G. (2006). Lerntheoretisch begründete Therapieverfahren bei der Lese-Rechtschreibstörung. In W. von Suchodoletz (Hrsg.): *Therapie der Lese-Rechtschreib-Störung (LRS). Traditionelle und alternative Behandlungsmethoden im Überblick* (S. 33-57). Stuttgart: (2. überarb. und erw. Aufl.) Kohlhammer.

Schulte-Körne, G., Deimel, W., Bartling, J. & Remschmidt, H. (1998). Die Bedeutung der auditiven Wahrnehmung und der phonologischen Bewußtheit für die Lese-Rechtschreibschwäche. *Sprache, Stimme, Gehör, 22*, 25-30.

Schulte-Körne, G., Remschmidt, H. & Hebebrand, J, (1993). Zur Genetik der Lese-Rechtschreibstörung. *Zeitschrift für Kinder- und Jugendpsychiatrie, 21*, 242-252.

Schulte-Körne, G., Schäfer, J., Deimel, W. & Remschmidt, H. (1997). Das Marburger Eltern-Kind-Rechtschreibtraining. *Zeitschrift für Kinder- und Jugendpsychiatrie, 25*, 151-159.

Schulte-Körne, G., Warnke, A. & Remschmidt, H. (2006). Zur Genetik der Lese-Rechtschreibschwäche. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 34 (6)*, 435-444.

Schumacher, E. (2007). Zum Übergang vom Kindergarten in die Grundschule – Die „neue Schuleingangsphase“. In M. Plieninger & E. Schumacher (Hrsg.): *Auf den Anfang kommt es an – Bildung und Erziehung im Kindergarten und im Übergang zur Grundschule*. Schwäbisch Gmünd: Rektorat der Pädagogischen Hochschule Schwäbisch Gmünd.

Schwarz, J. (2012). *Master of Science in Pflege. Modul Statistik*. Berner Fachhochschule Gesundheit. URL: http://www.schwarzpartners.ch/einfuehrung_statistik/Lehreinheit%205%20Regression.pdf (Stand: 02.08.2016).

Schwenck, C. & Schneider, W. (2003). Einflussfaktoren für den Zusammenhang von Rechen – und Schriftsprachleistungen im frühen Grundschulalter. *Kindheit und Entwicklung, 12 (4),* 212-221.

Scriven, M. (1967). *The Methodology of Evaluation.* Washington, CD: American Educational Research Association.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed developmental model of wordrecognition and naming. *Psychological Review, 96*, 523-568.

Seymour, P. H. K., Aro, M. & Erskine, J. M. (2003): Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143 – 174.

Shankweiler, D. Liberman, I. Y., Mark, L. S., Fowler, C. A. & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human learning and Memory, 5 (6),* 531-545.

Shankweiler, D. (1999). Words to meanings. *Scientific Studies of Reading, 3,* 113-127.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santaro, L. E. & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests. Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24 (1),* 19- 35.

Shinn, M. R. (1998). *Curriculum-based assessment. Assessing special children*. New York: Guilford Press.

Shinn, M. R. & Bamonto, S. (1998). Advanced applications of curriculum- based measurement: 'big ideas' and avoiding confusion, chapter 1. In M. R. Shinn (Ed.): *Curriculum-based assessment. Assessing special children. New York: Guilford Press.*

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D. & Collins V. L. (1992). Curriculum-Based Measurement of Oral Reading Fluency: A Confirmatory Analysis of its Relation to Reading. *School Psychology Review, 21 (3),* 459-479.

Shinn, M. R., Tindal, G. A., Spira, D., & Marston, D. (1987). Practice of learning disabilities as social policy. *Learning Disabilities Quarterly, 10,* 17–28.

Siegel, L. S. (1988a). Evidence that IQ scores are irrelevant to the definitions and analysis of reading disabilitiy. *Canadian Journal of Psychology, 42*, 201-215.

Siegel, L. S. (1988b). Definitional and theoretical issues research on learning disabilities. *Journal of Learning Disabilities, 21,* 264-266.

Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 469-479.

Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities, 2*5, 618-629.

Siegel, L. S. & Ryan, E. B. (1988). Development of grammatical sensitivity, phonological and short-term-memory skills in normally achieving and learning disabled children. *Developmental Psychology, 24*, 28-37.

Silberglitt, B. & Hintze, J. M. (2007). How much growth can we expect? A conditional analysis of R-CBM growth rates by level of performance. *Exceptional Children, 74*, 71-84.

Simanowski-Schulz, S. (2014). *Einfluss vorschulischer exekutiver Funktionen auf die Entwicklung von Lesen, Rechtschreiben und Rechnen in der Schuleingangsphase*. URL: http://geb.uni-giessen.de/geb/volltexte/2014/11105/ (Stand: 08.08.2016).

Simos, Fletcher, Bergmann, Breier, Foorman, Castillo, Davis, Fitzgerald & Papanicolaou (2002). Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology, 58*, 1203-1213.

Singleton, C. H. (1988). The early diagnosis of developmental dyslexia. *Support for Learning, 3 (2),* 108-121.

Sirch, K. (1975). *Der Unfug mit der Legasthenie*. Stuttgart: Klett.

Skowronek, H. & Marx, H. (1989a). The Bielefeld Longitudinal Study on early identification risks in learning read and write: Theoretical background and first results. In M. Brambring, F. Lösel, H. Skowronek: *Children at risk: Assessment, longitudinal research, and intervention* (p. 268-295). New York: de Gruyter.

Skowronek, H. & Marx, H. (1989b). Die Bielefelder Längsschnittstudie zur Früherkennung von Risiken der Lese-Rechtschreibschwäche: Theoretischer Hintergrund und erste Befunde. *Heilpädagogische Forschung, 15 (1)*, 38-49.

Snowling, M. van Wagtendonk, B. & Stafford, C. (1988). Object naming deficits in developmental dyslexia. *Journal of Research in Reading, 11*, 67-85.

Snowling, M. J., Gallagher, A. & Frith, U. (2003). Family risk of dyslexia is continuous: Individual differences in the precursors of reading skill. *Child Development, 74 (2)*, 358-373.

Solga, H. & Powell, J. (2006). Gebildet – Ungebildet. In S. Lessenich & F. Nullmeier (Hrsg.): *Deutschland – eine gespaltene Gesellschaft* (S. 175-190). Frankfurt a. M.: Campus.

Sonntag, W. (2010). Fördert induktives Denken die Gedächtnisstrategie des Kategorisierens bei lernbehinderten Sonderschülern? *Empirische Sonderpädagogik, 1*, 5-21.

Souvignier, E. & Förster, N. (2011). Effekte prozessorientierter Diagnostik auf die Entwicklung der Lesekompetenz leseschwacher Viertklässler. *Empirische Sonderpädagogik 3 (3),* S. 243-255.

Souvignier, E., Förster, N. & Salaschek, M. (2014). quop: ein Ansatz internet-basierter Lernverlaufsdiagnostik und Testkonzepte für Mathematik und Lesen. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik - Tests und Trends N.F. Band 12*. (S. 239- 256). Göttingen: Hogrefe.

Souvignier, E., Förster, N. & Schulte, E. (2014). Wirksamkeit formativen Assessments – Evaluation des Ansatzes der Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik - Tests und Trends N.F. Band 12* (S. 221-237). Göttingen: Hogrefe.

Spearman, C. (1927). *The abilities of man*. London: MacMillan.

Speck-Hamdan, A. (1998). Individuelle Zugänge zur Schrift. Schriftspracherwerb aus konstruktivistischer Sicht. In L. Huber, G. Kegel, A. Speck-Hamdan (Hrsg.): *Einblicke in den Schriftspracherwerb*. Braunschweig: Westermann.

Squires, G. & McKeown, S. (2006). *Supporting children with dyslexia*. London, New York: Continuum.

Stackhouse, J. & Wells, B. (1997). *Childrens speech and literacy difficulties. A psycholinguistic framework*. London: Whurr.

Stackhouse, J. & Wells, B. (2001). *Childrens speech and literacy difficulties. Identification and intervention*. London: Whurr.

Stanovich, K. E. (1982). Individual differences in the cognitive process of reading I: Word decoding. *Journal of Learning Disabilities, 15*, 485-493.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden variety poor reader: The phonological-core variable difference model. *Journal of Learning Disabilities, 15*, 590-612.

Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us estray? *Reading Research Quartely, 26*, 7-29.

Stanovich, K. E. (1994). Are discrepancy-based definitions of dyslexia empirically defensible? In K. van den Bos, L. S. Siegel, D. J. Bakker & D. L. Share (Eds.): *Current directions in dyslexia research*. Lisse: Swets & Zeitlinger.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York: Guilford.

Stanovich, K. E. & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24-53.

Stecker, P. M. & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice (Lawrence Erlbaum), 15 (3*), 128-134.

Steinbrink, C., Schwanda, S., Vogt, K. (2008). Zusammenhänge zwischen kognitiven Variablen und Lese-Rechtschreibleistungen bei Erstklässlern mit Schwierigkeiten im Rechtschreiben. *Nervenheilkunde, 7*, 644-650.

Steinbrink, C., Schwanda, S., Klatte, M. & Lachmann, T. (2010). Sagen Wahrnehmungsleistungen zu Beginn der Schulzeit den Lese-Rechtschreiberfolg in Klasse 1 und 2 voraus? – Zur prognostischen Validität der Differenzierungsproben 1 und 2 nach Breuer und Weuffen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42 (4)*, 188-200.

Sternberg, R. J., Grigorenko, E. L. & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly, 47*, 1-41.

Stock, C. (2009). Der Zusammenhang zwischen phonologischer Bewusstheit und Leseleistung bei Grundschulkindern. In W. Lenhard & W. Schneider (Hrsg.). *Diagnostik und Förderung des Leseverständnisses. Tests und Trends, N. F*. Band 7 (S. 81-95). Göttingen (u.a.): Hogrefe.

Stock, C., Marx, P. & Schneider, W. (2003). *BAKO 1-4. Basiskompetenzen für Lese-Rechtschreib-leistungen. Ein Test zur Erfassung der phonologischen Bewusstheit vom ersten bis vierten Grundschuljahr*. Göttingen: Beltz Test GmbH.

Stock, C., Schneider, W. (2008a). *Deutscher Rechtschreibtest für das erste und zweite Schuljahr (DERET 1-2).* Bern, Huber.

Stock, C., Schneider, W. (2008b). *Deutscher Rechtschreibtest für das dritte und vierte Schuljahr (DERET 3-4)*. Bern, Huber.

Stocker, H. (2015). *Methoden der empirischen Wirtschaftsforschung. Kapitel 10: Multikollinearität.* URL: https://www.uibk.ac.at/econometrics/einf/kap08.pdf (Stand: 17.8.2016).

Strathmann, A. M. (2014). Lernverlaufsdiagnostik Mathematik für zweite bis vierte Klassen (LVD-M). In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik (Tests und Trends N.F. Band 12)* (S. 203-220). Göttingen: Hogrefe.

Strathmann, A.M. & Klauer, K.J. (2008). Diagnostik des Lernverlaufs. Eine Pilotstudie am Beispiel der Entwicklung der Rechtschreibkompetenz. *Sonderpädagogik, 38 (1),* 5–24.

Strathmann, A.M. & Klauer, K.J. (2010). Lernverlaufsdiagnostik: Ein Ansatz zur längerfristigen Lernfortschrittsmessung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 42 (2),* 111–122.

Strathmann, A.M. & Klauer, K.J. (2012). *Lernverlaufsdiagnostik Mathematik. LVD-M.* Göttingen: Hogrefe Verlag.

Strathmann, A., Klauer, K. J. & Greisbach, M. (2010). Lernverlaufsdiagnostik – Dargestellt am Beispiel der Entwicklung der Rechtschreibkompetenz in der Grundschule. *Empirische Sonderpädagogik, 1*, 64-77.

Strehlow, U. (1994). Katamnestic studies of dyslexia. *Acta paedopsychiatrica, 56*, 219-228.

Strehlow, U. (1998). Der Verlauf der umschriebenen Lese-Rechtschreibschwäche. *Sprache, Stimme, Gehör, 22,* 31-33.

Strehlow, U. & Haffner, J. (2002). Definitionsmöglichkeiten und sich daraus ergebende Häufigkeit der umschriebenen Lese- bzw. Rechtschreibstörung – theoretische Überlegungen und empirische Befunde an einer repräsentativen Stichprobe junger Erwachsener. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 30 (2*), 113-126.

Strehlow, U., Kluge, R., Moller, H. & Haffner, J. (1992). Der langfristige Verlauf der Legasthenie über die Schulzeit hinaus: Katamnesen aus einer kinderpsychiatrischen Ambulanz. *Zeitschrift für Kinder- und Jugendpsychiatrie, 20,* 254-265.

Suchodoletz, W. v., Berwanger, D. & Mayer, H. (2004). Die Bedeutung auditiver Wahrnehmungs-schwächen für Pathogenese der Lese-Rechtschreibstörung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 32 (1)*, 19-27.

Suchodoletz, W. von (2006). Spannungsfeld zwischen etablierten und alternativen Behandlungs-verfahren. In W. von Suchodoletz (Hrsg.): *Therapie der Lese-Rechtschreib-Störung (LRS). Traditionelle und alternative Behandlungsmethoden im Überblick* (S.15-32). Stuttgart (2. überarb. und erw. Aufl.): Kohlhammer.

Suchodoletz, W. (2007). Möglichkeiten und Grenzen von Prävention. In W. Suchodoletz (Hrsg.): *Prävention von Entwicklungsstörungen*. Göttingen (u.a.): Hogrefe.

Suhr, K. (2005). *Förderung rechtschreibschwacher Schülerinnen und Schüler. Konzeption, Implementation und Evaluation eines multidimensionalen Förderprogramms*. Hamburg: Dr. Kovač.

Svensson, I. & Jacobson, C. (2005). How persistent are phonological difficulties? A longitudinal study of reading retarded children. *Dyslexia, 12*, 3-20.

Swanson, H.L. & Berninger, V.W. (1996). Individual differences in children´s working memory and writing skill. *Journal of Experimental Child Psychology, 63*, 358-396.

Tacke, G., Nock, H. & Staiber, W. (1987). Rechtschreibförderkurse in der Schule. Wie erfolgreich sind sie, und welche Faktoren tragen zu Leistungsverbesserungen bei? *Zeitschrift für Pädagogische Psychologie, 1*, 45-52.

Tacke, G. (2008). Frühe Leseförderung zur Vorbeugung von Leseschwäche. In J. Borchert, B. Hartke & P. Jogschies (Hrsg.): *Frühe Förderung entwicklungsauffälliger Kinder und Jugendlicher*. Stuttgart: Kohlhammer.

Tarazona, W. (2015). *Fortgeschrittene Statistik – Logistische Regression*. Zentrale Methodenlehre, Europa Universität Flensburg (unveröffentlichte PP).

Tarelli, I., Valtin, R., Bos, W., Bremerich-Vos, A. & Schwippert, K. (2012). IGLU 2011. Wichtige Ergebnisse im Überblick. In W. Bos, I. Tarelli, A. Bremerich-Vos & K. Schwippert (Hrsg.): *IGLU 2011. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 11-25). Münster: Waxmann.

Thole, W., Fölling-Albers, M. & Roßbach, H.-G. (2008). Die „Pädagogik der Kindheit" im Fokus der Wissenschaften. In W. Thole, H.-G. Roßbach, M. Fölling-Albers & R. Tippelt (Hrsg.): *Bildung und Kindheit. Pädagogik der frühen Kindheit in Wissenschaft und Lehre* (S. 17-30). Opladen & Farmington Hills: Verlag Barbara Budrich.

Thomas, K., Schulte-Körne, G. & Hasselhorn, M. (2015). Stichwort – Entwicklungsstörungen schulischer Fertigkeiten. *Zeitschrift für Erziehungswissenschaften, 18,* 431-451.

Thorndike, R. L. (1963). *The concepts of over and under achievement*. New York: Columbia University Bureau of Publications.

Tietze, W. (2010). Bildungspsychologie des Vorschulbereichs. In C. Spiel, B. Schober, P. Wagner & R. Reimann (Hrsg.). *Bildungspsychologie* (S. 50-69). Göttingen (u.a.): Hogrefe.

Tiu, R. D., Thompson, L. A. & Lewis, B. A. (2003). The role of IQ in a component model of reading. *Journal of Learning Disabilities, 36*, 424-436.

Torgesen, J. K., Kistner, J. A. & Morgan, S. (1987). Component processes in working memory In J. Borkowski & J. Day (Eds.): *Memory and cognition in special children: Comparative approaches*. (p. 49-86). Norwood, NJ: Albex.

Torgesen, J. K. & Bryant, B. (1994). *Test Of Phonology Awareness: Examiner's Manual*. Austin: Pro-Eds.

Toth, G. & Siegel, L. S. (1994). A critical evaluation of the IQ-achievement discrepancy based definition of dyslexia. In K. van den Bos, L. S. Siegel, D. J. Bakker & D. L. Share (Eds.): *Current directions in dyslexia research*. Lisse: Swets & Zeitlinger.

Tramontana, M. G., Hooper, S. R. & Selzer, S. C. (1988). Research on the preschool prediction of later academic achievement: A review. *Developmental Review, 8,* 89-146.

Treiman, R. & Zukowski, A. (1991). Levels of phonological awareness. In S. Brady & D. Shankweiler (Eds.). *Phonological processing in literacy. A tribute to Isabelle Liberman* (p. 67-83). Hillsdale, N. J.: Erlbaum.

Trempler, D. (1976). *Legasthenie – neue Wege der Heilung*. Freiburg: Herder.

Treutlein, A., Roos, J. & Schöler, H. (2007). *Zur prognostischen Validität des Heidelberger Auditiven Screenings in der Einschulungsdiagnostik HASE - Abschlussbericht des Projektes EVEr*. Heidelberg: Pädagogische Hochschule. URL: http://www.ph-heidelberg.de/wp/Schoeler /seiten/ever.htm (Stand: 16.01.2013).

Tröster, H.(2009). *Früherkennung im Kindes- und Jugendalter. Strategien bei Entwicklungs-, Lern- und Verhaltensstörungen*. Göttingen (u.a.): Hogrefe.

Tunmer, W. E. & Hoover, W. A. (1992). Cognitive and linguistic factors in learning to read. In P. B. Gough, L. E. Ehri, R. Treiman (Eds.): *Reading Acquisition* (p. 175-214). Hillsdale New York: Lawrence Erlbaum Associates.

Tunmer, W. E. & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. Journal of *Educational Psychology, 77*, 417-427.

Universität Zürich (2016a). *Methodenberatung – Einfaktorielle Varianzanalyse (mit Messwiederholung).* URL: http://www.methodenberatung.uzh.ch/datenanalyse/unterschiede/zentral/evarianzmessw.html (Stand: 30.07.2016).

Universität Zürich (2016b). *Methodenberatung –Logistische Regressionsanalyse.* URL: http://www.methodenberatung.uzh.ch/datenanalyse/zusammenhaenge/lreg.html (Stand: 31.07.2016).

Valerius, S. & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungen im Nested-Factor-Modell. Zeitschrift für Pädagogische Psychologie, 29, 101-108.

Valtin, R. (1972). *Empirische Untersuchungen zur Legasthenie*. Hannover: Schroedel.

Valtin, R. (1974a). *Rechtschreibleistung und Rechtschreibfehler*. Fernstudienlehrgang Legasthenie, Studienbegleitheft 1, 58-86.

Valtin, R. (1974b). *Legasthenie – Theorien und Untersuchungen*. Weinheim: Beltz.

Valtin, R. (1975). Ursachen der Legasthenie: Fakten oder Artefakte? *Zeitschrift für Pädagogik, 21*, 407-418.

Valtin, R. (1978). Legasthenie – ein überholtes Konzept? In Deutsche Forschungsgemeinschaft (Hrsg.): *Zur Lage der Legasthenieforschung* (S. 27-33). Boppard: Boldt.

Valtin, R. (1988). Schriftspracherwerb als Entwicklungsprozess. *Grundschule, 20*, 12-16.

Valtin, R. (2004). Das Konstrukt Legasthenie – Wem schadet es? Wem nützt es? In G. Thomé (Hrsg.): *Lese-Rechtschreib-Schwierigkeiten (LRS) und Legasthenie* (S. 56-64). Weinheim und Basel (2. Aufl.).: Beltz.

Valtin, R. (2006). Der medizinische Ansatz der Legasthenie und seine Problematik. In B. Hofmann & A. Sasse (Hrsg.): *Legasthenie. Lese-Rechtschreibstörungen oder Lese-Rechtschreibschwierigkeiten? Theoretische Konzepte und praktische Erfahrungen mit Förderprogrammen* (S. 44-60). Frankfurt: Deutsche Gesellschaft für Lesen und Schreiben.

Valtin, R. (2010). *Phonologische Bewusstheit – eine notwendige Voraussetzung beim Lesen- und Schreibenlernen?* URL: http://www.leseforum.ch/myUploadData%5Cfiles%5C2010_2_Valtin_PDF.pdf (Stand: 05.10.2010)

Valtin, R. & Naegele, I. (2001). Lesen lernt man nur durch Lesen (sinnvoller Texte). In R. Valtin & I. Naegele (Hrsg.): *LRS - Legasthenie in den Klassen 1-10*, Band1 (S. 203-209). Weinheim (5. neu ausgest. Aufl.): Beltz.

VanDerHeyden, A. M. (2005). Intervention-driven assessment practices in early childhood/early intervention: Measuring what is possible rather than what is present. *Journal of Early Intervention, 28*, 28-33.

VanDerHeyden, A. M. & Snyder, P. S. (2006). Integrating frameworks from early childhood intervention and school psychology to accelerate growth for all young children. *School Psychology Review, 35 (4),* 519-534.

VanDerHeyden, A. M., Witt, J. C., Naquin, G. & Noell, G. (2002). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30 (3),* 363-382.

Van der Wissel, A. & Zegers, F. E. (1985). Reading retardation revisited. *British Journal of Developmental Psychology, 3,* 3-9.

Vaughn, S., Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18(3),* 137-146.

Vellutino, F. R. (1979). *Dyslexia: Theory and Research*. Cambridge, MA: MIT Press.

Vellutino, F. R. (1980). Dylexia: Perceptual deficiency or perceptual inefficiency. In J. F. Kavanagh & R. L. Venezky (Eds.): *Orthography, reading and dyslexia* (pp. 251-270). Baltimore: University Park Press.

Vellutino, F. R. (1990). *Dyslexia. Theory and research*. Cambridge: The MIT Press.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J. & Scanlon, D. (2004). Specific reading disabilities (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45 (1),* 2-40.

Vellutino, F. R. & Scanlon, D. M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from longitudinal and experimental study. *Merrill-Palmer Quarterly, 33*, 321-363.

Vellutino, F. R., Scanlon, D. M., Sispay, E. R. Small, S. G., Pratt, A., Chen, R. & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Learning Disabilities, 33 (3)*, 223-238.

Vellutino, F. R., Scanlon, D. M. & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers. *Journal of Learning Disabilities, 33,* 223-238.

Vellutino, F. R., Scanlon, D. M. & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology, 59*, 76-123.

Vellutino, F. R., Scanlon, D. M. & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology, 59,* 76-123.

Vellutino, F. R., Steger, J. A. & Kaman, M. (1975). Visual form perception in deficient and normal readers as a function of age and orthographic linguistic familiarity. *Cortex, 11*, 22-30.

Voß, S. (2013). Curriculumbasierte Messverfahren im mathematischen Erstunterricht. Zur Güte und Anwendbarkeit einer Adaption US-amerikanischer Verfahren im deutschen Schulsystem. URL: http://rosdok.uni-rostock.de/file/rosdok_disshab_0000001156/rosdok_derivate_0000005279/Dissertation_Voss_2014.pdf (Stand: 03.08.2016).

Voß, S., Blumenthal, Y., Sikora, S., Mahlau, K. Diehl, K. & Hartke, B. (2014). Rügener Inklusionsmodell (RIM) – Effekte eines Beschulungsansatzes nach dem Response to Intervention-Ansatz auf die Rechen –und Leseleistungen von Grundschulkindern. *Empirische Sonderpädagogik, 2,* 114-132.

Voß, S. & Hartke, B. (2014). Curriculumbasierte Messverfahren (CBM) als Methode der formativen Leistungsdiagnostik im RTI-Ansatz. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik (Tests & Trends, NF Bd. 12*, S. 83-99). Göttingen: Hogrefe.

Voß, S., Mahlau, K., Sikora, S., Blumenthal, Y., Diehl, K. & Hartke, B. (2015). Evaluationsergebnisse des Projekts “Rügener Inklusionmodell (RIM) – Präventive und Integrative Schule auf Rügen (PISaR)“ nach vier Schuljahren zum Messzeitpunkt 2014. Rostock: Universität Rostock. URL: http://www.uni-rostock.de/uploads/media/RIM-Evaluationsbericht2015-druck.pdf (Stand: 07.08.2016).

Vukovic, R. K. & Siegel, L. S. (2006). The Double-Deficit Hypothesis: A Comprehensive Analysis of the Evidence. *Journal of Learning Disabilities, 39*, 25-47.

Wagner, R. K. (1988). Causal relation between the development of phonological processing abilities and the acquisition of reading skills: A meta-analysis. *Merrill-Palmer Quarterly, 34*, 261-279.

Wagner, R. & Torgesen, J. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Waligora, K. (2003). Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC), Rezension. *Praxis der Kinderpsychologie und Kinderpsychiatrie 52 (9),* 744-747.

Walter, J. (o.J.). *Berechnung Klassifikationsvalidität*. (nur intern verfügbar)

Walter, J. (1996). *Förderung bei Lese- und Rechtschreibschwäche. Grundlagenforschung, methodische Konsequenzen, Praxisbeispiele und mediendidaktische Anregungen auf der Basis empirischer Forschungsmethoden*. Göttingen (u. a.): Hogrefe.

Walter, J. (1996b). Der Spracherfahrungsansatz für den Anfangsunterricht: Empirisch-experimentell abgesicherte Effekt und didaktisch-methodische Konsequenzen aus den USA. *Sonderpädagogik, 26,* 136-143.

Walter, J. (2000). Lesen und Schreiben. In J. Borchert (Hrsg.): *Handbuch der sonderpädagogischen Psychologie* (S. 380-404). Göttingen (u. a.): Hogrefe.

Walter, J. (2002a). Differenzielle Effekte des Trainings des phonologischen Wissens auf das Lesen- und Schreibenlernen: Ergebnisse der international angelegten Meta-Analyse von Ehri et al. *Heilpädagogische Forschung, 26(1)*, 38-49.

Walter, J. (2002b). „Einer flog übers Kuckucksnest" oder welche Interventionsformen erbringen im sonderpädagogischen Feld welche Effekte? Ergebnisse ausgewählter us-amerikanischer Meta- und Mega-Analysen. *Zeitschrift für Heilpädagogik, 53*, 442-450.

Walter, J. (2005a). *Erklärungsansätze für Legasthenie*. Unterlagen für das Seminar „LRS-Theorie“ im SS2005 an der Universität Flensburg (unveröffentlichtes Manuskript).

Walter, J. (2005b). Diskrepant oder nicht-diskrepant: Ist das noch die Frage? Über eine problematische Untergruppenbildung bei lese-rechtschreibschwachen Kindern und deren diagnostisch-konzeptionelle Überwindung. *Sonderpädagogik, 35,* 63-79.

Walter, J. (2007a). Phonologische Bewusstheit. In J. Walter & F. B. Wember (Hrsg.): *Sonderpädagogik des Lernens. Handbuch Sonderpädagogik*, Band 2 (S. 479-503). Göttingen (u. a.): Hogrefe.

Walter, J. (2007b). Meta- und Megaanalyse als Erkenntnismethoden zur Darstellung von Trainingseffekten bei Schülern mit sonderpädagogischem Förderbedarf. In J. Walter & F. B. Wember (Hrsg.): *Sonderpädagogik des Lernens. Handbuch Sonderpädagogik,* Band 2 (S. 871-896). Göttingen (u. a.): Hogrefe.

Walter, J. (2007c). *Markierung der Programmänderung bei Alyssa*. Unterlagen für das Seminar „LRS-Theorie“ im SS2005 an der Universität Flensburg (unveröffentlichtes Manuskript).

Walter, J. (2008a). Adaptiver Unterricht erneut betrachtet: Über die Notwendigkeit systematischer formativer Evaluation von Lehr- und Lernprozessen und die daraus resultierende Diagnostik und Neudefinition von Lernstörungen nach dem RTI- Paradigma. *Zeitschrift für Heilpädagogik, 59*, 202-215.

Walter, J. (2008b). Curriculumbasiertes Messen (CBM) als lernprozessbegleitende Diagnostik: Erste deutschsprachige Ergebnisse zur Validität, Reliabilität und Veränderungssensibilität eines robusten Indikators zur Lernfortschrittsmessung. *Heilpädagogische Forschung, 34*, 62-79.

Walter, J. (2009a). Eignet sich die Messtechnik „MAZE“ zur Erfassung von Lesekompetenzen als lernprozessbegleitende Diagnostik? *Heilpädagogische Forschung, 35*, 62-75.

Walter, J. (2009b). Theorie und Praxis Curriculumbasierten Messens (CBM) in Unterricht und Förderung. *Zeitschrift für Heilpädagogik, 60*, 162-170.

Walter, J. (2010a). *Lernfortschrittsdiagnostik Lesen (LDL). Ein curriculumbasiertes Verfahren*. Göttingen: Hogrefe.

Walter, J. (2010b). Lernfortschrittsdiagnostik am Beispiel der Lesekompetenz (LDL): Messtechnische Grundlagen sowie Befunde über zu erwartende Zuwachsraten während der Grundschulzeit. *Heilpädagogische Forschung, 36,* 162-176.

Walter, J. (2011a). Die Entwicklung eines auch computerbasiert einsetzbaren Instruments zur formativen Messung der Lesekompetenz. *Heilpädagogische Forschung, 37(3),* 106-126.

Walter, J. (2011b). Die Messung der Entwicklung der Lesekompetenz im Dienste der systematischen formativen Evaluation von Lehr- und Lernprozessen. *Zeitschrift für Heilpädagogik, 6,* 204-217.

Walter, J. (2014a). Lernfortschrittsdiagnostik Lesen (LDL) und Verlaufsdiagnostik sinnerfassenden Lesens (VSL): Zwei Verfahren als Instrumente einer formativ orientierten Lesediagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik. Tests und Trends, Jahrbuch der pädagogisch-psychologischen Diagnostik*, N.F., Bd.12, S. 165-202. Göttingen: Hogrefe. URL (Langfassung): http://193.174.11.180/HPI/DService/Tests_Trends_Walter_2014_Info.pdf (Stand: 03.08.2016).

Walter, J. (2014b). *Evaluation einer Screening-Prozedur zur Bestimmung des Risikostatus von Grundschülern bezüglich zukünftiger Lernschwächen*. Veröffentlicht in Zeitschrift für Heilpädagogik, 65 (7), 244-256, veröffentlicht in Langfassung unter URL: http://193.174.11.180/HPI/DService/Walter_2014_Evaluation_Mehrdim_Screening_ZfH_Lang.pdf (Stand: 31.07.2016)

Walter, J. & Schulz, L. (2013). Jenseits der phonologischen Informationsverarbeitung: Kann die Prognose von Lese- und Rechtschreibschwierigkeiten durch eine konzeptionelle Erweiterung und methodologische Veränderung des Vorhersagemodells verbessert werden? *Heilpädagogische Forschung, 39 (2),* 63-86.

Warnke, A. (1996). Umschriebene Lese-Rechtschreibschwäche aus kinder- und jugendpsychiatrischer Sicht. In S.-M. Behrndt & M. Steffen: *Lese-Rechtschreibschwäche im Schulalltag* (S. 21-44). Frankfurt am Main (u. a.): Lang.

Warnke, A. (1999). Reading and spelling disorders: clinical features and causes. *European Child & Adolescent Psychiatry, 8 (3)*, 2-12.

Warnke, A., Hemminger, U. & Plume, E. (2004). *Leitfaden Kinder- und Jugendpsychotherapie, Band 6. Lese-Rechtschreibstörungen.* Göttingen (u.a.): Hogrefe.

Warnke, A., Hemminger, U., Roth, E. & Schneck, S. (2002). *Legasthenie – Leitfaden für die Praxis*. Göttingen: Hogrefe-Verlag.

Warnke, A. & Roth, E. (2000). Umschriebene Lese-Rechtschreibstörung. In F. Petermann (Hrsg.). *Lehrbuch der Klinischen Kinderpsychologie und –psychotherapie* (S. 453-476). Göttingen: Hogrefe.

Wayman, M. M., Wallace, T., Wiley, H.I., Tichá, R. & Espin, C. A. (2007). Literature Sytnthesis on Curriculum-Based Measurement in Reading. The Journal of Special Education, 41 (2), 85 – 120.

Weber, E. (1972). *Grundriss der biologischen Statistik*. Jena (7. Aufl.): Fischer-Verlag.

Weber, J. (2003). *Lese- und Rechtschreibschwierigkeiten und Legasthenie. Verursachungsfaktoren und Fördermöglichkeiten*. Hamburg: Dr. Kovač.

Weber, B. (o. J.). *Lernwege. Früherkennung und –förderung schriftsprachlich relevanter Kompetenzen*. URL: http://www.legakids.net/eltern-lehrer/fruehfoerderung/lernwege/lernwege-fachtext/ (Stand: 20.10.2010).

Weber, J., Marx, P. & Lenhard, W. (2006). Prognose und Prävention von schulischen Leseproblemen. *PÄD forum: unterrichten erziehen, 35/35*, 289- 292.

Weber, J., Marx, P. & Schneider, W. (2007). Die Prävention von Lese-Rechtschreibschwierigkeiten bei Kindern mit nichtdeutscher Herkunftssprache durch ein Training der phonologischen Bewusstheit. *Zeitschrift für Pädagogische Psychologie, 21 (1)*, 65-75.

Weinert, F. E. (1977). Legasthenie: Defizitäre Erforschung defizienter Lernprozesse? *Psychologie in Erziehung und Unterricht, 24,* 164-173.

Weinert, F. E. & Schneider, W. (1987). *The Munich Longitudinal study on the Genesis of Individual Competencies (LOGIC)*. Report No. 2: Documentation of assessment procedures used in waves one to three. München: Max-Planck-Institut für psychologische Forschung.

Weinert, F. E. & Schneider, W. (1992). *The Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC)*. Report No. 8. Assessment Procedures and Results of Wave Five. München: Max-Planck Institut für Psychologische Forschung.

Weinert, F.E. & Schrader, F.-W. (1986). Diagnose des Lehrers als Diagnostiker. In H. Petillon, J. W. L. Wagner & B. Wolf. (Hrsg.): *Schülergerechte Diagnose* (S. 11-29). Weinheim: Beltz.

Weiß, R. & Osterland, J. (1997*). Grundintelligenztest Skala 1 (CFT 1).* Göttingen (5. rev. Aufl.): Hogrefe.

Wendeler, J. (1986). Prognose der Lese-Rechtschreibschwäche. *Psychologie in Erziehung und Unterricht, 33,* 10-16.

Wendeler, J. (1988). Prognose der Lese-Rechtschreib-Schwäche im Anfangsunterricht der Grundschule. *Zeitschrift für Heilpädagogik, 39 (11),* 770-781.

Wentura, D. (2005). *Multivariate Datenanalyse. Skriptum zur Veranstaltung Forschungsmethoden: Multivariate Datenanalyse* (SS 2005) Fachrichtung Psychologie, Universität des Saarlandes, Saarbrücken. URL: www.uni-saarland.de/ fak5/excops/download/SCRIPT05.pdf (Stand: 12.01.2013)

Werner, B. (o. J.). *Früherkennung und -förderung schriftsprachlich relevanter Kompetenzen*. URL: www.legakids.net/fileadmin/user_upload/Downloads/Fruehfoederung/Lernen-Fachtext.pdf (Stand: 18.03.2012).

Wesson, C. L., King, R. P., & Deno, S. L. (1984). Direct and frequent measurement of student performance: If it's good for us, why don't we do it? *Learning Disabilities Quarterly, 7*, 45-48.

WHO (2010). *ICD-10*. URL: http://apps.who.int/classifications/icd10/browse/2010/en (Stand: 20.01.2013).

WHO (2015). *ICD-10.* URL: http://www.dimdi.de/static/de/klassi/icd-10-who/kodesuche/onlinefassungen/htmlamtl2016/block-f80-f89.htm (Stand: 23.06.2017).

Wilbert, J. (2014). Instrumente zur Lernverlaufsmessung: Gütekriterien und Auswertungsherausforderungen. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.): *Lernverlaufsdiagnostik* (Tests & Trends, NF Bd. 12, S. 281-308). Göttingen: Hogrefe.

Wilbert, J. & Linnemann, M. (2011). Kriterien zur Analyse eines Tests zur Lernverlaufsdiagnostik. *Empirische Sonderpädagogik, 3*, 225-242.

Wild, K.-P. & Krapp, A. (2001). Pädagogisch-psychologische Diagnostik. In A. Krapp & B. Weidenmann (Hrsg.): *Pädagogische Psychologie* (S. 513-535). Weinheim (4. vollst. überarb. Aufl.): Beltz.

Wiley, I.H. & Deno, S.L. (2005). Oral Reading and Maze Measures as Predictors of Success for English Learners on a State Standards Assessment. *Remedial and Special Education, 26 (4),* 207–214.

Wimmer, H. (1993). Characteristics of dyslexia in a regular writing system. *Applied Psycholinguistics, 14*, 1-33.

Wimmer, H. & Goswami, U. (1994). The influence of orthographic consistency on reading development: word recognition in English and German children. *Cognition, 51,* 91-103.

Wimmer, H., Hartl, M. & Moser, E. (1990). Passen „englische" Modelle des Schriftspracherwerbs auf „deutsche" Kinder? Zweifel an der Bedeutsamkeit der logographischen Stufe. Zeitschrift für Entwicklungspsychologie und *Pädagogische Psychologie, 22 (2),* 136-154.

Wimmer, H. & Hartl, M. (1991). Erprobung einer phonologisch, multisensorischen Förderung bei jungen Schülern mit Lese-Rechtschreib-Schwierigkeiten. *Heilpädagogische Forschung, 17*, 74-79.

Wimmer, H., Landerl, K. Linortner, R. & Hummer, P. (1991). The relationship of phonemic awareness to reading acquisition: More consequence than precondition but still important. *Cognition, 40*, 219-249.

Wimmer, H., Landerl, & Schneider, W. (1994). The role of rhyme awareness in learning to read a regular orthography. *British Journal of Developmental Psychology, 12*, 469-484.

Wimmer, H. & Mayringer, H. (1999). Kognitive Defizite lese-rechtschreibschwacher Kinder. *Kindheit und Entwicklung, 8 (3),* 141-146.

Wimmer, H. & Mayringer, H. (2002). Dysfluent reading in the absence of spelling difficulties: a specific disability in regular orthographies. *Journal of Educational Psychology, 92 (2),* 272-277.

Wimmer, H., Mayringer, H. & Landerl, K.(2000). The double-deficit-hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology, 92 (4),* 668-680.

Windfuhr, K. L. & Snowling, M. J. (2001): The relationship between paired associate learning and phonological skills in normally developing readers. *Journal of Experimental Child Psychology, 80*, 160 – 173.

Wirtz, M. (2004). Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann. *Rehabilitation, 43*, 109-115.

Wolf, M. & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91 (3)*, 415-538.

Wolf, M. & Obregon, M. (1992). Early naming deficits, developmental dyslexia, and a specific deficit hypothesis. *Brain and Language, 42*, 219-247.

Wooldridge, J. M. (2003). *Introductory Economics: A modern approach*. Mason: South-Western.

World Helath Organization (Hrsg.) (2007). *International Statistical Classification of Diseases and Related Health problems.* URL: http://www.who.int/classifiations/apps/icd/icd10online/ (Stand: 05.05.2011).

Wright, J. (o. J.). *Curriculum-based measurement. A manual for teachers* [Online].URL: www.interventioncentral.org (Stand: 12.12.2007)

Wustmann, C. (2004): *Resilienz. Widerstandsfähigkeit von Kindern in Tageseinrichtungen fördern*. Weinheim: Beltz.

Wygotski, L. S. (1977). *Denken und Sprechen*. Frankfurt/M.: Fischer.

Wygotski, L. S. (1987). Unterricht und geistige Entwicklung im Schulalter. In: L. S. Wygotski (Hrsg.): *Ausgewählte Schriften, Bd. 2. Arbeiten zur psychischen Entwicklung der Persönlichkeit*. Köln: Pahl-Rugenstein Verlag.

Yeh, S. S. (2009). Class size reduction or rapid formative assessment? A comparison of cost-effectiveness. *Educational Research Review, 4*, 7-17.

Yell, M. L., Deno, S. L., & Marston, D. B. (1992). Barriers to implementing curriculum-based measurement. *Diagnostique, 18 (1),* 99-112.

Yeo, S. (2010). Predicting Performance on State Achievement Tests Using Curriculum-Based Measurement in Reading: A Multilevel Meta-Analysis. *Remedial and Special Education, 31 (6),* 412–422.

Ziegler, J., Perry, C., Coltheart, M. (2000). The DRC model in visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology, 12(3),* 413-430

Ziegler, J. & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29.

Zielinski, W. & Schneider, W. (1986). Diagnostische Möglichkeiten bei Lese- und Rechtschreibschwierigkeiten – Folgerungen aus der Forschung. In K. Ingenkamp, R. Horn & R. S. Jäger (Hrsg.). *Test und Trends 5 – Jahrbuch der Pädagogischen Diagnostik* (S. 38-62). Weinheim: Beltz.

Zimmermann, A. (1980). *Legasthenie und schriftsprachliche Kommunikation. Ansätze zur Neuorientierung eines fragwürdigen Konzepts*. Weinheim: Beltz.

XI. Anhang

Der Anhang der vorgelegten Arbeit befindet sich aufgrund des großen Seitenumfangs in einem beiliegenden zusätzlichen Band.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbstständig und ohne Hilfsmittel angefertigt habe; die aus fremden Quellen (einschließlich elektronischer Quellen, dem Internet und mündlicher Kommunikation) direkt oder indirekt übernommenen Gedanken sind ausnahmslos unter genauer Quellenangabe als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe sogenannter Promotionsberaterinnen / Promotionsberater in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar Geld oder geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Flensburg, 25.06.2017

Lea Schulz