

Interaktives Lernen an Stationen im Primarbereich

- Eine zweistufige quasiexperimentelle Evaluationsstudie der Langzeitwirksamkeit
eines naturwissenschaftlichen Bildungsprojektes

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

Vorgelegt von

Dipl.-Päd. Sören Asmussen

geboren am 22.02.1981 in Flensburg

erster Gutachter: Herr Prof. Dr. Lutz Fiesser

zweiter Gutachter: Herr Prof. Dr. Volker Müller-Benedict

eingereicht am: 12.04.2007

Für Helge

Vorwort:

„Erklärungen in Form von Vorträgen liebe ich nicht. Junge Leute geben wenig Acht darauf und behalten sie kaum. Dinge und Sachen! Ich kann es nicht genug wiederholen, dass wir den Worten zuviel Bedeutung beimessen. Mit unserer schwatzhaften Erziehung schaffen wir nur Schwätzer“ (Rousseau 1998, S.174).

Rousseaus Formulierung aus dem 17. Jahrhundert lässt sich bis heute als Anspruch an die Gestaltung von Lehr-Lernprozessen auffassen. Sie stellt eine prägnante Zusammenfassung der grundlegenden Prinzipien des genetischen Lernens nach Wagenschein (1992) dar. Seine Forderung pädagogische Prozesse konsequent vom Kind aus zu denken, deckt sich mit der Aussage des obigen Zitats aus dem Emil.

Als Erwachsene und pädagogisch Tätige neigen wir dazu diese Prinzipien im Bereich der naturwissenschaftlichen Elementarbildung nicht anzuwenden. Vielmehr deuten wir den Wunsch der Kinder nach einer Erklärung all zu oft trivial: Geben vorschnelle und vermeindlich kindgerechte Erklärungen. Sowohl die praktischen Erfahrungen, als auch die Ergebnisse der Physikdidaktik, die dem Physikunterricht kurz- und langfristig nur eine geringe Wirkung attestiert (vgl. beispielsweise Baumert, Lehmann 1997; Brämer 1980 und Dengler 1995), belegen wie wenig gewinnbringend dieses Vorgehen ist.

Vielmehr sollen die Kinder von den „Dingen und Sachen“ (vgl. oben) selbst lernen, das heißt in Bezug auf die naturwissenschaftliche Elementarbildung einen Versuch selbständig und eigenverantwortlich durchführen. Auf der Basis ihrer Beobachtungen gelangen sie, in einer ständigen Rückkopplung an das Experiment, zu altersangemessenen und subjektiv tragfähigen Erklärungen, die einen Einstieg in das wirklich tiefe Verstehen naturwissenschaftlicher Zusammenhänge bieten.

Dieses Plädoyer für die Selbsttätigkeit und Selbstständigkeit der Schülerinnen und Schüler darf nicht missverstanden werden: Das Ziel ist nicht die Abschaffung von Erklärungen. Vielmehr soll das Denken der Kinder nicht durch vorzeitig gegebene Erklärungen abgebrochen werden.

Die vorliegende Arbeit stellt den Versuch einer weiteren Fundierung der Prinzipien des genetischen Lernens dar, indem deren Wirksamkeit empirisch anhand des Beispiels der MINIPHÄNOMENTA überprüft wird. Es kann gezeigt werden, dass der vermeintlich langsame und beschwerliche Weg des genetischen Lernens, bei den Schülerinnen und Schülern zu vielfältigen Änderungen geführt hat, die als eine Vertiefung der Physikalischen Bildung gedeutet werden müssen. Es wurde daher tatsächlich, um es erneut mit den Worten Rousseaus zu sagen, Zeit verloren, um sie zu gewinnen.

Ein derartig aufwendiges Projekt wäre ohne die Unterstützung einer Vielzahl von Menschen nicht möglich gewesen. Allen voran möchte ich Herrn Prof. Fiesser danken, der die Arbeit in allen Stadien ihrer Entstehung fachlich und persönlich umfassend betreute. Mein Dank gilt weiterhin den Mitarbeitern des Institutes, insbesondere Herrn Dr. Kiupel und Herrn Stachowitz, die den Prozess der Entstehung der Arbeit ebenfalls durch vielfältige Hilfestellungen unterstützten.

Das Anfertigen einer Dissertation stellt auch für das private Umfeld eine erhebliche Herausforderung dar. Ich danke Jania, Helge, Elisabeth und Johanna ausdrücklich für die Diskussionen, die Ratschläge und die Toleranz, die ich ihnen besonders an Wochenenden und zu nächtlicher Zeit abverlangen durfte.

Zuletzt möchte ich den beteiligten Schulen, das heißt den Schülerinnen und Schülern, den Lehrerinnen und Lehrern und den Eltern danken, ohne deren produktive Mitarbeit das empirische Projekt zum Scheitern verurteilt wäre.

Flensburg im April 2007

Sören Asmussen

Zusammenfassung der Ergebnisse

In the following study the long-term effectiveness of a scientific project for primary school education called MINIPHÄNOMENTA is researched thoroughly. To do so a two-stage experimental design has been used.

The results of the tests show serious deficiencies on part of the pupils regarding to scientific education prior to participation in the MINIPHÄNOMENTA project. The objectives regarding to dealing with science and especially with physical and chemical experiments in school as they are defined in the curriculum are not met by a vast majority of pupils. The results of the prätest are disilluioning. Taking this into account the results of the posttests and the step two of the evaluation must be deemed to be highly encouraging.

It can be shown that the MINIPHÄNOMENTA project has far-reaching influence on the physical education of the pupils moreover it can be shown that the exerted influence remains valid over a long period of time.

In der folgenden Arbeit sollen die Langzeitwirkungen der MINIPHÄNOMENTA¹, einem naturwissenschaftlichen Bildungsprogramm für die Primarstufe, untersucht werden. Dies geschieht anhand von zwei Untersuchungsstufen, die beide - dies ergibt sich aus der Definition des Wirksamkeitsbegriffes - zwingend experimentell zu konstruieren sind. Die Untersuchungsstufe I stellt das Zentrum der Auswertung dar. Es handelt sich hierbei um einen Zwei-Gruppen-Prätest-Posttest-Plan mit zwei Messwiederholungen. Die Wirkungen der MINIPHÄNOMENTA auf die abhängigen Variablen können so über einen Zeitraum von sechs Monaten umfassend beschrieben werden. Bei der Untersuchungsstufe II handelt es sich um einen Ein-Gruppen-Prätest-Posttest-Plan. Durch ihn können die Wirkungen der MINIPHÄNOMENTA explorativ über einen Zeitraum von zweieinviertel Jahren erfasst werden.

Ausgehend von den durch den Physikunterricht zu fördernden Kompetenzbereichen (Sach-, Selbst-, Methoden- und Sozialkompetenz) werden die folgenden abhängigen Variablen der Untersuchung eingeführt:

¹ Bei der MINIPHÄNOMENTA handelt es sich um ein komplexes pädagogisches Gesamtangebot mit dem Ziel die naturwissenschaftliche Bildung in der Primarstufe zu fördern. Das Zentrum des Projekts stellen 53 Experimentierstationen dar, an denen die Schülerinnen und Schüler, ohne externe Vorgaben, interessengeleitet im Rahmen der Schulpausen arbeiten können.

Variable	Erfassungsbereich	Erhebungsmethode
deskriptive Aspekte	Erinnerung der Schülerinnen und Schüler an die Stationen der MINIPHÄNOMENTA	Fragebogenmethode
Einstellung gegenüber Phänomen und ihrer Bearbeitung (kurz: Einstellung)	Einstellung gegenüber Phänomenen auf der Basis des Drei-Komponenten-Modells. Eine Einstellung besteht aus drei Dimensionen: <ul style="list-style-type: none"> • Kognitive Dimension • Affektive Dimension • verhaltensbezogene Dimension 	Fragebogenmethode
Formale Kompetenz	Hier werden drei Aspekte erfasst: <ul style="list-style-type: none"> • Logische Kompetenzen • Experimentelle Kompetenzen • Spezifisch naturwissenschaftliche Denkweisen 	Verhaltensbeobachtung und Interview
Allgemeines Experimentierverhalten	Verhalten in einer offenen Experimentiersituation: <ul style="list-style-type: none"> • Vertiefung in den Versuch • Fähigkeit zur Gruppenarbeit • Formale Aspekte des Experimentierens 	Verhaltensbeobachtung und Interview
Aktuelle Motivation zur tiefergehenden Bearbeitung physikalischer Fragestellungen (kurz: Motivation)	Motivation zur tiefergehenden Beschäftigung mit einem Phänomen. Sie wird durch drei Dimensionen operationalisiert: <ul style="list-style-type: none"> • Interesse • Misserfolgsbefürchtungen • Kompetenzerwartungen 	Fragebogenmethode
Wissen	Fähigkeit eine subjektiv tragfähige Erklärung auf der Basis von Beobachtungen zu entwickeln	Verfahren in Anlehnung an die Methode der Concept-Maps

Tabelle 1: Zusammenfassung der Ergebnisse;
- Übersicht über die erfassten abhängigen Variablen
Der Verfasser

Die MINIPHÄNOMENTA wird als wirksam betrachtet, sofern sich die Schülerinnen und Schüler an die Stationen erinnern, ihre [Einstellung gegenüber dem Experimentieren]², ihre [formalen und experimentellen Kompetenzen], ihre [Motivation] naturwissenschaftliche Probleme zu lösen und ihr [Wissen] verbessern konnten. Dabei wird Wirksamkeit als polytomes Merkmal mit mehreren Stufen gefasst: hoch wirksam, wirksam, teilweise wirksam und unwirksam. Die Ergebnisse der abhängigen Variablen werden anhand der zum statistischen Prüfverfahren gehörenden Effektgröße und des Umfangs der Wirkung den genannten Wirksamkeitsstufen zugewiesen.

Forschungsmethodisch wird darüber hinaus zwischen zwei Messebenen der Wirksamkeit unterschieden:

- 1 Ebene I: Wirksamkeit als positive Entwicklung der abhängigen Variablen in den Posttests in Relation zu den Ergebnissen im Prätest.
- 2 Ebene II: Die Prä-Posttest-Differenz (siehe Ebene I) wird hier zusätzlich am Verlauf der abhängigen Variablen in der Kontrollgruppe relativiert.

² Zur besseren Übersicht sind die Namen der Variablen im Textverlauf durch eckige Klammern hervorgehoben.

Es muss hier betont werden, dass es sich bei den Wirksamkeitsebenen nicht um ein Merkmal handelt, dass nach dem Grad der Wirksamkeit differenziert. Vielmehr beschreiben die beiden Ebenen die Wirksamkeit der MINIPHÄNOMENTA in methodisch unterschiedlich kontrollierter Weise.

Im Vorfeld der MINIPHÄNOMENTA müssen die Ausprägungen auf den genannten abhängigen Variablen als überaus niedrig beurteilt werden: Die Schülerinnen und Schüler waren experimentell unerfahren: Die eigenständige Durchführung eines Versuches, das Beobachten der Ergebnisse sowie das Nachdenken über mögliche Erklärungen fiel ihnen schwer. Weiterhin sanken die [Motivation naturwissenschaftliche Fragestellungen zu bearbeiten] und die [Einstellung gegenüber dem Experimentieren] mit steigendem Alter. Ebenfalls bedenklich sind die deutlichen geschlechtsspezifischen Unterschiede, insbesondere bei den Variablen [Einstellung] und [Motivation]. Damit werden die Anforderungen des Lehrplanes HSU nicht eingelöst. Keines der von diesem Lehrplan definierten Lernziele wurde von den Schülerinnen und Schülern erreicht.

In beiden Posttests der Untersuchungsstufe I und in der Untersuchungsstufe II konnte der große Einfluss der MINIPHÄNOMENTA auf die abhängigen Variablen nachgewiesen werden. Die folgende Tabelle gibt die Wirksamkeit des treatments (MINIPHÄNOMENTA) im Bezug auf die erfassten Variablen wieder:

Variable	Untersuchungsstufe I	Untersuchungsstufe I	Untersuchungsstufe II
	Posttest I	Posttest II	
	Ebene I	Ebene II	Ebene I
Deskriptive Aspekte	Die Wirksamkeit der MINIPHÄNOMENTA kann immer nur in Relation zu den Ergebnissen des Prätests ermittelt werden. Da hier aus forschungspraktischen Gründen die Erhebung unterbleiben muss, wird für die Variable keine Wirksamkeitsstufe zugeschrieben. Vielmehr wird die Erinnerung an die Stationen in dieser Arbeit als Bedingung der Wirksamkeit der MINIPHÄNOMENTA gefasst.		
	Es konnte festgestellt werden, dass die Schülerinnen und Schüler sich zu den unterschiedlichen Zeitpunkten umfassend an die Stationen des Projektes erinnern.		
Einstellung	HOCHWIRKSAM	WIRKSAM	WIRKSAM
Formale Kompetenz	HOCHWIRKSAM	HOCHWIRKSAM	HOCHWIRKSAM
Allgemeines Experimentierverhalten	HOCHWIRKSAM	HOCHWIRKSAM	HOCHWIRKSAM
Motivation	HOCHWIRKSAM	TEILWEISE WIRKSAM	Siehe Erläuterungen im Textverlauf
Wissen	HOCHWIRKSAM	HOCHWIRKSAM	HOCHWIRKSAM
Gesamt	HOCHWIRKSAM	WIRKSAM	WIRKSAM

Tabelle 2: Zusammenfassung der Ergebnisse;
- Die Wirksamkeit der MINIPHÄNOMENTA auf die erfassten abhängigen Variablen
Der Verfasser

Grundlage der Wirksamkeit von ASIP ist die [Erinnerung der Schülerinnen und Schüler an die Stationen]: Diese ist sowohl im Bezug auf die Quantität als auch auf die Qualität zu jedem Erhebungszeitpunkt gegeben. Es gelingt den Kindern eine Vielzahl von Stationen wieder zu erkennen und im Bezug auf ihren Aufbau, die beobachtbaren Effekte sowie die experimentellen Abläufe korrekt zu beschreiben.

Ein besonders herausragender Effekt der MINIPHÄNOMENTA besteht in der umfassenden Förderung der [formalen Kompetenzen] der Schülerinnen und Schüler. Sie arbeiteten nach der Teilnahme an der MINIPHÄNOMENTA länger und konzentrierter an einem Experiment und gingen dabei insgesamt deutlich erfahrener vor als im Vorfeld des Projektes. Darüber hinaus wurden sie durch die Teilnahme an dem Projekt, dies zeigen die Ergebnisse der Variablen [Wissen], in die Lage versetzt auf der Basis eigener Beobachtungen zu subjektiv tragfähigen Erklärungen zu gelangen. Auch die [Einstellung der Schülerinnen und Schüler gegenüber Phänomenen und ihrer Bearbeitung] konnte durch die MINIPHÄNOMENTA deutlich positiver gestaltet werden. Insbesondere der ‚verhaltensbezogene Anteil‘³ der Einstellung wurde entscheidend gefördert. Die Kinder erleben damit das Experimentieren als einen aktiven Prozess, den sie gestalten können und wollen.

Entsprechend der Kriterien muss der Variablen [Motivation] in der Untersuchungsstufe II die Wirksamkeitsstufe UNWIRKSAM zugewiesen werden. Dieses Urteil muss durch drei Ergebnisse relativiert werden:

1. Die [Motivation] der Schülerinnen und Schüler wurde aufgrund methodischer Probleme lediglich in der Klassenstufe vier erhoben. Es liegen somit keine Daten für die gesamte Stichprobe vor.
2. Zudem kann festgestellt werden, dass die [Motivation] die einzige Variable ist, in der im Prätest keine unterdurchschnittlichen Werte erzielt wurden. Damit besteht bei den Schülerinnen und Schülern auch schon vor der MINIPHÄNOMENTA die Bereitschaft über physikalische Fragestellungen nachzudenken.
3. Darüber hinaus ist, dies legen die Daten der Untersuchungsstufe I nahe, der geringe Effekt der MINIPHÄNOMENTA in Bezug auf die [Motivation] der Schülerinnen und Schüler vor allem dadurch bedingt, dass ihre ‚Kompetenzerwartungen‘ infolge der Teilnahme an der MINIPHÄNOMENTA zurückgingen. Diese Tatsache könnte dadurch hervorgerufen werden, dass bei ASIP keinen externen Bewertungskriterien existieren, anhand derer die Kinder ihre eigene Kompetenz exakt einschätzen können.

Insgesamt kann damit die Langzeitwirksamkeit der MINIPHÄNOMENTA festgestellt werden.

³ Die Dimensionen der abhängigen Variablen sind im Textverlauf zur besseren Übersicht in einfache Anführungen gesetzt.

Gliederung

Vorwort

Zusammenfassung der Ergebnisse

1 Theoretischer Teil

1.1 Einführung in die Untersuchung S. 15 - 17

1.2 Die Ausgangssituation:

Physikalische Bildung in Deutschland

1.2.1 Zur Diskussion über den Begriff
der Physikalischen Bildung

1.2.1.1 Der pragmatische Bildungsbegriff
der PISA-Studie S. 18 - 20

1.2.1.2 Ein erweiterter Bildungsbegriff:
Bildung als „Stärkung des Ich“ S. 21 - 26

1.2.2 Die Physikalische Bildung im Spiegel
empirischer Studien

1.2.2.1 Einleitung S. 27 - 28

1.2.2.2 Das Interesse S. 29 - 33

1.2.2.3 Das Wissen S. 34 - 37

1.2.3 Physikalische Bildung in der Primarstufe S. 38 - 39

1.3 Die MINIPHÄNOMENTA als ein Beitrag zur Physikalischen Bildung

1.3.1	Das Projekt	S. 40 - 43
1.3.2	Der theoretische Begründungszusammenhang der MINIPHÄNOMENTA	
1.3.2.1	Einleitung	S.44 - 45
1.3.2.2	Die Theorie Martin Wagenscheins als pädagogische Grundlage des interaktiven Lernens	S. 46 - 49
1.3.2.3	Die Theorie der kognitiven Entwicklung von Piaget unter der besonderen Berücksichtigung der Kritik und Weiterentwicklungen	S. 50 - 54
1.3.2.4	Interaktive Stationen vor dem Hintergrund aktueller Theorien des selbstgesteuerten Lernens	S. 55 - 57
1.3.3	Ergebnisse der bisherigen Forschung zum Projekt	S. 58 - 60

1.4 Die Grundlagen der abhängigen Variablen der Untersuchung: Zentrale Definitionen und Theorien

1.4.1	Die abhängigen Variablen im Untersuchungskontext	S. 61 - 64
1.4.2	Deskriptive Aspekte	S. 65 - 66
1.4.3	Einstellung gegenüber Phänomenen und ihrer Bearbeitung	S. 67 - 68
1.4.4	Formale Kompetenz	S. 69 - 70
1.4.5	Allgemeines Experimentierverhalten	S. 71 - 73
1.4.6	Aktuelle Motivation zur tiefgehenden Bearbeitung physikalischer Fragestellungen	S. 74 - 77
1.4.7	Wissen	S. 78 - 79

2 Methodische Aspekte

- 2.1 Der Begriff der Wirksamkeit und seine methodischen Implikationen S. 80 - 83
- 2.2 Das Untersuchungsdesign S. 84 - 86
- 2.3 Die Stichprobe
 - 2.3.1 Stichprobenziehung in der Untersuchungsstufe I S. 87 - 89
 - 2.3.2 Kennwerte der Experimentalgruppe I und der Kontrollgruppe S. 90 - 91
 - 2.3.3 Stichprobenziehung in der Untersuchungsstufe II S. 92 - 92
 - 2.3.4 Die Experimentalgruppe II S. 93 - 94
 - 2.3.5 Repräsentativitätsprobleme S. 95 - 96
- 2.4 Die Methoden und Untersuchungsinstrumente
 - 2.4.1 Einleitung S. 97 - 99
 - 2.4.2 Fragebogen zur Erfassung der Einstellung gegenüber Phänomenen und ihrer Bearbeitung S. 100 - 104
 - 2.4.3 Fragebogen zur Erfassung deskriptiver Aspekte S. 105 - 107
 - 2.4.4 Kategorienschema zur Ermittlung der formalen Kompetenz S. 108 - 110
 - 2.4.5 Kategorienschema zur Ermittlung des allgemeinen Experimentierverhaltens S. 111 - 114
 - 2.4.6 Der modifizierte Fragebogen zur Erfassung der aktuellen Motivation S. 115 - 120
 - 2.4.7 Wissensdiagnostik S. 121 - 125
- 2.5 Die Stufen der Wirksamkeit S. 126 - 131
- 2.6 Konkretisierung der Fragestellung: Die Hypothesen der Untersuchung S. 132 - 132

3 Darstellung der Ergebnisse

3.1	Einleitung in die Auswertung	S. 133 - 134
3.2	Ergebnisse der Untersuchungsstufe I	
3.2.1	Der Prätest	
3.2.1.1	Die Vergleichbarkeit von Experimental- und Kontrollgruppe im Vorfeld des treatments	S. 135 - 139
3.2.1.2	Der Stand vor ASIP: Ergebnisse der Prätestuntersuchung	S. 140 - 169
3.2.1.3	Zusammenfassung der Ergebnisse	S. 170 - 172
3.2.2	Der Posttest I	
3.2.2.1	Die Wirksamkeit von ASIP drei Monate nach der Teilnahme	S. 173 - 201
3.2.2.2	Zusammenfassung der Ergebnisse	S. 202 - 203
3.2.3	Der Posttest II	
3.2.3.1	Die Wirksamkeit von ASIP sechs Monate nach der Teilnahme	S. 204 - 225
3.2.3.2	Zusammenfassung der Ergebnisse des Posttest II	S. 226 - 227
3.2.4	Zusammenfassende Bewertung der Ergebnisse der Untersuchungsstufe I	S. 228 - 229

3.3	Ergebnisse der Untersuchungsstufe II	
3.3.1	Die Wirksamkeit von ASIP zweieinviertel Jahre nach der Teilnahme	S. 230 - 241
3.3.2	Zusammenfassung der Ergebnisse: Auswertung im Hinblick auf die Hypothesen	S. 242 - 243
3.4	Erweiterung des Wirksamkeitsbegriffes	S. 244 - 246
3.5	Ausblick	S. 247 - 248

4 Anhang

- 4.1 Fragebogen zur Erfassung der deskriptiven Daten
- 4.2 Fragebogen zur Erfassung der Einstellung
- 4.3 Fragebogen zur Erfassung der Einstellung (Endform)
- 4.4 Fotos und Beschreibungen der Versuche zur Ermittlung der formalen Kompetenz
- 4.5 Kategorienschema formale Kompetenz
- 4.6 Kategorienschema formale Kompetenz (Endform)
- 4.7 Kodieranweisung zum Kategorienschema formale Kompetenzen (für die Endform)
- 4.8 Beobachtungsbogen allgemeines Experimentierverhalten
- 4.9 Beobachtungsbogen allgemeines Experimentierverhalten (Endform)
- 4.10 Kodieranweisung zum Beobachtungsbogen allgemeines Experimentierverhalten (für die Endform)
- 4.11 Liste der zur Ermittlung des allgemeinen Experimentierverhalten verwendeten Versuche
- 4.12 Überarbeitete Version des FAM
- 4.13 Überarbeitete Version des FAM (Endform)

Literaturverzeichnis

Curriculum vitae

Erklärung

1 Theoretischer Teil

1.1 Einführung in die Untersuchung

Das Ziel der vorliegenden empirischen Studie besteht in der Analyse der Langzeitwirkung eines naturwissenschaftlichen Bildungsprojektes für die Primarstufe: der MINIPHÄNOMENTA. Hierbei handelt es sich um ein physikalisches Erfahrungsfeld aus derzeit 53 Versuchsstationen⁴. Zum besseren Verständnis des folgenden Theorieteils nenne ich zunächst einige zentrale Merkmale des Projekts:

Die zentralen Elemente der MINIPHÄNOMENTA bilden die genannten Versuchsstationen, die jeweils ein naturwissenschaftliches, zumeist physikalisches Phänomen in spannender und interessanter Weise darstellen. Sie werden zunächst für 14 Tage leihweise in den Fluren der teilnehmenden Schulen aufgestellt. Dabei sind die Stationen in ein pädagogisches Gesamtkonzept eingebunden, das mit der Fortbildung von Lehrerinnen und Lehrern im Bereich der naturwissenschaftlichen Elementarbildung beginnt. In einer Wochenendveranstaltung lernen sie das Projekt, die Stationen und den vertretenen pädagogischen Ansatz praxisnah und handlungsorientiert kennen. Hiernach sollen sie als Multiplikatoren an der durch sie vertretenen Schule wirken. Um die Nachhaltigkeit der MINIPHÄNOMENTA auch über den Zeitraum der Ausleihe der Stationen hinaus sicher stellen zu können, sollen die Exponate von den Eltern und den Lehrkräften der Schule nachgebaut werden. So finden sie dauerhaften Einzug in den unterrichtlichen und schulischen Alltag.

Die Versuche sind möglichst anschaulich und schlicht gehalten. Dieser Aufbau erleichtert den Schülerinnen und Schülern den Umgang mit den Stationen und ermöglicht ihnen die Konzentration auf das Wesentliche, das naturwissenschaftliche Phänomen und den Umgang mit ihm. An den Stationen befinden sich keinerlei Erklärungen oder Experimentieranweisungen, so dass die Schülerinnen und Schüler den Umgang mit den Versuchen völlig selbstbestimmt gestalten können. Dieser Charakter der Versuchsstationen wird auch dadurch unterstrichen, dass die MINIPHÄNOMENTA weniger ein Unterrichtsmedium darstellt, als vielmehr eine alternative Pausenbeschäftigung für die Schülerinnen und Schüler.

⁴ Die Begriffe Versuch, Station, Experiment und Exponat sowie Kombinationen aus ihnen werden in dieser Arbeit bedeutungsgleich verwendet. Dabei muss zwischen zwei Arten von Versuchen unterschieden werden:

1. Stationen des Projektes MINIPHÄNOMENTA: Die MINIPHÄNOMENTA besteht derzeit, dies wurde bereits im obigen Textverlauf angedeutet aus 53 Versuchsstationen (vgl. hierzu insbesondere den Abschnitt 1.3.1 in dieser Arbeit).
2. Versuche die zur Erhebung der Langzeitwirksamkeit der MINIPHÄNOMENTA im Rahmen dieser Arbeit herangezogen werden: Zur Erhebung der Langzeitwirksamkeit der MINIPHÄNOMENTA auf den unterschiedlichen Variablen ist eine Vielzahl von weiteren Versuchen erforderlich. An ihnen werden die untersuchten Schülerinnen und Schüler in unterschiedlichen Settings zumeist beobachtet oder interviewt.

Welche Art von Station gemeint ist wird im Textzusammenhang deutlich.

Bisherige Forschungsergebnisse bestätigen den großen Erfolg von ASIP⁵: Die Schülerinnen und Schüler arbeiteten konzentriert und über lange Zeiträume an den Stationen. Sie experimentierten zumeist gemeinsam und diskutierten Lösungen und mögliche Erklärungen in der Gruppe. Effekte der Stationen auf die Begriffsbildung der Schülerinnen und Schüler im naturwissenschaftlichen Bereich konnten mit der Hilfe von Concept-Maps nachgewiesen werden. Weiterhin gelang es durch das Projekt, die Eltern aktiv in schulische Prozesse einzubeziehen (vgl. Sauer 2005, Holst 2005, John 2004).

Bei allen bisherigen Untersuchungen zur MINIPHÄNOMENTA handelt es sich um querschnittliche Erhebungen. Um den Erfolg des Projektes langfristig und umfassend einschätzen zu können, muss die Wirksamkeit von ASIP längsschnittlich untersucht werden. Wie ist es also um die Langzeitwirksamkeit des Projektes bestellt? Lassen sich nach unterschiedlich langen Zeiträumen, im Vergleich mit einer Kontrollgruppe, sicher auf ASIP zurückführbare Effekte nachweisen? Dies ist die zentrale Frage der vorliegenden Untersuchung.

Mit dieser Arbeit wird die Untersuchung der Langzeitfolgen der MINIPHÄNOMENTA intendiert.

Hierzu wird ein zweistufiges quasiexperimentelles Evaluationsdesign verwendet mit dem der kausale Einfluss des treatments der MINIPHÄNOMENTA auf die abhängigen Variablen zu unterschiedlichen Zeitpunkten überprüft wird.

Im **ersten** Teil der Arbeit werde ich die theoretischen Grundlagen verdeutlichen. Ausgehend vom Begriffen der Physikalischen Bildung und empirischen Ergebnissen wird die Notwendigkeit physikalischer Bildungsprojekte hergeleitet. Darauf aufbauend wird das Projekt MINIPHÄNOMENTA (Konzeption, theoretischer Hintergrund und bereits vorliegende Forschungsergebnisse) ausführlich beschrieben. Es folgt die Herleitung und Definition der abhängigen Variablen. Abschließend werden die Hypothesen der empirischen Untersuchung konkretisiert.

Der **zweite** Teil befasst sich mit der Methodik der Arbeit. Basierend auf dem Begriff der Wirksamkeit wird in diesem Abschnitt auf drei Themengebiete eingegangen: die Stichprobenziehung und Zusammensetzung, die Untersuchungsinstrumente sowie auf das grundsätzliche Vorgehen bei der Erhebung und Auswertung der Daten.

⁵ Die Begriffe ASIP (**A**dvancing **S**cience through **I**nformed **P**arents) und MINIPHÄNOMENTA bezeichnen im Rahmen dieser Arbeit das gleiche Projekt. Die erste Bezeichnung unterstreicht dabei die aktive Einbindung der Eltern in das Projekt. Im Rahmen einer Informationsveranstaltung werden sie aufgefordert für die Schule ihrer Kinder ein kleines Experimentierfeld mit Versuchen aus der MINIPHÄNOMENTA nachzubauen.

Im **dritten** Teil der Arbeit werden die Ergebnisse der beiden Untersuchungsstufen ausgewertet. Dabei wird zunächst die prinzipielle Wirksamkeit des treatments in Bezug auf die abhängige Variable untersucht. Darauf aufbauend wird spezielleren Fragestellungen, wie zum Beispiel dem Einfluss differentialpsychologischer Aspekte nachgegangen.

An dieser Stelle noch einige Anmerkungen zur formalen Darstellungsweise in der Arbeit:

- Zusammenfassungen im Textverlauf sind durch graue Kästen mit dicken Linien hervorgehoben.
- Zentrale Definitionen sind schwarz mit dünnen Linien umrandet.
- Die erhobenen abhängigen Variablen sind im Textverlauf durch eckige Klammern hervorgehoben.
- Die Dimensionen der abhängigen Variablen sind durch einfache Anführungszeichen gekennzeichnet.

1.2 Die Ausgangssituation: Physikalischen Bildung in Deutschland

1.2.1 Zur Diskussion über den Begriff der Physikalischen Bildung

1.2.1.1 Der pragmatische Bildungsbegriff der PISA-Studie

Der Begriff der Physikalischen Bildung ist komplex und heterogen. In der Geschichte der physikdidaktischen Forschung und der empirischen Bildungsforschung hat es eine Vielzahl höchst unterschiedlicher Begriffsbestimmungsversuche gegeben (vgl. John 1995, S. 25 ff.). An dieser Stelle beginne ich, aufgrund ihrer aktuellen politischen und wissenschaftlichen Relevanz, mit den Definitionen der PISA-Studie. Darauf aufbauend werde ich zeigen, dass die in der PISA-Studie angewandte Konzeption zentrale Defizite für den Kontext dieser Arbeit aufweist. Der Begriff der Physikalischen Bildung wird durch sie zu eng definiert. Eine Vielzahl wichtiger Prozesse, vor allem in Bezug auf das hier zentrale interaktive Lernen (vgl. Abschnitt 1.4.3), lassen sich durch den dort verwandten Begriff der Physikalischen Bildung nicht erfassen. Aus diesem Grund wird in dem folgenden Kapitel ein umfassenderer Begriff der Physikalischen Bildung entwickelt, der Bezug nimmt auf das traditionelle Verständnis von Bildung des deutschen Idealismus (vgl. Reble 2002, S. 174 ff.). Abschließend werden einige ausgewählte empirische Ergebnisse zu dem eingeführten Begriff der Physikalischen Bildung geschildert.

In Bezug auf die Definitionen der PISA-Studie muss festgestellt werden, dass der Begriff der Physikalischen Bildung an keiner Stelle explizit definiert wird, lediglich der Begriff der naturwissenschaftlichen Bildung, einige Autoren sprechen von Grundbildung, wird eingeführt. Diese Definition geht über den Begriff der Physikalischen Bildung hinaus und schließt die Chemie, Biologie und Geowissenschaften mit ein (vgl. Deutsches PISA-Konsortium 2005, S. 103 ff.)

Grundlage des Verständnisses von naturwissenschaftlicher Bildung ist das „literacy-Konzept“: Aus dem Bereich der Lese- und Schreibfähigkeit kommend, betont dieses Konzeption die Funktion von Bildung. Bildung in unterschiedlichen Bereichen, hier dem Bereich der Naturwissenschaften, soll den Menschen in die Lage versetzen an der Kultur teilzuhaben. Sie stellen eine normative Zielvorgabe dar, die - basierend auf demokratisch-partizipativen Gedanken - von jeder oder jedem erreicht werden soll, damit sie oder er in die Lage versetzt wird, sich in einer immer mehr von Technik und Naturwissenschaften geprägten Welt zu orientieren und Entscheidungen zu verstehen, beziehungsweise kompetent zu kritisieren. (vgl. Rost, Prenzel, Senkbeil, Groß 2004, S. 25 ff.). Die Definition von Bildung über ihre Funktion erscheint, betrachtet man die 200 Jahren andauernde deutsche Diskussion um den Begriff der Bildung (vgl. Benner 2001, S. 63 ff.), wenig reflektiert. Dies räumen auch die Autoren der Studie ein. Sie betonen deutlich den pragmatischen Charakter der Definition, der zum einen für eine hohe internationale Vergleichbarkeit Sorge, zum anderen zur empirischen Testkonzeption gut geeignet sei.

Basierend auf dem genannten Konzept wird naturwissenschaftliche Bildung als „die Fähigkeit, naturwissenschaftliches Wissen anzuwenden, naturwissenschaftliche Fragen zu erkennen und aus Belegen Schlussfolgerungen zu ziehen, um Entscheidungen zu verstehen und zu treffen, welche die natürliche Welt und die durch menschliches Handeln an ihr vorgenommenen Veränderungen betreffen“ (ebd. 2004, S. 28) verstanden.

In Bezug auf die naturwissenschaftliche Bildung können, nach der obigen Definition, vier Dimensionen unterschieden werden:

1. Kenntnisse über naturwissenschaftliche Begriffe
2. Vorstellungen über naturwissenschaftliche Denk- und Vorgehensweisen
3. Kenntnisse über die Besonderheiten der Naturwissenschaften
4. Wissen über die Beziehungen zwischen Naturwissenschaften und Gesellschaft

Von diesem Begriff der naturwissenschaftlichen Bildung ausgehend werden zunächst fünf Kompetenzstufen definiert. Dabei wird die bloße Wiedergabe von einfachen naturwissenschaftlichen Begriffen als niedrigste Kompetenzstufe beurteilt. Die höchste Kompetenz besteht in der Entwicklung und Anwendung von Modellvorstellungen auf deren Basis Hypothesen gebildet und überprüft werden. (vgl. Rost, Prenzel, Carstensen, Senkbeil, Groß 2004, S. 42)

Weiterhin werden kognitive Teilkompetenzen⁶ definiert, die nicht stufenförmig angelegt sind, sondern vielmehr einzelne Kompetenzbereiche beschreiben, die bei jeder Schülerin oder bei jedem Schüler individuell unterschiedlich ausgeprägt sind. Mit ihrer Hilfe soll beschrieben werden, aus welchen Teilkompetenzen sich naturwissenschaftliche Bildung zusammensetzt und in welchen Bereichen deutsche Schülerinnen und Schüler besondere Probleme haben, beziehungsweise in Bezug auf welche Bereiche sie besonders kompetent sind (vgl. Rost, Prenzel, Carstensen, Senkbeil, Groß 2004, S. 41). Jede der von den Schülerinnen und Schülern zu lösende Aufgabe, überprüft dabei die Fähigkeiten in einem oder mehreren kognitiven Kompetenzbereichen. Sie lauten im Einzelnen:

⁶ PISA 2000 ist so konzipiert, dass zusätzlich zu den internationalen Vergleichsuntersuchungen, nationale Ergänzungen eingebracht werden können. Die Erhebung der kognitiven Teilkompetenzen stellt hier die zentrale deutsche nationale Erweiterung dar.

Nummer	Kompetenzbereich	Definition
1	Faktenwissen	Heranziehen von deklarativem Wissen zur Aufgabenlösung
2	Grafikverständnis	Das Entnehmen relevanter Informationen aus einer Grafik.
3	Mentale Modelle	Nutzung eines mentalen Modells zur Bearbeitung einer Aufgabe.
4	Schlüsse ziehen	Auf der Basis gegebener Informationen angemessene Schlüsse ziehen.
5	Sachverhalte verbalisieren	Einen Sachverhalt sprachlich angemessen beschreiben (vgl. Rost, Prenzel, Carstensen, Senkbeil, Groß 2004, S. 17).

Tabelle 1.2.1.1.1: Der pragmatische Bildungsbegriff der PISA-Studie;
- kognitive Kompetenzbereiche der nationalen Erweiterung der Pisa-Studie
Der Verfasser

1.2.1.2 Ein erweiterter Bildungsbegriff: Bildung als „Stärkung des Ich“

In diesem Abschnitt soll der Begriff der Physikalischen Bildung für den Kontext des interaktiven Lernens (siehe unten in diesem Abschnitt) erweitert werden. Ausgehend von der pragmatischen Konzeption des Begriffes der naturwissenschaftlichen Bildung der PISA-Studie ist eine Vielzahl von Erweiterungen und Ergänzungen erforderlich. Zur besseren Übersicht über den folgenden Textverlauf möchte ich die Argumentationsfigur zunächst kurz umreißen:

Ich beginne mit der allgemeinen Kritik am Begriff der naturwissenschaftlichen Bildung der PISA-Studie, wie sie in der einschlägigen Literatur genannt wird. Darauf aufbauend definiere ich den Begriff des interaktiven Lernens und zeige anhand eines Gedankenexperimentes weitere Defizite des genannten Begriffes der naturwissenschaftlichen Bildung der PISA-Studie in Bezug auf die hier zu diskutierende Untersuchung. Es folgen Bezüge zum Begriff der Bildung des deutschen Idealismus. Hier werde ich zeigen, dass die Physikalische Bildung ein Teil der allgemeinen Bildung im Sinne Humboldts ist, die zentral der Stärkung des Individuums dient. Abschließend wird der entwickelte erweiterte Begriff der Physikalischen Bildung mit dem Begriff der naturwissenschaftlichen Bildung der PISA-Studie nochmals verglichen.

Es muss zunächst betont werden, dass der Begriff der naturwissenschaftlichen Bildung der PISA-Studie deutlich weiter gefasst ist, folglich einen größeren Gegenstandsbereich umfasst, als der Begriff der Physikalischen Bildung⁷. Da der größte Teil der Stationen der MINIPHÄNOMENTA dem Bereich der Physik zuzurechnen ist, erscheint der genannte Begriff der Bildung im Hinblick auf den Untersuchungsgegenstand wenig passend.

Ein weiterer Kritikpunkt betrifft das Konzept der gesellschaftlichen Teilhabe. Dengler (1995, nach dem Deutschen PISA-Konsortium 2001, B, S. 196 ff.) kritisiert zu Recht, dass die hierfür erforderlichen Kompetenzen sehr umfangreich sein müssten. Es kann von Laien nicht erwartet werden, dass sie den Stand jüngerer naturwissenschaftlicher Forschungen verstehen.

⁷ Die Unterscheidung der Physik von anderen Naturwissenschaften kann auf zwei Ebenen vollzogen werden:

1. Zunächst auf der Stufe der Alltagstheorien: Im weiteren Verlauf des Abschnittes 1.2 in dieser Arbeit werde ich zeigen, dass sich die Physik wesentlich durch ihre starke Verbindung zur Mathematik wahrgenommen wird. Dadurch ist sie bei den befragten Untersuchungsteilnehmerinnen und Untersuchungsteilnehmerinnen die unbeliebteste Naturwissenschaft.
2. Weiterhin auf der Ebene der Fachwissenschaft: Physik als wissenschaftliche Disziplin lässt sich definieren als „die Lehre von solchen Naturvorgängen, die experimenteller Erforschung, messender Erfassung und mathematischer Darstellung zugänglich sind und allgemeinen Gesetzen unterliegen, insbesondere von Erscheinungs- und Zustandsformen der Materie, ihrer Struktur und ihren Eigenschaften“ (vgl. Fiesser 1990, S. 7). Sie unterscheidet sich wesentlich von anderen naturwissenschaftlichen Disziplinen durch ihre Orientierung am Experiment. Ihr Ziel besteht damit nicht nur in der Beschreibung von Sachverhalten, sondern vorrangig in ihrer Erklärung, durch allgemeine, oft mathematisch formulierte Gesetzmäßigkeiten.

Miller (1997) betont darüber hinaus, dass „die Notwendigkeit der postulierten Kompetenzen für die Bewältigung des Lebens keineswegs evident sei“ (zitiert nach Deutsches PISA-Konsortium 2001, B, S 196).

Neben dieser allgemeinen Kritik an der Definition der naturwissenschaftlichen Bildung in der PISA-Studie wurde im Abschnitt 1.2.1.1 festgestellt, dass dieser Begriff der Bildung für den Bereich des interaktiven Lernens zu eng gewählt ist. Um hier den Analysestand weiter zu vertiefen, soll zunächst der Begriff des interaktiven Lernens definiert werden.

Kiupel (1999) fasst interaktives Lernen als einen Prozess in dem der Lernende mit der Experimentierstation, zum Beispiel aus dem Versuchsfeld der MINIPHÄNOMENTA umgeht. Zwischen dem Lernenden und der Station entwickelt sich dabei ein Wechselverhältnis. Der Lernende konzipiert Ideen, Ansätze und Theorien. Diese überprüft er an der empirischen Realität, dem Exponat. Dies führt sowohl zur Verwerfung von Konzepten als auch zur Generierung neuer Ansätze. Der Prozess der Interaktion endet, wenn der Lernende mit dem Ergebnis zufrieden ist. Diese Zufriedenheit kann schnell erreicht sein, der Prozess kann aber auch, bei hohem Interesse und hoher Motivation, sehr lange andauern. Es handelt sich folglich um einen selbstbestimmten Lernvorgang, bei dem keine externen Lernziele oder Bewertungskriterien existieren. In der psychologischen und pädagogischen Literatur hat es sich eingebürgert solche selbstregulierten, nicht von außen durch formelle Vorgaben gelenkten Lernvorgänge, als „informelles Lernen“ zu bezeichnen (vgl. Zimbardo, Gerrig 1999, S. 205 ff.). Die Prozesse des interaktiven Lernens an den Stationen werden in dieser Arbeit in dem Abschnitt 1.3.2 genauer beschrieben. Die Begriffe des informellen beziehungsweise interaktiven Lernens beschreiben an dieser Stelle lediglich eine erste formale Beschaffenheit des Lernprozesses.

Unter interaktivem Lernen versteht man einen informellen Lernprozess, der auf einer dialogischen „Kommunikation“ zwischen dem Lernenden und der Station beruht.

Nach der Einführung dieses Begriffes unternehme die Leserin oder der Leser ein Gedankenexperiment (vgl. Fiesser 1990, S. 326 ff.): Sie oder er stelle sich die folgende Situation vor: Ein Kind streift durch die Exponate der MINIPHÄNOMENTA, probiert das „Kinorad“ aus, drückt einmal auf den „cartesianischen Taucher“, probiert die drei Kugeln des „Drei-Zeiten-Pendels“ in Phase schwingen zu lassen, schweift dann ab und widmet sich anderen Pausenbeschäftigungen. In den folgenden Pausen macht das Kind etwas Anderes, um sich dann wieder dem Experimentierfeld zuzuwenden. Was hat es nun gelernt?

Viele Antworten sind hier möglich:

- Gesetz des mathematischen Pendels
- Bewegungsgesetze der Kreisbewegung
- Rezeptorträgheit des Auges
- Weckung von Interesse als die Basis für eine weitere Auseinandersetzung mit Phänomenen
- Änderung der Einstellung gegenüber Naturwissenschaft und Technik
- Sammeln wichtiger sinnlicher Erfahrungen
- Formal-logische Kompetenzen, wie zum Beispiel die korrekte Trennung der Variablen

Die Entscheidung für eines dieser Lernziele erscheint unmöglich, da der Lernvorgang interindividuell höchst unterschiedlich verläuft. Auch eine Rangreihe nach der Wichtigkeit des Erreichens der Lernziele ist nicht denkbar, da kein Kriterium ableitbar ist, anhand dessen die Rangreihe gebildet werden könnte. Es kann letztendlich an dieser Stelle nur festgestellt werden, dass - um es mit den Worten Wagenscheins (1992, S. 27 ff.) zu sagen - interaktives Lernen in seinem Ergebnis und Prozess nie programmierbar ist, sondern immer Dunkelheit vor sich hat.

Aus den Definitionen und dem Gedankenexperiment wird der umfassende Charakter des interaktiven Lernens deutlich. Es findet eine Vielzahl von Prozessen statt, die sich nicht mit der auf den Begriff des Wissens fokussierten Konzeption der naturwissenschaftlichen Bildung der PISA-Studie in Einklang bringen lassen. Bezug nehmend auf das obige Beispiel können hier genannt werden: Einstellungsbildung gegenüber Naturwissenschaften und Technik, Motivation, sich mit naturwissenschaftlichen Fragestellungen auseinander zu setzen, Prozesse der Handlungsregulation an der Station und die Sammlung wichtiger sinnlicher Erfahrungen.

Basierend auf dem Begriff des interaktiven Lernens konnte gezeigt werden, dass der Bildungsbegriff der PISA-Studie für die MINIPHÄNOMENTA zu eng gewählt ist.

Der Begriff der Physikalischen Bildung muss daher für den Kontext des interaktiven Lernens deutlich erweitert werden. Dies geschieht, wie bereits in Abschnitt 1.2.1.2 angedeutet, durch die Anknüpfung an die Ideen des deutschen Idealismus (vgl. Reble 2002, S. 174 ff.).

Das elementare Bedürfnis des Menschen nach Bildung entspringt seiner Anthropologie. Kant formuliert diesen Zusammenhang wie folgt: „Die Natur hat gewollt: dass der Mensch alles was über die mechanische Anordnung seines tierischen Daseins geht, gänzlich aus sich selbst herausbringe, und keiner anderen Glückseligkeit, oder Vollkommenheit teilhaftig werde, als die er sich selbst, frei von Instinkt durch eigene Vernunft verschafft hat.“ (ebd. 1968, S. 36). Gehlen formuliert diese zentrale Besonderheit anders, wenn er vom Menschen als dem einzig Freigelassenen der Schöpfung spricht und ihn als organisch mittellos bezeichnet, meint damit aber letztlich das Gleiche (ebd. 1993, S. 6 ff.). Wir sind damit, im Gegensatz zum Tier, als eine *constitutio humana* auf Bildung angewiesen. Nur durch sie kann der Mensch zum Menschen werden.

Wir sind also auf Bildung angewiesen: Sie ist integraler Bestandteil des menschlichen Seins. Doch was ist unter dem Begriff der Bildung zu verstehen? Hier ist eine Vielzahl von Antworten möglich. Ich gehe lediglich auf den klassischen Bildungsbegriff Humboldts ein, der zum einen in hervorragender Art und Weise für die Prozesse des interaktiven Lernens genutzt werden kann, zum anderen die deutsche Bildungsdiskussion und das Schulwesen bis heute wie kein anderer geprägt hat (vgl. Benner 2001, S. 62 ff.).

Dieser definiert das Ziel der menschlichen Bildung, das bei ihm zugleich das allgemeine Ziel des menschlichen Seins ist, als eine Verknüpfung des „ichs mit der Welt zu der allgemeinsten, regesten und freiesten Wechselwirkung“ (ebd. 1980, S. 235 – 236). Wie aber kann dieses Ziel erreicht werden? Der Mensch muss sich die Welt und die Vielzahl der Gegenstände in ihr näher bringen, „diesem Stoff die Gestalt seines Geistes aufdrücken und beide einander ähnlicher machen“ (ebd. 1980, S. 237). Bildung ist damit notwendig ein aktiver Vorgang des Individuums. Sie ist immer Selbstbildung, die auf der Interaktion des Individuums mit der Umwelt basiert. Aus dem obigen Zitat wird ein weiterer besonderer Charakter des Begriffes der Bildung deutlich: Bildung ist Ziel und Prozess zugleich.

Durch die bislang angeführten Zitate könnte der Eindruck entstehen, Humboldt verstehe unter Bildung einen vorrangig rezeptiven Vorgang. Der Mensch steht in Bezug auf die Bildung jedoch unter einem „doppelten Imperativ“ (vgl. Reble, S. 190). Bildung umfasst notwendigerweise immer zwei Dimensionen: Rezeption und Kreation. Beide zusammen bezeichnet Humboldt als „ganze Bildung“ (ebd. 1980, S. 239).

Physikalische Bildung ist in Bezug auf den dargestellten Bildungsbegriff zunächst als eine Einengung des Gegenstandsbereiches zu verstehen. Sie ist zunächst diejenige Bildung in Bezug auf die Aspekte der Physik.

Verlassen wir damit den Bereich der allgemeinen Bildung? Dem Gedanken Humboldts folgend ist dies nicht der Fall: Unter der allgemeinen Bildung versteht Humboldt eine Stärkung des Menschen insgesamt, die spezielle Bildung hingegen soll nur Fertigkeiten und Fähigkeiten zu ihrer Anwendung enthalten (vgl. ebd. 1982, S. 188 ff.). Physikalische Bildung verstehe ich, bezugnehmend auf diese Klassifizierung, als Teil der allgemeinen Bildung. Sie bietet für das Individuum vor allem drei Emanzipationsmöglichkeiten, die zentral zur ihrer/seiner Stärkung beitragen können:

1. Schulung allgemeiner Fähigkeiten: Während der Beschäftigung mit physikalischen Inhalten, vor allem während des Experimentierens trainiert das Individuum allgemeine Fähigkeiten, die ihr oder ihm in anderen Lebensbereichen von Nutzen sein können. Man denke hier zum Beispiel an die Fähigkeit des genauen Beobachtens von Sachverhalten, das gezielte Experimentieren mit dem notwendigen Geschick, die präzise Dokumentation von Sachverhalten usw. Im Gegensatz zu anderen Disziplinen vereint gerade die Physik das praktische Handeln mit kognitiven Prozessen und kann so wesentlich zu elaborierten Handlungsstrategien beitragen.
2. Ausbildung logischer Kompetenzen: Die Orientierung an der Methodik des Experimentes hat natürlich nicht nur Einfluss auf das Handeln, sondern auch auf kognitive Aspekte: Das kontrollierte Nachdenken über physikalische Fragestellungen führt zur Entwicklung zentraler Konzeptionen, wie zum Beispiel dem Begriff der Variablen, der kausalen Verursachung, der Unabhängigkeit usw. Diese Konzepte erhöhen auch in anderen Lebensbereichen die Rationalität des Gedachten und tragen damit entscheidend zur Klärung von Sachverhalten bei.
3. Erhöhung der Selbstwirksamkeit: Die erfolgreiche Lösung physikalischer Aufgaben vermittelt dem Individuum Kompetenzerfahrungen, die zu einer Erhöhung der Selbstwirksamkeit führen. Dieser Gedankengang wird durch die weiteren Erörterungen (insbesondere in den Abschnitten 1.2.2.2 und 1.2.2.5) evident, in denen herausgearbeitet wird, dass es sich bei der Physik in der Einschätzung von Untersuchungsteilnehmerinnen und Untersuchungsteilnehmern um ein „schweres“ Fach handelt. Insofern kommt Erfolgen in diesem Bereich ein besonderes Gewicht zu.

Physikalische Bildung weist damit über ihren eigenen Gegenstandsbereich, eben den der Physik, hinaus. Sie enthält allgemeine Merkmale, die zentral zur Stärkung des Individuums beitragen.

Physikalische Bildung wird durch die Interaktion des Individuums mit der Welt, idealerweise durch das tätige Experimentieren aufgebaut. Sie ist ein Teil der allgemeinen Bildung, dient damit zentral der Stärkung des Individuums.

Physikalische Bildung ist mehr als reines Wissen.

Sämtliche Erfahrungen, die die Schülerinnen und Schüler an den Versuchsstationen machen, und Kompetenzen, die sie erwerben, sind als ein Ausdruck physikalischer Bildung zu sehen.

An dieser Stelle sei nach dem Verhältnis der beiden Begriffe der Physikalischen Bildung der PISA-Studie und seiner Erweiterung in dieser Arbeit gefragt. Bislang könnte der Eindruck entstanden sein, beide würden sich diametral gegenüber stehen. Dies ist jedoch keineswegs der Fall.

Zunächst muss festgestellt werden, dass der Begriff der physikalischen beziehungsweise der naturwissenschaftlichen Bildung der PISA-Studie ein Teil des vorgeschlagenen erweiterten Begriffes der Physikalischen Bildung ist. Die am Begriff des Wissens orientierte Definition der PISA-Studie geht damit in dem hier vorgeschlagenen Begriff der Physikalischen Bildung auf. Weiterhin, und dieser Punkt ist wesentlich bedeutender, sind zentrale Elemente des erweiterten Begriffes der Physikalischen Bildung als Vorbedingung für das Erreichen der naturwissenschaftlichen beziehungsweise Physikalischen Bildung, wie sie in der PISA-Studie definiert wird, zu sehen: Die Entwicklung von Interesse an physikalischen Fragestellungen, die Motivation sich mit derartigen Problemen auseinander zu setzen usw. All dies sind Vorbedingungen, die erreicht werden müssen, damit sich Schülerinnen und Schüler in angemessener Weise mit physikalischen Fragestellungen auseinander setzen können. Der erweiterte Begriff der Physikalischen Bildung geht über die Definition der PISA-Studie hinaus, indem er die Bedingungen, das heißt die Basis einer späteren, inhaltlich orientierten Physikalischen Bildung, wie sie von der PISA-Studie erfasst werden, mit einschließt.

1.2.2 Die Physikalische Bildung im Spiegel empirischer Studien

1.2.2.1 Einleitung

Physikalische Bildung ist, dies machten die Erörterungen der vorherigen Abschnitte deutlich, ein überaus komplexes Phänomen. In der von mir vorgeschlagenen entzieht es sich geradezu einer exakten operationalen Definition. Dennoch können einzelne Aspekte operationalisiert werden, zum Beispiel das Wissen zu physikalischen Fragestellungen und das Interesse an der Physik. Es muss jedoch betont werden, dass diese Variablen zwar Indikatoren Physikalischer Bildung sind, diese jedoch nicht umfassend empirisch beschreiben (vgl. Abschnitt 1.4.1 in dieser Arbeit).

Es existiert eine Vielzahl von Studien, die sich mit den vier in der Tabelle 1.2.2.1.1 genannten Themenbereichen befassen. Leider sind die Stichproben, dies zeigt die Tabelle ebenfalls, wenig homogen, so dass kein umfassendes Bild zu einer bestimmten Bevölkerungsgruppe vorliegt, vielmehr existieren Ergebnisse zu einzelnen Personengruppen:

Untersuchungsbereich	Untersuchung	Beschaffenheit der Stichprobe
Einstellung	Dengler 1995	Studenten bei der Einschreibung, Patienten einer Poliklinik, Besucher des deutschen Museums, Schüler eines Gymnasiums
Motivation	Baumert, Lehmann, Lehrke 1997	Schülerinnen und Schüler der siebten und achten Klassenstufe aller Schularten
Interesse	Hoffmann, Häußler, Lehrke 1998	Schülerinnen und Schüler der fünften bis zehnten Klassenstufe aller Schularten
Kompetenzen	Baumert, Lehmann, Lehrke 1997; Deutsches PISA Konsortium 2001	Schülerinnen und Schüler der siebten und achten Klassenstufe aller Schularten, junge Erwachsene

Tabelle 1.2.2.1.1: Physikalische Bildung im Spiegel empirischer Studien;
- Beschaffenheit der Stichproben der Untersuchungen
Der Verfasser

Gesichert ist die Datenlage entsprechend der Übersicht in Bezug auf Schülerinnen und Schüler der Klassenstufen fünf bis zehn. Beispielhaft schildere ich im Folgenden die Ergebnisse zu den Untersuchungsbereichen Interesse und Wissen. Eine vollständige Diskussion aller Aspekte erscheint nicht notwendig, da die wesentlichen Probleme in Bezug auf die Physikalische Bildung bereits durch diese beiden Aspekte deutlich werden.

Die Ergebnisse zu den folgenden beiden Aspekten der Physikalischen Bildung können als Indikatoren für die geringe Wirksamkeit des physikalischen Unterrichtes verstanden werden, auch wenn an dieser Stelle natürlich betont werden muss, dass in einer postmodernen, pluralen Mediengesellschaft weit mehr Faktoren Einfluss auf den Stand der Physikalischen Bildung haben, als nur der physikalische Unterricht. Dennoch ist er das einzige setting, in dem Schülerinnen und Schülern physikalische Inhalte in institutionalisierter und systematisch aufgearbeiteter Form angeboten werden. Insofern ist es evident anzunehmen, dass er im Vergleich zu anderen Angeboten den umfassendsten Einfluss auf den Stand der Physikalischen Bildung hat. Dem folgenden Abschnitt vorausgreifend, kann das weitgehende Scheitern des herkömmlichen naturwissenschaftlichen Unterrichtes an den Schulen festgestellt werden (vgl. Fiesser 1990, S. 14 ff.).

1.2.2.2 Das Interesse

Bei der Durchsicht gängiger Kompendien der allgemeinen Psychologie (vgl. Zimbardo, Gerrig 1999; Popp 1995; Müsseler, Prinz 2002) fällt auf, dass der Begriff des Interesses in der Psychologie offensichtlich keine zentrale Rolle spielt. Dies liegt an der mangelnden Bestimmtheit des Begriffes: Die Inhalte reichen von ökonomisch-juristischen Deutungen (ökonomisches Interesse, staatliches Interesse usw.) bis zu persönlichen Deutungen wie zum Beispiel Affinität zu bestimmten Themen (vgl. Sieger 1993, S. 15ff.).

Die Problematik des Begriffes wird auch bei spezifisch-psychologischen Definitionen deutlich: Todt definiert in einer älteren Publikation Interessen als „Verhaltens- oder Handlungstendenzen (Dispositionen), die relativ überdauernd sind, [...], die gerichtet sind auf verschiedene Gegenstands-, Tätigkeits- oder Erlebnisbereiche [...]. Bei den Verhaltens- und Handlungstendenzen können zu unterschiedlichen Zeiten und unter unterschiedlichen Bedingungen die kognitiven und die affektiven oder die konnotativen Bedeutungen im Vordergrund stehen.“ (ebd. 1978, S.14).

Unter dem Begriff des Interesses wird in neueren Publikationen Ähnliches verstanden.

Interesse ist ein „Konstrukt einer Disposition zu Verhaltens- oder Handlungstendenzen, die auf Gegenstände, Tätigkeiten oder Erlebnisse gerichtet sind und interindividuell variieren nach Intensität [...].“ (Arnold, Eysenck, Meili 1997).

Der Beginn der Definition durch Todt und die Definition nach Arnold Eysenck und Meili sind nahezu deckungsgleich. In der ersten Definition wird darüber hinaus deutlich, dass ähnlich dem Drei-Komponenten-Modell der Einstellungen (vgl. Rosenberg, Hovland 1960) und Emotionen (vgl. Ulich, Mayring 1992, S. 28ff.) auch Interessen eine solche dreigliedrige Struktur aufweisen sollen. Hier ergibt sich jedoch ein Problem: Wie sind nach dieser Definition Interessen und Einstellungen noch zu unterscheiden? Sind Einstellungen doch ebenfalls definiert als „die Bereitschaft auf kognitiver, affektiver und verhaltensbezogener Ebene bestimmte Reaktionen zu zeigen“ (Hannover, Bettge 1993, S.16).

Auch wenn beide Definitionen implizieren, es gäbe weitgehende Einigkeit darüber, was unter dem Konstrukt des Interesses in der Psychologie zu verstehen sei, schwierig sei lediglich die Abgrenzung zu anderen Konzepten, so muss diese Einigkeit angezweifelt werden. In der Persönlichkeitspsychologie (vgl. Asendorpf 1999, S. 36ff.) versteht man unter Interessen relativ stabile Dispositionen, sogenannte „traits“. Das Interesse gegenüber unterschiedlichen Bereichen lässt sich anhand von psychologischen Tests messen. In der pädagogischen Psychologie hingegen versteht man unter Interesse vielmehr eine aktuelle Befindlichkeit, so zum Beispiel das Interesse an einer bestimmten Aufgabe (vgl. Edelmann 2000, S. 245 ff.).

Trotz dieser Schwierigkeit ist der Begriff des Interesses in dem hier zu untersuchenden pädagogischen Kontext von großer Bedeutung. Sowohl in seiner situativen, als auch in seiner traitpsychologischen Fassung ist er, für die intrinsische Motivation sich mit physikalischen Inhalten auseinander zu setzen, ausschlaggebend (Zimbardo, Gerrig 1999, S. 345ff.; Edelman 2000, S. 240 ff.).

Das Interesse an Naturwissenschaften und vor allem das Interesse an Physik wurde in der IPN-Interessenstudie (vgl. Hoffmann, Häußler, Lehrke 1998) umfassend untersucht.

Bei dieser Untersuchung handelt es sich um eine Längsschnittstudie. Die Erhebung begann im Jahr 1984, mit insgesamt 51 Klassen der Jahrgangsstufe fünf aus allen Schularten. Bis 1989 wurde die Erhebung jährlich wiederholt. Durch dieses Vorgehen konnten Daten aus dem Zeitraum der fünften bis zehnten Klasse gewonnen werden.

Die IPN-Interessenstudie orientiert sich am trait-Konzept des Interessenbegriffes, fasst das Interesse an Naturwissenschaften folglich als ein relativ stabiles Merkmal auf.

Die Ergebnisse wurden in anderen Studien, wie zum Beispiel der TIMSS (vgl. Baumert, Lehmann 1997, S.161 ff.) bestätigt, so dass von ihrer Gültigkeit ausgegangen werden kann. Ich konzentriere mich in meinen Ausführungen auf die deskriptiven Befunde zu den fünf Interessenvariablen. Diese sind:

Variable	Definition
Sachinteresse	Interesse am Gegenstandsbereich Physik
Freizeitinteresse	Interesse, sich in der Freizeit mit physikalischen Sachverhalten zu befassen
Berufsinteresse	Interesse einen physikalisch-technischen Beruf zu ergreifen
Fachinteresse	Interesse am Schulfach Physik
induziertes Interesse	Durch den Physikunterricht hervorgerufenen Interesse

Tabelle 1.2.2.2.1: Das Interesse;
- Definitionen der Interessenvariablen
Der Verfasser

Die folgende Grafik stellt die fünf durchschnittlich erreichten Punktwerte der Interessenvariablen in Abhängigkeit von der Klassenstufe dar. Das Sachinteresse wurde als einzige Interessenvariable bereits ab der 5 Klasse erhoben. Die Erhebung der übrigen Variablen beginnt ab der Klassenstufe 7:

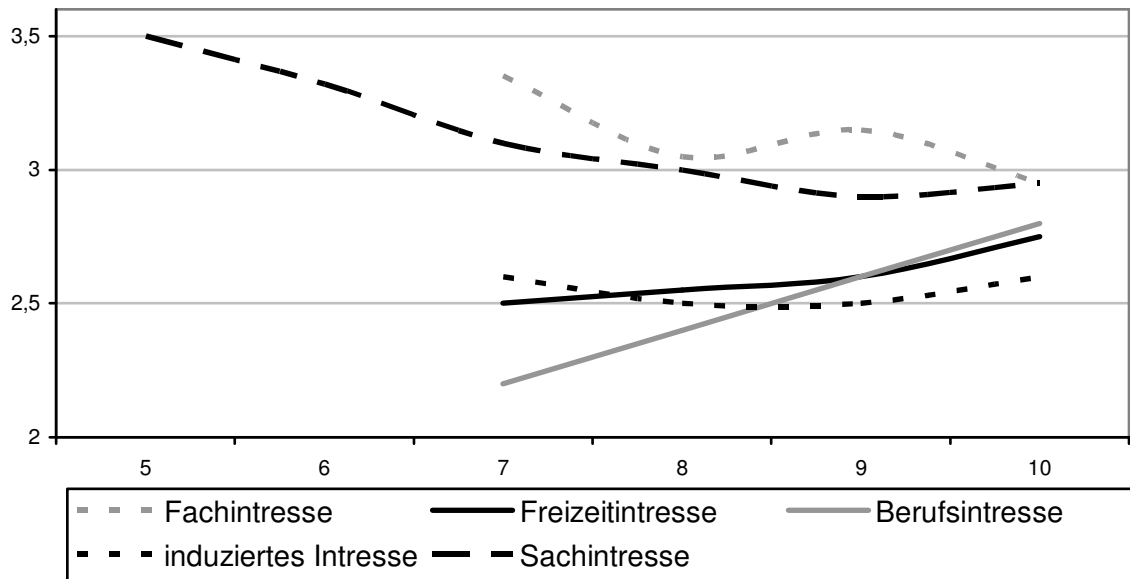


Diagramm 1.2.2.2.1: Das Interesse
 - deskriptiver Verlauf der Interessensvariablen⁸:
 Nach Hoffmann, Häußler und Lehrke 1998

In der älteren physikdidaktischen Forschung stellt Todt in Bezug auf das Interesse der Schülerinnen und Schüler fest: „Sobald die Schüler unser Fach kennen lernen, verlieren sie das Interesse daran“ (ebd. 1991, zitiert nach Kiupel S.23). Entsprechend dem Modell der Kieler Interessenstudie soll dieses ernüchternde Bild nun differenziert werden:

Fachinteresse:

Das anfangs relativ hohe Fachinteresse an Physik sinkt bis zur zehnten Klassenstufe um 0,5 Zählerwerte auf einen Wert von 2,9 ab. Noch deutlicher wird die Lage wenn man das Fachinteresse nach dem Geschlecht differenziert. Das Fachinteresse der Mädchen ist statistisch bedeutsam geringer als das der Jungen. Während das Fachinteresse bei Jungen nach dem Absinken in der siebten und achten Klassenstufe bis zur zehnten Klasse fast wieder das Ausgangsniveau erreicht, bleibt das Interesse der Mädchen konstant auf dem niedrigen Niveau der vorherigen Klassenstufen.

Weiterhin konnte festgestellt werden, dass Jungen zwar das Fach Physik als die uninteressanteste Naturwissenschaft einstufen, die Unterschiede im Interesse in Bezug auf die unterschiedlichen naturwissenschaftlichen Fächer aber statistisch nicht bedeutsam sind. Ganz anderes bei den Mädchen: Das Interesse schwankt bei ihnen in Bezug auf die unterschiedlichen naturwissenschaftlichen Fächer signifikant: Biologie ist am beliebtesten, Physik am unbeliebtesten.

⁸ Die Werte der Grafik wurden durch den Autor zeichnerisch aus Werten der Publikation der IPN-Interessenstudie ermittelt. Sie liegen zwischen 1 und 4.

Freizeitinteresse:

Bezüglich des Interesses, sich in der Freizeit mit naturwissenschaftlich-technischen Inhalten zu befassen, konnten - beruhend auf 13 Items - drei Faktoren extrahiert werden:

1. Informationsinteresse in der Freizeit an Physik und Technik
2. Interesse am spielerischen Umgang mit Physik und Technik
3. Praktisches Interesse an technischen Sachverhalten

Für alle drei Faktoren ergeben sich sehr niedrige Werte. So liegt zum Beispiel das Interesse an einem spielerischen Umgang mit Physik und Technik bei Mädchen in der siebten Klassenstufe bei nur 1,8. In der 10 Klasse ist es um weitere 0,4 Zählerpunkte auf 1,4 gefallen. Das Interesse von Jungen ist höher. Insgesamt scheint aber eine private Beschäftigung mit physikalischen Inhalten äußerst selten zu sein.

Berufsinteresse:

Das Interesse an einem physikalischen Beruf steigt in der Zeitspanne von der siebten bis zur zehnten Klasse deutlich an. Die Autoren der Studie lassen diesen Effekt uninterpretiert. Möglich wäre eine Erklärung durch die Spezifizierung der Berufinteressen in dieser Alterspanne (vgl. Oerter, Montada 1995, S. 423ff.). Hier sind die Unterschiede zwischen Mädchen und Jungen noch deutlicher: Ist das Interesse der Jungen in der siebten Klasse schon deutlich höher und steigt weiter, sinkt das der Mädchen mit steigender Klassenstufe stetig.

Induziertes Interesse:

Das durch den Physikunterricht induzierte Interesse an der Physik ist gering. Bei Jungen und Mädchen fällt das Interesse zunächst ab, um in der neunten und zehnten Klassenstufe wieder anzusteigen. Die Werte sind insgesamt niedrig, nahe dem Mittelwert der Skala. Der Unterschied zwischen Mädchen und Jungen ist auch hier statistisch bedeutsam.

Sachinteresse:

Das Sachinteresse fällt von der fünften Klasse bis zur zehnten Klasse von einem hohen auf einen knapp über dem Durchschnitt liegenden Wert. Die Unterschiede zwischen Mädchen und Jungen fallen hier deutlich geringer aus, als zum Beispiel beim Fachinteresse. Im weiteren Verlauf gleichen sich die beiden Kurven an. Das Sachinteresse variiert nicht zwischen den Schulformen. In der TIMSS wurden sogar noch geringere Werte bezüglich des Sachinteresses von Schülerinnen und Schülern in der achten Klasse gemessen. Diese Werte liegen deutlich unterhalb des Neutralitätspunktes der Skala (vgl. Baumert, Lehmann 1998, S. 169).

Zusammenfassend bewerte ich vorrangig zwei Ergebnisse als problematisch:

1. Die immensen Unterschiede im Interesse zwischen Mädchen und Jungen:
Während sich die Interessenlage von Jungen nach der siebten und achten Klasse nach einem Abfall weitgehend stabilisiert, bleiben die unterschiedlichen Formen des Interesses bei den Mädchen größtenteils niedrig oder sinken zum Teil weiter. Ihr Interesse an physikalisch-technischen Fragestellungen ist nur gering und der Physikunterricht scheint dieses Interesse nicht zu fördern. Im Gegenteil: Es nimmt durch ihn weiter ab. Offenbar gelingt es im Rahmen des Physikunterrichts nicht, die Interessen der Mädchen aufzugreifen und auszubauen.
2. Das allgemein niedrige Niveau des Interesses an der Physik: Vor allem den kontinuierlichen Abfall des Sachinteresses (besonders ausgeprägt in den Klassenstufen fünf bis sieben) sehe ich als problematisch an.

1.2.2.3 Das Wissen

Eine ausführliche Diskussion des dieser Arbeit zu Grunde liegenden Wissensbegriffes erscheint an dieser Stelle nicht erforderlich. Sie wird im Abschnitt 1.4.7 vorgenommen, so dass hier nur eine kurze Definition geliefert werden soll:

Unter Wissen über einen bestimmten Inhalt verstehe ich im Rahmen dieser Arbeit die Summe der im Gedächtnis verfügbaren Propositionen und Verknüpfungen, die in der Form eines semantischen Netzwerkes gespeichert sind. Es soll zwischen deklarativem und prozeduralem Wissen unterschieden werden.

Born und Euler diagnostizieren im Hinblick auf die naturwissenschaftlichen Kompetenzen von Jugendlichen in einer, in den 70er Jahren erschienen Untersuchung „die sehr hoch gesteckten Lehrziele mancher Curricula und Rahmenlehrpläne seien (der Verfasser) [...] Etikettenschwindel“ (ebd. 1978 S. 76). Ältere Untersuchungen scheinen dieses Bild zu bestätigen. Daumenlang (1969) verglich das Wissen von 13-jährigen über physikalische Alltagsphänomene mit dem von jungen Erwachsenen. Er stellte fest, dass kein Effekt der schulischen Beschäftigung mit physikalischen Inhalten nachzuweisen sei. Die Erklärungen der 13-jährigen unterschieden sich nicht signifikant von denen der jungen Erwachsenen.

Brämer (1980, S. 15) verweist auf methodische Probleme der Studie, vor allem auf die Forderung nach einer physikalischen Erklärung, hin. Betrachtet man weitere Untersuchungen, zum Beispiel die von Dengler (1995, S. 27 ff.), so wird deutlich, dass die Untersuchung von Daumenlang weitere Schwierigkeiten aufweist. Der Physikunterricht der Mittelstufe und im noch stärkeren Maße der der Oberstufe, hat nur wenig Bezug zum Alltag. Insofern könnten die scheinbar leichten, lebensweltnahen Aufgaben für die Schülerinnen und Schüler und die jungen Erwachsenen besonders schwer gewesen sein, da der Einfluss des Physikunterrichts auf etwas gemessen wurde, auf das dieser aufgrund seiner Beschaffenheit keinen Einfluss hat.

Neben einer Diskussion der methodischen Probleme der Untersuchung muss gefragt werden, ob die ermittelten Ergebnisse heute noch aktuell sind. Zur Klärung dieser Frage sollen wichtige Ergebnisse der beiden großen Studien TIMSS und PISA herangezogen werden. An dieser Stelle ergibt sich jedoch ein weiteres Problem: Die Ergebnisse der Studien sind nicht nach der üblichen Fächersystematik gegliedert⁹. Der naturwissenschaftliche Kompetenzbereich umfasst vielmehr die Disziplinen Physik, Chemie, Biologie und Geowissenschaften insgesamt. Fächerspezifische Aufteilungen der Ergebnisse sind nur für wenige und sehr spezielle Fragestellungen vorhanden. In den genannten Studien wird somit ein Gesamturteil zur naturwissenschaftlichen Bildung erhoben.

Eine direkte Verwendung der Ergebnisse ist aus meiner Sicht in Bezug auf die Untersuchungsfrage nicht möglich: Zentral unterscheidet sich die Physik von den anderen Naturwissenschaften durch ihre starke Verbindung mit der Mathematik (vgl. Dengler 1995, S. 27 ff.; Abschnitt 1.2.2.2 in dieser Arbeit). Insofern sollen die Ergebnisse zum Kompetenzbereich Naturwissenschaften der PISA-Studie hier nicht mit einbezogen werden. Dadurch würde ein verzerrtes Bild der Sachlage erzeugt werden. Es bestünde die Gefahr einer Überschätzung der naturwissenschaftlichen Fähigkeiten der Schülerinnen und Schüler.

Möglich wird die Verwendung der Ergebnisse der PISA-Studie durch die Publikation „PISA 2000 Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland“ (vgl. Deutsches PISA-Konsortium 2001, A). Hier werden die fachspezifischen Kompetenzen in Ansätzen aufgeschlüsselt. Leider bleiben dabei Physik und Chemie aufgrund methodischer Probleme weiterhin undifferenziert. Die biologische Kompetenz wird hier jedoch einzeln angegeben. Aufgrund der relativen Nähe der Disziplinen Physik und Chemie können die Ergebnisse dennoch berücksichtigt werden¹⁰.

⁹ Dieser Umstand resultiert aus der Tatsache, dass die naturwissenschaftlichen Fächer im europäischen und im nicht europäischen Ausland zumeist gemeinsam, als integriertes naturwissenschaftliches Fach unterrichtet werden. Insofern erscheint eine Aufteilung aus internationaler Sicht tatsächlich wenig sinnvoll.

¹⁰ In der Trias der drei klassischen Naturwissenschaften nimmt insbesondere die Biologie eine Sonderstellung ein. Dies beweisen zahlreiche Untersuchungen im Hinblick auf unterschiedliche Aspekte: Dengler (1995) konnte nachweisen, dass die Einstellung gegenüber der Biologie positiv ist. Sie wird als interessant und lebensnah empfunden. Auch in der PISA-Studie und der TIMSS konnte die Sonderstellung der Biologie bestätigt werden. Die Schülerinnen und Schüler sind in diesem Fach höher motiviert und zeigen bessere Leistungen (vgl. Deutsches PISA-Konsortium 2001; Baumert, Lehmann 1997).

Bevor mit der Schilderung von Ergebnissen begonnen werden kann, möchte ich zunächst einige Anmerkungen zu den konzeptionellen Grundlagen der Studie machen: Untersucht wurden in der PISA-Studie (2000) 15- bis 16-jährige Schülerinnen und Schüler aller Schularten. Der internationale Aufgabenteil für den Bereich der Naturwissenschaften beanspruchte dabei eine Bearbeitungszeit von 60 Minuten. Aus 13 Themengebieten wurden insgesamt 35 Aufgaben gestellt. Durch die Rasch-Skalierung des Tests und das damit verbundene Multi-Matrix-Sampling gelang es, die effektive Bearbeitungszeit der Naturwissenschaftsaufgaben durch die einzelne Schülerin oder den einzelnen Schüler auf 30 Minuten zu kürzen, indem nicht jede Schülerin oder jeder Schüler alle Aufgaben bearbeiteten. Die Punktwerte des Tests sind normalverteilt. Sie wurden auf ein arithmetisches Mittel von 500 Testpunkten und eine Standardabweichung von 100 Punkten geeicht.

Im Bereich Physik/Chemie erreichten die Schülerinnen und Schüler im Durchschnitt der Bundesländer (ohne Berlin und Hamburg) lediglich 491 Testpunkte. Damit befindet sich der Durchschnitt aller deutschen Schülerinnen und Schüler lediglich in der Kompetenzstufe II (funktional naturwissenschaftliches Alltagswissen) (vgl. Deutsches PISA-Konsortium 2001 A, S. 16 ff.). Damit gelingt es ihnen lediglich einfache naturwissenschaftliche Fragen zu benennen, einfache Schlussfolgerungen zu ziehen und naturwissenschaftliches Alltagswissen anzuwenden.

Die Gymnasiasten schnitten bei diesem Test etwas besser ab. Sie erreichten im Durchschnitt die Kompetenzstufe III (funktional naturwissenschaftliches Wissen): Ihnen gelingt es, naturwissenschaftliche Konzepte anzuwenden, um Erklärungen zu geben und Vorhersagen zu treffen.

Alle weiteren Daten der Studien beziehen sich leider auf den Gesamtbereich der naturwissenschaftlichen Bildung, können hier folglich nicht verwandt werden. Bereits die beiden genannten Ergebnisse belegen das niedrige Kompetenzniveau deutscher Schülerinnen und Schüler in diesem Bereich. Sie befinden sich im internationalen Vergleich im Bereich des unteren Durchschnitts der Rangreihe.

Diese Befundlage kann durch die Ergebnisse der TIMSS vertieft werden: Zunächst zu den konzeptionellen Grundlagen der Studie: Es gibt zwei Messzeitpunkte: In der siebten und achten Klassenstufe. Den Schülerinnen und Schülern wurden 70 offene und geschlossene Fragen vorgelegt, die in insgesamt 90 Minuten bearbeitet werden sollten. Die verwendete Skala wurde auf einen Mittelwert von 500 Testpunkten standardisiert, mit einer Standardabweichung von 100 Testpunkten. Die Aufgaben entstammen den Lernzielen zu den wichtigsten, in der siebten und achten Klasse unterrichteten, Bereichen der Physik. Wichtig dabei ist, dass keine Formeln erfragt und keine Rechenoperationen abverlangt wurden. Befragte Lehrerinnen und Lehrer bewerteten die Fragen damit als eher leicht.

Auch in Bezug auf diese Studie besteht das oben geschilderte Problem der Kombination der naturwissenschaftlichen Fächer zu einem großen Bildungsbereich. Nach Fächern aufgegliederte Ergebnisse liegen nur in wenigen Fällen vor. Die verfügbaren Ergebnisse seien im Folgenden beschrieben:

Vergleicht man die Leistungen deutscher Schülerinnen und Schüler mit denen aus anderen Ländern, so kann festgestellt werden, dass deutsche Schülerinnen und Schüler mit im Durchschnitt 57% richtig gelösten Aufgaben im unteren Durchschnitt der Rangreihe stehen.

Besonders bedenklich ist weiterhin das geringe Verständnis von Schülerinnen und Schülern für die Grundprinzipien des naturwissenschaftlichen Arbeitens. Nur etwa 5% der Achtklässler verfügen über einigermaßen zutreffende Vorstellungen darüber was naturwissenschaftliches Experiment ausmacht.

Die Leistungsentwicklung in Abhängigkeit von der Schulformen und der Klassenstufe scheint auf den ersten Blick erfreulich. Es ist ein stetiger Anstieg der Kompetenzen in allen Schulformen vom Beginn der siebten bis zum Ende der achten Klassen zu beobachten. Entsprechend der unterschiedlichen Aufgaben der Schulform beginnen die Schülerinnen und Schüler bei unterschiedlichen Leistungswerten. Auch der Anstieg der Leistungen ist - entsprechend der Schulform - unterschiedlich stark ausgeprägt. Vergleicht man den Leistungsanstieg mit dem anderer Fächer, zum Beispiel aus dem sprachlichen Bereich, so fällt auf, dass der Lernerfolg im sprachlichen Bereich, vor allem auf dem Gymnasium und auf der Realschule deutlich ausgeprägter ist.

Ein letzter bedenklicher Punkt sind die Leistungsunterschiede im Fach Physik in Bezug auf das Geschlecht. Mädchen erzielen signifikant niedrigere Testleistungen als Jungen.

Insgesamt müssen die Kompetenzen deutscher Schüler im Fach Physik als besorgniserregend eingeschätzt werden:

- Nur etwa 5% deutscher Schülerinnen und Schüler haben in der SEK I Kenntnisse über naturwissenschaftliches Arbeiten.
- Relativ geringe Leistungssteigerung, vor allem am Gymnasium und an der Realschule in Relation zu anderen Fächern
- Statistisch bedeutsame Unterschiede in der Leistung von Jungen und Mädchen
- Insgesamt befinden sich die meisten deutschen Schülerinnen und Schüler in den Fächern Physik und Chemie auf der Kompetenzstufe II.

1.2.3 Physikalische Bildung in der Primarstufe

Nachdem der Begriff der Physikalischen Bildung definiert wurde und nachdem gezeigt werden konnte, dass der Physikunterricht in Bezug auf die beiden untersuchten Variablen Wissen und Interesse wenig wirksam ist, muss an dieser Stelle gefragt werden, ob und wie physikalische Phänomene Einzug in den Unterricht der Primarstufe finden. Denn neben der Notwendigkeit den „Physikunterricht neu zu denken“ (Braun 1995, S. 14 ff.) scheint eine frühe Beschäftigung mit physikalischen Inhalten geboten. Hierüber sind sich nahezu alle Autoren der Physikdidaktik und der empirischen Bildungsforschung einig (vgl. beispielsweise Dengler 1995, S. 28f.; Rost, Prenzel, Carstensen, Senkbeil, Groß 2004, S. 117 ff.).

Diese Forderung wird auch durch die Ergebnisse der Evaluation des Programms „Versuch macht klug“, einer Kooperation der Universität Flensburg, der Nordmetall-Stiftung und der Vereinigung Hamburger Kindertagesstätten gestützt: Die Kinder verweilten über große Zeiträume an den Exponaten¹¹, beobachteten sorgfältig, suchten aktiv nach eigenen Erklärungen und diskutieren ihre Ergebnisse in Kleingruppen. Anfängliche Befürchtungen die Stationen könnten zu Spielobjekten umfunktioniert werden, erwiesen sich als unbegründet. Eindrucksvoll ließen sich auch die große Konzentration, mit der die Kinder arbeiteten, die überraschend elaborierten geistigen Fähigkeiten, sowie die Kompetenz zur Gruppenarbeit nachweisen (vgl. Fiesser, Philippi, Schließmann 2005, S. 16 ff.). Insgesamt deuten die Ergebnisse darauf hin, dass die Kinder ihre Physikalische Bildung an den Stationen entscheidend ausbauen konnten.

Insgesamt sprechen die Ergebnisse dafür, dass Kinder im Alter von 3 bis etwa 10 Jahren ein besonderes Interesse an naturwissenschaftlichen Fragestellungen haben. Es scheint in Bezug auf derartige Inhalte im Kindesalter eine „sensible Periode“ (vgl. Oerter, Montada 1995, S. 84ff.) vorzuliegen.

Wie greift die Grundschule diese Interessenlage der Schülerinnen und Schüler nun auf? Die Lage ist paradox: Unterrichtszieldefinitionen und schulischer Alltag klaffen weit auseinander.

¹¹ Bei den hier angesprochenen Exponaten handelt es sich um interaktive Experimentierstationen, die sich am Vorbild der MINIPHÄNOMENTA (vgl. Fiesser 2005) orientieren.

Der Lehrplan des Faches Heimat- und Sachunterricht des Ministeriums für Bildung und Frauen des Landes Schleswig-Holstein für die Primarstufe (vgl. <http://www.schooloffice-sh.de/download/lehrplan/primar/hsu.pdf> 05.09.2006, 12.00 Uhr) wird dieser Aufgabe gerecht. In ihm wird als ein Unterrichtsschwerpunkt der Umgang mit physikalisch-chemisch-biologischen Fragestellungen definiert. Dieses Ziel wird in den Lernfeldern fünf (Natur und Umwelt) und sechs (Technik, Medien und Wirtschaft) konkretisiert: Die Schülerinnen und Schüler sollen sich am Ende der zweiten Klassenstufe in der „[...] Begegnung mit naturwissenschaftlichen Phänomenen angemessen verhalten [...]“ (vgl. ebd. S.116). Bis zum Ende der vierten Klasse sollen sie einfache naturwissenschaftliche Gesetzmäßigkeiten erklären können, sowie in die Lage sein, selbstständig und/oder unter Anleitung zu experimentieren.

Wird dieser Lehrplananspruch nun auch in Bezug auf die naturwissenschaftliche Bildung im Unterricht in der Primarstufe umgesetzt? Die Untersuchungen von Sauer (2005, S. 23 ff.) ergaben, dass dies nicht der Fall ist. Vielmehr kommt er zu dem ernüchternden Ergebnis: Der größte Teil der Schülerinnen und Schüler hat in der Primarstufe nie experimentiert. Wie so oft in der Physikdidaktik festgestellt, handelt es sich in Bezug auf die Unterrichtsinhalte und Lernziele offenbar auch in der Primarstufe nicht um eine realistische Beschreibung des Unterrichtsalltages, sondern vielmehr um einen „Etikettenschwindel“.

1.3 Die MINIPHÄNOMENTA als ein Beitrag zur Physikalischen Bildung

1.3.1 Das Projekt

Die MINIPHÄNOMENTA besteht derzeit aus 53 transportablen Experimentierstationen. In Anlehnung an Fiesser (1990, S. 26 ff.) können diese in sechs Kategorien eingeteilt werden.

Stationen hinsichtlich:

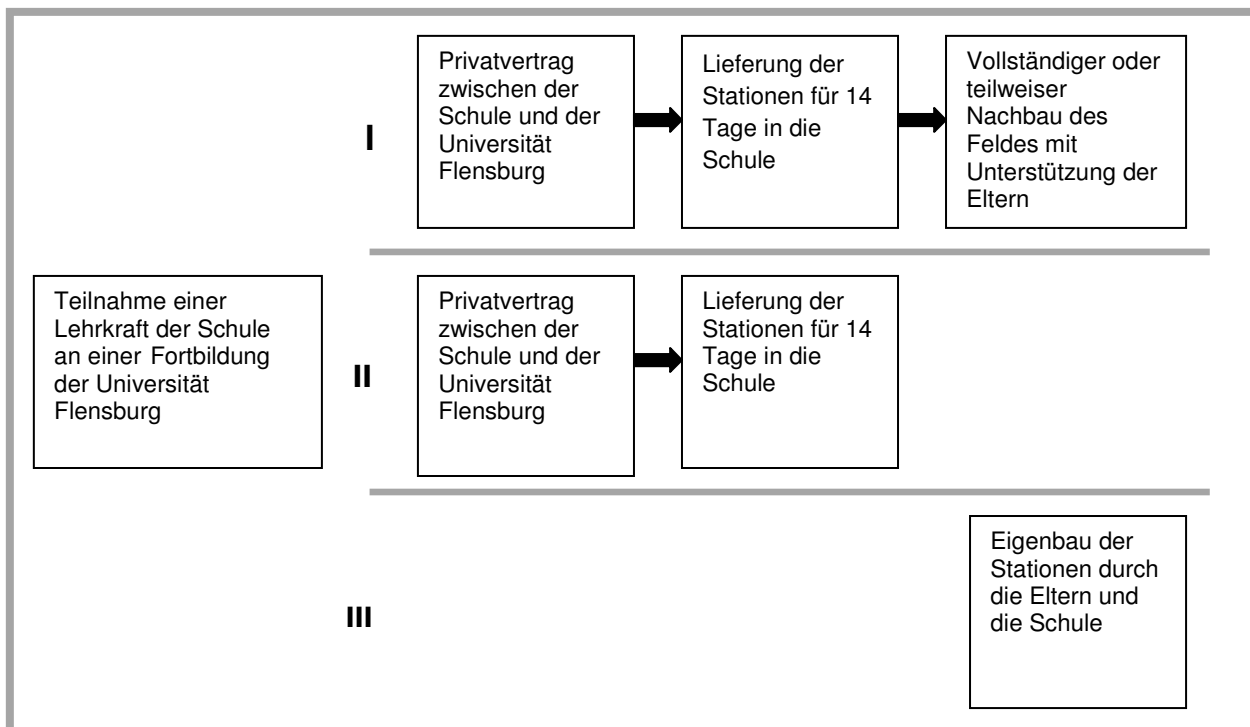
1. der Schulung und Bewusstmachung der Sinne
2. einer hohen ästhetischen Qualität
3. der Quantifizierung des eignen Körpers und der Sinneswahrnehmungen
4. Wahrnehmungstäuschungen
5. physikalischer Phänomene im engeren Sinne
6. chemischer Versuche zu „Stoffeigenschaften“

Dabei dienen die Stationen nicht primär dem unterrichtlichen Einsatz sondern stehen den Schülerinnen und Schülern zum Experimentieren im Rahmen ihrer Pausenbeschäftigung zur Verfügung. ASIP zeichnet sich wesentlich durch seinen Angebotscharakter aus. Die Kinder sind nicht verpflichtet, sich mit den Stationen überhaupt beziehungsweise in einer bestimmten Art und Weise zu befassen. Vielmehr ist jede Form der Beschäftigung mit ihnen intrinsisch motiviert und selbstgesteuert.

Neben den in Abschnitt 1.3.2.2 zu nennenden Spezifikationen von interaktiven Stationen, nennt Holst (2005, S. 70 ff.) folgende Merkmale, die gerade für den Einsatz der Exponate in der Schule im Rahmen der MINIPHÄNOMENTA bedacht werden sollten:

- Einfacher und preisgünstiger Nachbau
- Haltbarkeit des Exponats
- Beachtung der anwendbaren Sicherheits- und Entsorgungsrichtlinien
- Die Station muss sicher und frei im Raum stehen können.
- Die Stationen müssen einfach transportiert und gut verstaut werden können.
- Der für die Stationen notwendige Betreuungsaufwand darf nur gering sein. Dies ist besonders für „Chemie-Stationen“ schwer zu realisieren.

Wünschen sich Lehrerinnen und Lehrer, Eltern oder Schülerinnen und Schüler die MINIPHÄNOMENTA in der Schule, so gibt es mehrere mögliche Vorgehensweisen. Die folgende Grafik stellt Variablen und Konstanten dar:



Grafik 1.3.1.1: Die MINIPHÄNOMENTA;
- Möglichkeiten der Umsetzung von ASIP in den Schulen
Der Verfasser

Die Grundlage der MINIPHÄNOMENTA besteht in der Teilnahme von ein bis zwei Lehrerinnen oder Lehrern an einer zweitägigen Fortbildung. Hier sollen sich die Kursteilnehmerinnen und Kursteilnehmer praxisnah mit dem Konzept von ASIP auseinandersetzen. Neben dem Kennen lernen des pädagogischen Ansatzes der MINIPHÄNOMENTA bauen die Lehrerinnen und Lehrer in Kleingruppen eine erste Versuchsstation aus dem Experimentierfeld des Projektes. Ausgestattet mit diesem Wissen und ersten Erfahrungen sind die so fortgebildeten Lehrkräfte Multiplikatoren an ihrer jeweiligen Schule für die durch die MINIPHÄNOMENTA vertretenen Konzepte. Nach diesem ersten verpflichtenden Teil gibt es drei Möglichkeiten für die weitere Teilnahme am Projekt. Am häufigsten wird im Anschluss an die Fortbildung das Experimentierfeld der MINIPHÄNOMENTA für vierzehn Tage ausgeliehen. Entsprechend der Platzverhältnisse der Schule werden bis zu 53 Stationen in den Schulfluren aufgestellt. Im Rahmen des Vertrages zwischen der Universität Flensburg und der ausleihenden Schule verpflichtet sich diese zur Einhaltung bestimmter Rahmenbedingungen, wie zum Beispiel zum Sicherstellen von Aufsichten, zum Nachkauf von Verbrauchsmaterialien, zur Veranstaltung eines Elterninformationsabends usw. Insbesondere der Elternabend ist von herausragender Bedeutung: Zum einen sollen die Eltern über das Projekt informiert werden, zum anderen können so Elternteile für den Nachbau der Stationen gewonnen werden (Variante I). Durch dieses Vorgehen kann es gelingen Physikalische Bildung dauerhaft in den Alltag der Schule einzubeziehen. Schlägt dieser Versuch fehl, gelingt also nur der Nachbau sehr weniger beziehungsweise keiner Stationen, liegt Variante II vor. Diese Situation entsteht jedoch relativ selten. Ein eigenständiger Nachbau der Stationen ohne vorherige Ausleihe ist ebenfalls selten (Variante III).

Eine Besonderheit der MINIPHÄNOMENTA ist die aktive Einbeziehung von Eltern in den Bau der Exponate. Die Erfahrungen zeigen, dass die Eltern in einem großen Umfang in den Nachbau der Stationen und damit auch in das schulische Geschehen eingebunden werden können. So ergibt sich die Möglichkeit das ansonsten oft schwierige Verhältnis von Eltern und Schule durch positive Impulse nachhaltig zu verbessern (vgl. John 2004).

Eine umfassende Beschreibung der Bauweise, der Funktion und des theoretisch-physikalischen Hintergrundes aller Exponate erscheint hier nicht notwendig. Beide Aspekte sind bei Sauer (2005, S. 57 ff.) und Holst (2005, S. 73 ff.) bereits umfassend erörtert worden, so dass sie hier nicht wiederholt werden müssen. Vielmehr soll in dieser Arbeit, bevor ich mit der Schilderung des lerntheoretischen Hintergrundes fortfahre, eine Station exemplarisch beschrieben werden. Wegen ihrer charakteristischen Eigenschaften entscheide ich mich für das Experiment „Farben mischen“.

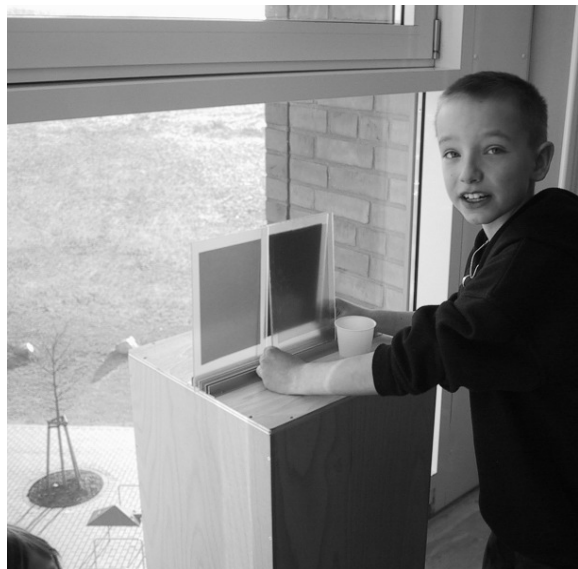


Bild 1.3.1.1: Die MINIPHÄNOMENTA;
- die Farbrahmen
Institut für Physik und Chemie und ihre Didaktik

Basierend auf der oben genannten Klassifikation muss das Exponat der Kategorie fünf (Physikalische Phänomene im engeren Sinne) zugeordnet werden. Mit ihm können die Schülerinnen und Schüler den Vorgang der subtraktiven Farbmischung experimentell nachvollziehen. Das Experiment befindet sich wie alle Exponate der MINIPHÄNOMENTA auf einem Podest, so dass die Schülerinnen und Schüler auf einer für sie angenehmen Höhe arbeiten können. Bei der Größe des Exponates wurde weiterhin beachtet, dass sie zum einen klein genug sind, um sicher transportiert zu werden, zum anderen groß genug, dass mehrere Kinder, im Falle des hier beschriebenen Exponates bis zu vier Schülerinnen und Schüler, gemeinsam experimentieren können. Der Austausch der Kinder über den Versuch, den die Möglichkeit eines gemeinsamen Experimentierens eröffnet, ist gerade für das Lernen an den Stationen besonders wichtig.

Die Stationen der MINIPHÄNOMENTA stellen insgesamt ein aufbereitetes Lernsetting dar, das sich hervorragend mit dem von Rousseau geprägten Begriff der negativen Erziehung in Einklang bringen lässt. Emil wird weitestgehend nicht über Belehrungen erzogen sondern vielmehr durch ein vollständig pädagogisch durchgestaltetes setting. So findet aus der Sicht des Emil keine Erziehung statt. Vielmehr ist die pädagogische Gesamtsituation so gestaltet, dass dem Kind alle Reaktionen auf sein Tun als natürliche Konsequenzen erscheinen, die vordergründig nicht mit der Person des Erziehenden in Verbindung stehen. Die Versuche der MINIPHÄNOMENTA sind das Ergebnis eines Designprozesses mit dem Ziel den Schülerinnen und Schülern ein naturwissenschaftliches Phänomen derart aufzubereiten, dass es für sie handhabbar und erfahrbar wird. Ausgangspunkt ist dabei der Begriff der Transparenz, der hier wie folgt definiert wird:

- Es befinden sich keine versteckten Hebel oder Mechanismen an der Station, vielmehr soll das zugrunde liegende Phänomen, hier die subtraktive Farbmischung, verständlich und fassbar gemacht werden.
- Die Stationen sind insgesamt schlicht gehalten. Lediglich die Teile mit denen die Kinder experimentieren sollen, hier die Farbrahmen, sind hervorgehoben.
- Die Schülerinnen und Schüler sollen am Exponat konkret tätig werden können. Es sollen eine Vielzahl von Variablen, hier die Stellung der Farbrahmen zueinander, möglichst frei veränderbar sein.
- An den Stationen befinden sich keinerlei Experimentieranleitungen oder sonstige Vorgaben wie mit dem Versuch umgegangen werden soll. Vielmehr soll der Umgang der Schülerinnen und Schüler mit den Stationen intuitiv erfolgen. Damit dies möglich ist, müssen die oben genannten Punkte erfüllt sein.
- Die Schülerinnen und Schüler sollen sich während des Experimentierens ihren eigenen Umgang mit dem Exponat erschließen und auf dieser Basis über subjektiv tragfähige Erklärungen für die auftretenden Phänomene nachdenken.

Mittlerweile konnte der Umfang des Projektes erheblich ausgeweitet werden. Die MINIPHÄNOMENTA begann 2004 in Schleswig-Holstein als kleines, lokal relativ begrenztes Projekt. Bis heute hat sich die Konzeption in Deutschland enorm ausgeweitet, so dass in vielen Bundesländern für Schulen die Möglichkeit der Teilnahme besteht.

Bei der MINIPHÄNOMENTA handelt es sich um ein physikalisches Bildungsprojekt, das den Schülerinnen und Schülern Erfahrungen durch offene Experimentierstationen ermöglichen möchte. Das Projekt versteht sich dabei als eine Antwort auf die Ergebnisse der PISA-Studie und der Probleme bei der Initiierung und Umsetzung physikalischer Bildungsprozesse in der Primarstufe.

1.3.2 Der theoretische Begründungszusammenhang der MINIPHÄNOMENTA

1.3.2.1 Einleitung

Das Ziel des folgenden Kapitels besteht darin, die theoretischen Grundlagen der MINIPHÄNOMENTA herauszuarbeiten. Dabei muss betont werden, dass es keine einzelne, grundlegende Theorie der MINIPHÄNOMENTA gibt. Vielmehr existieren vielfältige theoretische Bezüge, die in zwei Aspekten geordnet werden können:

1. Die Beschaffenheit der Experimentierstationen der MINIPHÄNOMENTA
2. Das Lernen an den Stationen der MINIPHÄNOMENTA

Ad 1: Die Beschaffenheit der Experimentierstationen lässt sich aus der reformpädagogischen Konzeption Martin Wagenscheins ableiten. Zusammenfassend kann festgestellt werden, dass eine Stationen ein Phänomen aus der Lebenswelt der Schülerinnen und Schüler auf interessante und spannende Art und Weise darstellen soll. Die Gestaltung des Versuches soll die Kinder zum konkreten Handel, das heißt zum Experimentieren anregen. Dazu muss sie klar gestaltet sein und eine Vielzahl von konkreten Handlungsalternativen bieten. Diese Theorie betont darüber hinaus die Notwendigkeit der offenen Gestaltung von physikalischen Lehr-Lernprozessen, indem sich der Lehrende deutlich zurücknimmt, sämtliche Prozesse konsequent vom Kind aus denkt und die Lernenden selbst aktiv werden lässt.

Ad 2: Bei dem Lernen an den Stationen handelt es sich um einen hochkomplexen Lernvorgang. Im bisherigen Textverlauf wurde dieser lediglich durch die beiden Begriffe des interaktiven Lernens und des informellen Lernens beschrieben. Der Erstgenannte betont dabei den „dialogischen Prozess“ zwischen dem Lernenden und der Station in der Situation des Experimentierens: Es werden Theorien entworfen und an der empirischen Realität überprüft. Auf dieser Basis werden neue Ideen konzipiert usw. Dieser Prozess dauert solange an, bis der Lernende mit seiner Lösung zufrieden ist. Unter informellem Lernen wird dabei ein Lernprozess verstanden der weitgehend durch das Individuum gesteuert wird und frei ist von externen Bewertungskriterien. Beide Begriffe sind damit nicht disjunkt trennbar. Sie weisen vielmehr eine breite inhaltliche Schnittmenge auf, die in der Selbstregulation des Lernprozesses liegt. Der inhaltliche Schwerpunkt des interaktiven Lernens liegt darüber hinaus auf dem Interaktionsprozess des Lernenden mit der Station, der Begriff des informellen Lernens betont hingegen die Abwesenheit externer Bewertungskriterien.

In den folgenden Abschnitten gilt es diesen Erkenntnistand zu vertiefen. Dies geschieht durch zwei theoretische Ansätze: Kognitionspsychologie und Handlungstheorie.

Die kognitiven Prozesse während des interaktiven Lernens sollen durch eine kritische Auseinandersetzung mit der Theorie Piagets beschrieben werden. Auch wenn viele seiner Forschungsergebnisse nicht mehr dem aktuellen Stand der Entwicklungspsychologie entsprechen, so bietet seine Theorie, erweitert um aktuelle Ansätze, dennoch ein großes Erklärungspotential, das im Rahmen dieser Arbeit genutzt werden soll. Hier werde ich zeigen, dass die physikalische Begriffsbildung durch einen Äquilibrationsprozess beschrieben werden kann, infolge dessen subjektive Theorien aufgebaut oder verändert werden.

Ein weiterer Theorieaspekt beschreibt den Prozess der Handlungsregulation an den Stationen. Unter der Verwendung eines integrativen Modells wird hier der Prozess des informellen und selbstgesteuerten Lernens genauer gefasst. Beide Begriffe werden im Rahmen dieser Arbeit bedeutungsgleich verwendet. Sie beschreiben einen weitgehend selbstbestimmten und selbstregulierten Lernprozess, der frei von externer Steuerung oder äußeren Bewertungsmaßstäben ist.

Einschränkend muss hier in Bezug auf die beiden Theorieteile eingewandt werden, dass beide Theorien spekulativen Charakter haben, folglich in weiten Teilen nicht empirisch überprüft worden sind. Diese Tatsache ist vor allem dadurch begründet, dass beide Theorien in weiten Teilen derartig abstrakt sind, dass sie sich einer exakten Operationalisierung entziehen.

1.3.2.2 Die Theorie Martin Wagenscheins als pädagogische Grundlage des interaktiven Lernens

Was macht einen Pädagogen zum Reformpädagogen und seine Pädagogik zur Reformpädagogik? Wählt man die Nennung des Autors und seiner Pädagogik in gängigen Lehrbüchern als Kriterium, so muss man Wagenschein unterstellen, er sei kein Reformpädagoge (vgl. Scheibe 1994; Oelkers 2005). Trotz dieser Nichtnennung Wagenscheins, die an seiner sehr speziellen Fragestellung, der vorrangigen Konzentration auf die Didaktik der Naturwissenschaften, besonders der Physik, liegen mag, ist Wagenschein aus meiner Sicht ein Reformpädagoge. Dies sei am Beispiel seiner Arbeit an der Odenwaldschule verdeutlicht: Bei dieser Schule handelte es sich um ein „Landerziehungsheim“, das sich bewusst von den üblichen Schulen der Weimarer Zeit abgrenzte. Zunächst liegen, dies impliziert der Name, diese Schulen auf dem Land. Sie verbindet eine bestimmte kultur- und großstadtkritische Attitüde. Jugend wurde hier als eine Zeit verstanden, die ganz dem Jugendlichen und seinen Peers gehören sollte. Man wandte sich gegen die übliche Stofffülle und setzte an ihre Stelle das gründliche Erarbeiten weniger Inhalte. Die Lehrerinnen und Lehrer dieser Schulen verstanden sich explizit nicht als „Stoffvermittler“ sondern als Erzieher. Sie wollten Ansprechpartner für die Schülerinnen und Schüler sein, mit ihnen zusammen leben (vgl. Scheibe 1999, S. 111ff.). Wagenschein berichtet begeistert von dieser Zeit. „Dabei glaube ich ernstlich nicht, daß ich von heute aus auf diese zweite Jugendzeit in verklärtem Lichte sehe“ (Wagenschein 1989, S. 33; Hervorhebung nicht im Original).

Im Folgenden soll die Theorie Wagenscheins für das interaktive Lernen an Stationen der MINIPHÄNOMENTA fruchtbar gemacht werden. Um die Übersichtlichkeit zu steigern, geschieht dies anhand einzelner Stichpunkte:

1. Die Stofffülle wird, so Wagenschein, nur durch die beispielhafte Auswahl einzelner Gegenstände handhabbar (exemplarisches Prinzip). Genau dies geschieht beim interaktiven Lernen an den Stationen. In Bezug auf die MINIPHÄNOMENTA werden beispielhaft 53 unterschiedliche Phänomene durch Stationen erfahrbar gemacht. Das Versuchsfeld ist kein „programmierter Lehrgang“ in dem das Feld der Naturwissenschaften systematisch, wohl möglich nach der Systematik der Wissenschaftsdisziplin, Stufe für Stufe entfaltet wird. Vielmehr sollen sich die Schülerinnen und Schüler selbstständig diejenigen Versuche herausgreifen, für die sie sich besonders interessieren.

2. Die Auswahl der Stationen durch die Schülerinnen und Schüler erfolgt selbstgesteuert. Die Station muss, so eine Erkenntnis der Lernpsychologie, die richtige Mischung aus Bekanntem und Fremdem enthalten, um für die Schülerinnen und Schüler einen entsprechenden Aufforderungscharakter zu gewinnen. Durch sie soll ein spannendes und komplexes Naturphänomen dargestellt werden, das die Kinder interessiert. Dieser Aufforderungscharakter entspricht der von Wagenschein erwähnten fesselnden Eingangsfrage. Sie wird den Schülerinnen und Schülern im Falle interaktiver Stationen durch das Exponat gestellt, beziehungsweise die Schülerinnen und Schüler formulieren sie sich selbst, aufgrund der Beobachtung und der Handhabung des Exponates. So bleiben die Individualität und das subjektive Interesse im Lernvorgang erhalten. Beide sind die Ausgangspunkte für den Lernprozess.
3. Die Schülerinnen und Schüler werden im Forschungsprozess durch den Aufbau des Experimentes und eventuell, im Sinne des sozialen Lernens, durch die Klassenkameraden angeleitet. Keinesfalls werden systematische Erklärungen durch Tafeln oder Ähnliches gegeben. So entwickeln die Schülerinnen und Schüler eigene, subjektiv tragfähige Vorstellungen über das Phänomen und seine Erklärung. Die dabei auftretenden Probleme werden, nicht wie im Unterricht sonst üblich, aus dem Weg geräumt. Vielmehr sind sie Gegenstand der Motivation, weiter zu forschen und mehr zu entdecken.
4. Am Anfang steht das Phänomen. Die Schülerinnen und Schüler „be - greifen“ es. Sie probieren aus, stellen um, entwickeln eigene Anschlussexperimente, verknüpfen die Stationen selbstständig untereinander usw. In diesem Prozess können sie sich selbst Erklärungen erarbeiten. Diese sind nicht zwangsläufig erforderlich. Abhängig vom Vorwissen, dem Interesse und der Motivation kann der Umgang mit dem Experiment ohne Theorie oder Modellvorstellung als ausreichend empfunden werden.
5. Bei den interaktiven Stationen wird der Bezug zum Alltag der Schülerinnen und Schüler auf mehreren Ebenen besonders deutlich hervorgehoben. Zunächst handelt es sich zum größten Teil um Alltagsphänomene, Wagenschein nennt sie in Abgrenzung zu Laborphänomenen, Naturphänomene. Weiterhin formulieren die Schülerinnen und Schüler die Beobachtungen und die Erklärungen in ihrer eigenen Sprache, ohne die Verwendung naturwissenschaftlicher Fachtermini. Sie arbeiten meist in Gruppen, formal gleichberechtigt als Forscherteam. Es wird gemeinsam experimentiert und diskutiert. Diese Situation des Lernens ist weit näher an der natürlichen Interaktion der Kinder in der Gruppe der Gleichaltrigen als die des Unterrichtes.

6. Im Zentrum steht immer das „Be - greifbare“, der konkrete Umgang mit der Station, das heißt das aktive Handeln. Dieser Umgang mit den Dingen wird auch von Wagenschein als besonders wichtig herausgearbeitet. Nur so können die Kinder mit späteren Abstraktionen umgehen. Die Begriffe, Variablen und Konstanten sind dann nicht leer, sondern können mit konkreten Erfahrungen verbunden werden. Um dies zu ermöglichen, sind die Stationen so groß, dass Kinder mit ihnen gut umgehen können, stabil genug, um auch robusteren Experimenten stand zu halten und in sich transparent gehalten. So ist das Experiment keine „entfremdete Apparatur“ sondern begreifbar.
7. Nach Wagenschein sollte das pädagogische Gespräch sokratisch orientiert sein. In der Diskussion mit den Schülerinnen und Schülern soll der Hintergrund eines Experimentes sukzessive argumentativ und experimentell ergründet werden. Die MINIPHÄNOMENTA lehnt sich mit ihren Stationen an diese Idee an: Im Falle der interaktiven Station wurden die Lehrerin oder der Lehrer mit dem Exponat ausgetauscht. Die Schülerinnen und Schüler können ihre eigenen Ideen am Experiment überprüfen und in der Gruppe der Gleichaltrigen diskutieren.
8. In den Stationen spiegelt sich nicht nur das exemplarische Prinzip, sondern auch Wagenscheins Gesamtkonzeption der Inselbildung. Die Stationen stellen, wie bereits erwähnt, die Beispiele dar, anhand derer vertiefend gelernt wird. Zusätzlich hierzu sollte es nach der Ansicht Wagenscheins „luftige Brücken“, das heißt Orientierungen und Informationen geben. Diese werden im Rahmen von ASIP beiläufig generiert oder gegeben. Die Schülerinnen und Schüler verknüpfen die Experimente selbstständig, Eltern liefern Informationen, Inhalte werden im Unterricht aufgegriffen usw.
9. Aufbauend auf die Anthropologie¹² Wagenscheins nutzen die Stationen die natürliche Lust am Denken und persönlichem Wachstum der Kinder. Sie setzen beim Interesse des Kindes an und sind in der Lage, die in ihm angelegte Motivation, zu denken, zu forschen, zu beobachten, zu experimentieren, kurz seine Spontaneität, zu wecken.

¹² Wagenscheins Bild des Kindes entspricht dem der Reformpädagogik (vgl. Scheibe, S. 51ff.).

10. Die interaktiven Experimentierstationen fördern die allgemeine Bildung im Sinne Wagenscheins:

Formatio:

Wagenschein übersetzt den Begriff mit Findigkeit. „[...] wir brauchen Menschen, denen zu neuen Aufgaben etwas Klärendes einfällt [...]“ (ebd. 1992, S. 77). Durch das tätige Experimentieren wird diese Fähigkeit besonders gefördert: Wie funktioniert die Station? Wie lassen sich Probleme lösen? Was könnte die Erklärung für das Phänomen sein? All diese Fragen werden selbstgesteuert, d. h. intrinsisch motiviert bearbeitet.

Enracinement:

hierunter versteht Wagenschein Eingewurzeltheit, das Vertrauen auf die eigene Beobachtung. Bei den ASIP - Stationen liegen zunächst keine anderen Informationen vor, als die die aus dem Experiment gewonnen werden können. Die Stationen schulen also die Fähigkeit zu beobachten, und Schlüsse aus diesen Beobachtungen abzuleiten.

kritisches Vermögen:

Es sichert die Korrektheit von Ergebnissen. Die Fähigkeit, sich im Prozess des Experimentierens selbst zu hinterfragen, Lösungen zu diskutieren, zu verwerfen und zu erneuern, wird ebenfalls durch das Arbeiten an den Stationen gefördert, da Erklärungen immer im Zusammenhang mit Beobachtungen stehen müssen. Ist dies nicht der Fall, sind die Schülerinnen und Schüler durch das Experiment aufgefordert, eigene Denkmuster zu verändern.

1.3.2.3 Die Theorie der kognitiven Entwicklung von Piaget unter der besonderen Berücksichtigung der Kritik und Weiterentwicklungen

Das Lernen der Schülerinnen und Schüler wurde im Abschnitt 1.2.1.2 bereits als interaktives, informelles Lernen gekennzeichnet. In den folgenden beiden Abschnitten soll dieses Verständnis nun weiter vertieft werden. Ich beginne mit der Beschreibung der kognitiven Prozesse. Hierbei orientiere ich mich an der Tradition Piagets. Im Folgenden beschreibe ich die Ansätze des selbstgesteuerten Lernens, die sich integrativ mit den Prozessen der Handlungsregulation an den Stationen befassen.

Piagets Theorie der geistigen Entwicklung ist enorm umfangreich. Sie lässt sich aus der Sicht des Autors in drei wesentliche Aspekte gliedern:

1. Das grundsätzliche Modell der geistigen Entwicklung: Das Auqilibrationskonzept
2. Das Stufenmodell der geistigen Entwicklung: Die vier Stadien der kognitiven Entwicklung
3. Spezielle Elemente, zum Beispiel: Piagets Spieltheorie, Entwicklung physikalischer Konzepte

Diese Bereiche oder Teile von ihnen an dieser Stelle darzustellen, wäre unergiebig. Ich verweise auf Sauer (2005, S. 26 ff.) und Holst (2005, S. 13 ff.), die vor allem das Auqilibrationskonzept dezidiert dargestellt haben, sowie auf gängige Lehrbücher der Entwicklungspsychologie (vgl. hier vor allem Oerter, Montada 1995, S. 518 ff.; Buggle 2001), und nicht zuletzt auf die arbeiten von Piaget selbst.

Piaget gilt als der Begründer der kognitiven Entwicklungspsychologie. Seine Arbeiten zogen eine Vielzahl weiterer Forschungsaktivitäten nach sich. In diesen wurden seine Konzeptionen erweitert, korrigiert und bestätigt. Im Folgenden soll die kritische Auseinandersetzung mit Piaget anhand einiger Beispiele geschildert werden. Ziel ist es zum einen, die Arbeiten Piagets zu würdigen, zum anderen den Blickwinkel für neuere Forschungsergebnisse zum Thema der kognitiven Entwicklung zu erweitern.

Dabei muss betont werden, dass es sich bei der Kritik an Piaget um kein einfaches Unterfangen handelt. Er veröffentlichte insgesamt weit über 400 Texte. Jedoch nicht nur der Umfang des Werkes macht die Kritik schwer, sondern auch dessen Vielfalt. Stets geht es zwar um menschliche Erkenntnis und die Frage, wie diese sich im Laufe der Kindheit und Jugend entwickelt: Piaget befasste sich mit den unterschiedlichsten Aspekten, von einfachsten Abläufen bis hin zu hoch komplizierten Denkprozessen (vgl. Buggle 2001, S. 16 ff.). Zur besseren Übersicht stelle ich zunächst die Kritik an Piaget in Form einer Tabelle dar, um im Folgenden auf einige Punkte genauer einzugehen:

Nummer	Autor, Erscheinungsjahr	Kritischer Aspekt
1.	Buggle 2001 ¹³ , S. 110 f.	Forschungsmethodische Einwände: fehlende Stichprobenbeschreibungen, fehlende statistische Beschreibungen, die Methode des klinischen Interviews, geringer Stichprobenumfang.
2.	Buggle 2001, S. 106 f.	Probleme mit den Begrifflichkeiten: Sie sind zum Teil unpräzise: Dies führt oft zur Vagheit und Unbestimmtheit der Termini.
3.	Buggle 2001, S. 106 f.	Das Konzept ist wenig erklärend, sondern vorrangig deskriptiv.
4.	Oerter, Montada 1998, S. 559 ff., 662 ff.	Das Stadienmodell als Ganzes ist wenig bestätigt.
5.	Oerter, Montada 1998, S. 559 ff.	Differentielle Entwicklungsverläufe finden sowohl in der Deskription als auch in der Erklärung zu wenig Berücksichtigung.
6.	Keller 1998, S. 147 ff.; Oerter, Montada 1998, S. 201 ff., 662 ff.	Kleinkinder scheinen sehr viel kompetenter zu sein als Piaget angenommen hat: So verfügen sie schon sehr viel früher über das Konzept der Objektpermanenz, haben ein Verständnis für Kausalität usw.
7.	Keller 1998, S. 147 ff., 317 ff.; Oerter, Montada 1998, S. 559 ff.	Es werden Zweifel an der Existenz des präoperatorischen Denkens geäußert. Als Beispiele sollen hier genannt werden: <ul style="list-style-type: none"> • Zweifel am Konzept des kindlichen Egozentrismus • Kausales Denken im Vorschulalter • Lineare Ordnungen und transitive Schlüsse
8.	Oerter, Montada 1998, S. 662 ff.	Entwicklung ist nicht universell sondern bereichsspezifisch und erfordert bereichsspezifisches Wissen.

Tabelle 1.3.2.3.1: Die Theorie der kognitiven Entwicklung von Piaget;
- ausgewählte Kritikpunkte an den Konzeptionen Piagets
Der Verfasser

Im Folgenden werde ich auf die Punkte fünf, sieben und acht aufgrund ihrer hohen Relevanz für den Verlauf der Arbeit genauer eingehen:

¹³ In dem der Tabelle folgenden Textteil werde ich keine weiteren Literaturangaben machen, es sei den die verwendete Literatur geht über die in der Tabelle Genannte hinaus.

Ad 5: Das Modell Piagets ist der allgemeinen Entwicklungspsychologie zuzuordnen. Sein Entwicklungsbegriff kann als traditionell bezeichnet werden. Es gibt eine Veränderungsreihe, die aus mehreren Schritten besteht und auf einen Endzustand hin verläuft. Dieser ist in Relation zum Ausgangspunkt höherwertiger. Die Folge dieser Schritte ist unumkehrbar. Veränderungen, die mit den Schritten eintreten, lassen sich als qualitative, strukturelle Transformation beschreiben. Sie sind mit dem Lebensalter korreliert und universell. In diesem Zusammenhang möchte ich mich lediglich auf den letzten Punkt, die „Universalitätsannahme“ konzentrieren. Diese Sichtweise ist problematisch, da sie bestimmte Fragestellungen von vornherein ausklammert werden. Sie betreffen vor allem Aspekte der differentiellen Entwicklung. Anders formuliert: Gibt es interindividuelle Unterschiede in der kognitiven Entwicklung und wenn es sie gibt, wie lassen sie sich erklären? Vielfältige empirische Ergebnisse weisen darauf hin, dass die kognitive Entwicklung interindividuell tatsächlich höchst unterschiedlich verläuft.

Ad 7: Kinder im Vorschulalter sind nach Piaget nicht in der Lage, die Perspektive anderer Personen zu berücksichtigen. Zu dieser Erkenntnis kam er durch den „Drei-Berge-Versuch“. Dabei schaut ein Kind auf eine Berglandschaft. Gegenüber vom Kind sitzt eine Puppe. Nachdem das Kind seine eigene Perspektive beschrieben hat, soll es versuchen zu erklären, wie die Puppe die Berge sieht. Hierzu sind Kinder auf dem präoperationalen Niveau nicht fähig. Die Egozentrismusannahme wurde im Wesentlichen durch zwei Versuche widerlegt: Zum einen durch die „Schildkrötenaufgabe“. Bei ihr sitzen Kind und Versuchsleiter einander gegenüber. Zwischen ihnen liegt ein Bild, das die Seitenansicht einer Schildkröte zeigt. Der Versuchsleiter erklärt dem Kind mehrmals, wie er die Schildkröte sieht, wenn sie unterschiedlich gedreht wird. Nun wird eben diese erneut unterschiedlich gedreht und das Kind muss beantworten, wie es selber und wie der Versuchsleiter die Schildkröte sieht. Ab einem Alter von etwa vier bis fünf Jahren können Kinder beide Fragen korrekt beantworten. Sie verstehen also schon sehr viel früher als Piaget annahm, dass ein Objekt aus unterschiedlichen Perspektiven unterschiedlich aussieht. Weiterhin wurde eine Untersuchung durchgeführt, in der zwischen Kind und Versuchsleiter eine Trennwand steht. Stellt man eine Figur auf die Seite des Kindes so antworten 2 bis 3 Jährige korrekt, dass nur sie und nicht der Versuchsleiter diese Figur sehen können. Es bleibt also festzustellen, dass Perspektivenübernahme schon sehr viel früher möglich ist. Piaget verkannte dies, da sein „Drei-Berge-Versuch“ sehr komplex ist und eine Vielzahl von Anforderungen stellt, die über das Perspektivenproblem hinausgehen. Die Beispiele machen deutlich, dass die theoretischen Ergebnisse wesentlich von der eingesetzten Methode abhängen. Piaget verwandte oft komplexe Methoden und unterschätzte so die kognitiven Fähigkeiten des Kindes.

Ad 8: In neueren Theorien zur kognitiven Entwicklung wird angenommen, dass Veränderungen inhaltsspezifisch und bereichsgebunden sind; globale Veränderung z.B. in Form von Denkstadien werden nicht mehr postuliert. Theoretisch lässt sich diese Position gut in das „Experten-Novizen-Paradigma“ einordnen. Kinder sind damit im Vergleich zu Erwachsenen universelle Novizen, die in allen Bereichen über weniger Wissen verfügen. Entwicklung ist damit der Erwerb von spezifischem Inhaltswissen. Kinder müssen in allen Bereichen des Lebens diese „Expertise“ durchlaufen. Es bleibt die Frage, wie diese „Expertise“ kognitionstheoretisch aussieht. Oder anders formuliert: Kommt es durch den Erwerb bereichsspezifischen Wissens lediglich zu einer Anreicherung oder zu einer grundlegenden Umstrukturierung des Wissens. Die kindlichen Denkfehler weisen darauf hin, dass es sich eher um eine Umstrukturierung handelt, denn die Fehler sind erstaunlich Resistent gegen Belehrungen. Die Erklärung, die ein Kind gibt und die „richtige“ Erklärung unterscheiden sich damit nicht einfach durch mangelndes Wissen: Beide kommen aus „unterschiedlichen Welten“. Sie verwenden z.B. völlig unterschiedliche Begriffssysteme. Man denke hier nur an animistische Deutungen kleinerer Kinder. Man kann den Wissenserwerb des Kindes als einen Prozess des Theoriewandels verstehen.

Wie soll nun mit der weit reichenden Kritik an Piaget im Rahmen dieser Arbeit umgegangen werden? Sind die Weiterentwicklungen derartig bedeutend, dass das gesamte Konzept Piagets in Frage gestellt werden muss? Erinnern wir uns an dieser Stelle an die eingangs genannten Themenkategorien der Theoriebildung Piagets und wenden diese auf die dargestellten Kritikpunkte an:

Die meisten Autoren beziehen sich in ihrer Kritik an den Arbeiten Piagets auf das Stufenmodell. Hier ließe sich die Liste auch noch deutlich erweitern; so erreichen längst nicht alle Abiturienten und Abiturientinnen das Niveau des formalen Denkens (vgl. Fiesser 1990, S.345 ff.). Auch die Angaben Piagets zum Alter des Überganges von einer Stufe zur anderen konnten vielfach nicht bestätigt werden (vgl. Buggle 2001, S.102 ff.).

Das Aquilibrationsmodell ist von der heutigen psychologischen Forschung als umfassendes Modell der geistigen Entwicklung ebenfalls heftig kritisiert worden. Viele Aspekte sprechen dafür, dass Wissen und geistige Fähigkeiten sich bereichsspezifisch entwickeln (siehe Punkt acht in diesem Abschnitt). Wie entwickelt sich aber das Wissen im Bereich der Physik? Die Annahme, dass physikalisches Wissen lediglich in der Form einer Anreicherung von Kenntnissen entsteht, konnte weitgehend widerlegt werden (vgl. Carey 1991, nach Oerter, Montada 1995, S. 634 ff.). Vielmehr kommt es im Laufe der Entwicklung der Kinder zu weitreichende Transformationen des Begriffssystems. Diese Veränderungen können hervorragend mit den Begriffen der Akkommodation und der Assimilation beschrieben werden. Wobei an dieser Stelle zu betonen ist, dass sich Abläufe der Begriffsveränderung in vielen kleinen Schritten vollziehen und weniger als plötzlicher Umbruch.

Auch die Betrachtungen der folgenden physikdidaktischen Untersuchungen legen das Äquilibrationskonzept als theoretisches Modell für den Erwerb physikalischer Kenntnisse und Fähigkeiten in einer bereichsspezifischen Form nahe: Fiesser 1990; Sauer 2005; Holst 2005; Fiesser, Philippi, Schließmann 2005.

Dieser Ansatz wird nun um zwei weitere Konzeptionen erweitert, das „Experten-Novizen-Paradigma“ und den Ansatz der „subjektiven Theorien“ (vgl. Laucken 1974, Groeben, Wahl, Scheele 1988).

Das „Experten-Novizen-Paradigma“ ist in diesem Abschnitt (vgl. ad 8) bereits dargestellt worden, muss daher nicht nochmals erörtert werden.

Anders beim der Ansatz der „subjektiven Theorien“: Unter einer subjektiven Theorie verstehen Groeben, Wahl und Scheele (1988) ein kognitives System mit einer Argumentationsstruktur. Parallel zu einer wissenschaftlichen Theorie erfüllt die subjektive Theorie die Funktion der Erklärung, Prognose, Technologie usw. Kinder bauen sich damit zu unterschiedlichen physikalischen Bereichen (beziehungsweise Phänomenen) subjektive Theorien auf. Genauso wie wissenschaftliche Theorien auch, sind sie damit in der Lage Erscheinungen zu beschreiben, zu erklären und Handlungen zu planen, um die Prozesse in eine bestimmte, beabsichtigte Richtung zu steuern. Dabei sind diese subjektiven Theorien ihrer Natur nach nicht so exakt wie die der wissenschaftlichen Physik, aber beide weisen Gemeinsamkeiten in Bezug auf ihre Inhalte, ihre Struktur und ihre Funktion auf. Die Erweiterung einer subjektiven Theorie bezeichnet man in der Sprache Piagets als Assimilation, den Theoriewandel als Akkommodation. Damit erweitert dieser Ansatz das Experten-Novizen-Paradigma, indem er etwas über die formale Beschaffenheit der Gedanken der Kinder aussagt. Dieser Gedanke passt gut zu der Annahme Piagets, dass es im Denken von Menschen zwar qualitative, aber keine kategorialen Unterschiede gibt. Der einfachste und der komplexeste Gedanke hat nach Piagets Ansicht den gleichen Ursprung (vgl. Buggle 2001, S. 16 ff.)

Ich fasse das gesagte zusammen:

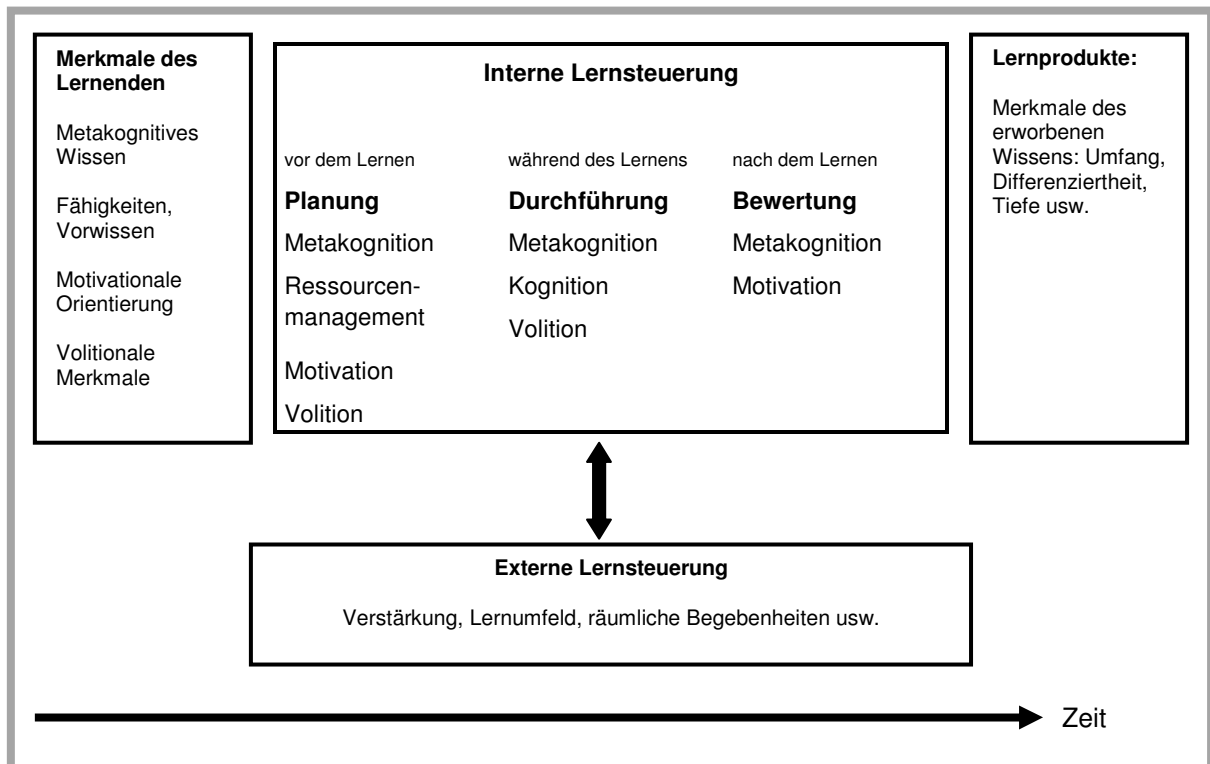
1. Physikalisches Wissen entwickelt sich bereichsspezifisch.
2. Kinder sind universelle Novizen, die in allen Bereichen und damit auch in der Physik eine Expertise durchlaufen müssen.
3. Physikalisches Wissen entwickelt sich in Bezug auf die Geschwindigkeit und den Umfang individuell unterschiedlich.
4. Die Entwicklung physikalischen Wissens lässt sich durch das Äquilibrationmodell beschreiben.
5. Das Individuum entwickelt zu physikalischen Sachverhalten subjektive Theorien, die in einigen Bereichen Analogie zu wissenschaftlichen Theorien aufweisen. Sie sind zugleich Möglichkeit als auch Grenze der Erkenntnis. Das Ergebnis zahlreicher Akkommodation ist der Wechsel der subjektiven Theorie.

1.3.2.4 Interaktive Stationen vor dem Hintergrund aktueller Theorien des selbstgesteuerten Lernens

Selbstgesteuertes Lernen ist eine „[...] komplexe Gesamthandlung [...] bei der der Handelnde die wesentlichen Entscheidungen, ob, was, wann, wie und worauf hin gelernt wird gravierend und folgenreich beeinflussen kann“ (Friedrich, Mandl 1997, S. 238).

Bislang wurde der Lernprozess an den Stationen als interaktiv und informell beschrieben. Der erstgenannte Begriff betont dabei den dialogischen Charakter zwischen Lernendem und Versuch, der zweite die Abwesenheit externer Bewertungskriterien und Institutionalisierung des Lernprozesses. Hier wird nun das Lernen an den Stationen durch einen weiteren Begriff charakterisiert. Die Selbstbestimmtheit des Lernens betont dabei die eigenständige Themenfindung, Initiierung, Organisation und Bewertung des Lernprozesses. Die drei Begriffe: Interaktives Lernen, informelles Lernen und selbstgesteuertes Lernen sind damit nicht disjunkt zu trennen. Vielmehr gibt es deutliche Überschneidungen der Begriffsräume. Dennoch werden durch jeden der drei genannten Begriffe unterschiedliche Schwerpunkte gesetzt.

Der Prozess des selbstgesteuerten Lernens an den Stationen kann durch das folgende integrative Modell beschrieben werden:



Grafik 1.3.2.4.1: Aktuelle Theorien des selbstgesteuerten Lernens; - Rahmenmodell des selbstgesteuerten Lernens Modifiziert nach Schiefele, Pekrun 1996, S. 271

Grundlage des Lernens sind die Merkmale des Lernenden, die interindividuell variieren. Das Modell nennt, ohne Anspruch auf Vollständigkeit, vier Faktoren:

1. Metakognitives Wissen: Der Begriff der Metakognition bezieht sich in diesem Zusammenhang auf die Fähigkeit zur Kontrolle kognitiver Vorgänge. Eben diese umfasst drei Komponenten: Fähigkeiten zur Planung, Überwachung und Kontrolle kognitiver Prozesse.
2. Fähigkeiten, Vorwissen: Das Vorwissen kann im Zusammenhang mit dem Wissensbegriff dieser Arbeit als der Umfang und die Beschaffenheit des semantischen Begriffsnetzes¹⁴ verstanden werden.
3. Motivationale Orientierung: Motivation wird im Zusammenhang dieser Arbeit nicht als habituelle Komponente, entsprechend dem trait-Ansatz aufgefasst, sondern vielmehr als aktuell generiertes Phänomen, das Prozesse initiiert, moderiert und beendet.
4. Volitionale Merkmale: Der Begriff der Volition bezieht sich nach Edelman (2000, S. 193 ff.) auf den Begriff des Handlungskonzeptes. Wie konkret durch die Person gehandelt wird ist kein vorrangig motivationaler Prozess, sondern ein volitionaler: Auf der Basis von kognitiven, affektiven und motivationalen Prozessen wird ein Handlungskonzept entwickelt.

Der eigentliche Lernvorgang ist das Ergebnis von zwei Prozessen, der internen und der externen Lernsteuerung. Die interne Lernsteuerung lässt sich in drei Aspekte gliedern: Die Planung der Handlung, ihre Durchführung und ihre Bewertung. Zur besseren Übersichtlichkeit sind in der obigen Darstellung Rückkopplungen nicht dargestellt. Es sei an dieser Stelle jedoch unterstrichen, dass selbstverständlich bei der Bewertung der Handlung ein Vergleich mit dem ursprünglichen Ziel stattfindet. Dabei sind folgende Fälle zu unterscheiden:

- Ziel und Ergebnis der Handlung stimmen überein: Die Handlung ist abgeschlossen.
- Ziel und Ergebnis der Handlung stimmen nicht oder nur teilweise überein: In diesem Fall sind mehrere Möglichkeiten denkbar: Modifizierung des Ziels, ganze oder teilweise Wiederholung der Handlung mit gleichem oder modifiziertem Handlungskonzept, Abwertung des Zieles usw.

Es sollte daher stets bedacht werden, dass Handlungen sich in ihrer Vollständigkeit nur in komplexen sequentiell und hierarchisch organisierten Regelkreisen darstellen lassen. Das obige Modell ist damit eine starke Vereinfachung der Sachlage.

¹⁴ Zum Wissensbegriff dieser Arbeit vergleiche Abschnitt 1.4.7

Die Planung der Handlung erfordert kognitive und metakognitive Kompetenzen. Sie wird durch Motivationsprozesse initiiert und es wird ein Handlungskonzept entworfen. In der Durchführung der Handlung treten motivationale Prozesse in den Hintergrund und volitionale in den Vordergrund. Bei der Bewertung der Handlung sind vorrangig metakognitive und motivationale Prozesse von Bedeutung. Die erzielten Ergebnisse werden rückblickend evaluiert.

Diese interne Lernsteuerung wird stets von der externen Lernsteuerung beeinflusst (und umgekehrt). So ist zum Beispiel die Umgebung, in der die MINIPHÄNOMENTA aufgestellt wird, von zentraler Bedeutung. Hiermit ist zum einen die räumlich-physikalische Umgebung gemeint, zum anderen auch die formalen Bedingungen des Settings Schule: Zum Beispiel die Nutzung der Stationen in den Pausen. Hinzu kommen die sozialen Bedingungen: Arbeitet das Kind zum Beispiel alleine an der Station oder zusammen mit anderen Kindern?

Das Ergebnis der Merkmale des Lernenden, der internen Lernsteuerung und der externen Lernsteuerung in ihrer vielfältigen Interaktion ist das Lernprodukt, in seiner individuell völlig unterschiedlichen Beschaffenheit.

Der besondere Reiz dieses Modells liegt in der Vielschichtigkeit seiner Annahmen. Mit Hilfe des Modells gelingt es nicht nur, die kognitiven Aspekte des Lernens an Stationen zu beschreiben, sondern vielmehr den gesamten Prozess der Handlung an der Station in seiner erörterten Vielschichtigkeit.

Durch die Theorie des selbstgesteuerten Lernens lassen sich die komplexen Prozesse der Interaktion der Schülerinnen und Schüler mit der Station integrativ erfassen.

1.3.3 Ergebnisse der bisherigen Forschung zum Projekt

Im Folgenden soll nun der Stand der Forschung zum Projekt dargestellt werden: Er basiert auf zwei Dissertationen und einer ersten Staatsarbeit für das Lehramt an Grund- und Hauptschulen. Aus den Arbeiten werde ich im Folgenden beispielhaft wichtige Ergebnisse zusammenstellen:

Ich beginne mit der Arbeit von Holst (2005): Der Autor befasst sich in seiner Dissertation mit der Überprüfung der Bildungswirksamkeit der ASIP-Stationen im Primar- und Orientierungsstufenbereich. Er verwendet zwei Forschungsmethoden:

1. Verhaltensbeobachtung an den Stationen,
2. Wissens- und Vorstellungsdiagnostik mit dem Computerprogramm CCMaP.

Ich beginne mit der Schilderung der Ergebnisse der Verhaltensbeobachtung: Die durchschnittliche Verweildauer der Kinder an den Stationen beträgt 1,3 Minuten. Dieser Wert mag zunächst relativ niedrig erscheinen. Hier muss jedoch der Kontext der Beobachtung berücksichtigt werden. Die Kinder hatten lediglich in den Pausenzeiten (häufig nur in den großen Pausen) Zugang zu den Experimenten.

Vergleicht man nun diesen Wert mit der durchschnittlichen Verweildauer von Schülerinnen und Schülern im Science-Center Phänomenta in Flensburg (unter einer Minute), so wird deutlich, dass es sich um einen hohen Wert handelt.

Die Schülerinnen und Schüler arbeiteten an den Experimentierstationen nur selten allein. Etwa zwei Drittel der Kinder näherten sich den Experimenten in Kleingruppen von drei bis fünf Mitgliedern. Sie diskutierten den Umgang mit den Experimenten intensiv. Oft stellten sich nach kurzer Zeit für das jeweilige Exponat Experten heraus, deren Rat bei Bedarf von den Mitschülerinnen und Mitschülern eingeholt wurde. Die Kinder erarbeiten in den Gruppen, oft im Prozess einer lebhaften Diskussion, eigene Erklärungen für die jeweiligen Phänomene.

Es stellte sich heraus, dass die Befürchtung, die Stationen könnten zu Spielobjekten umfunktioniert werden, unbegründet war. 70% der Schülerinnen und Schüler experimentierten gezielt. Sie manipulierten - ihrem kognitiven Entwicklungsstand entsprechend - Variablen, untersuchten Zusammenhänge und überprüften eigene Erklärungen.

Die positiven Ergebnisse der Verhaltensbeobachtung spiegelten sich auch in den Concept-Maps wieder: So lag der durchschnittliche Fehlerquotient, also der Quotient der Summe der richtigen und der falschen Propositionen lediglich bei 0,22. Die fachliche Bewertung wurde auf einer Skala von 1 bis 3 (3 = beste fachliche Bewertung, 1= schlechteste fachliche Bewertung) gemessen. Der Mittelwert aller Exponate liegt bei 2,0. Beide Ergebnisse zeigen den großen Lernerfolg der Kinder im Umgang mit den Stationen. Sie machen sehr wenige inhaltliche Fehler und die Qualität der Aussagen ist hoch.

Sauer (2005) untersuchte in seiner Dissertation den Einfluss der Experimentierstationen auf das naturwissenschaftlich-technische Lernen im Primarbereich. Der Autor führte insgesamt fünf Teiluntersuchungen durch, aus denen einige Ergebnisse referiert werden sollen:

Ich beginne mit den Schülerinterviews: 58% der Schülerinnen und Schüler gelang es nach dem Experimentieren, einen Versuch ihrer Wahl korrekt zu beschreiben. 20% der Befragten konnten dies zumindest ansatzweise richtig tun. 91% der Schülerinnen und Schüler sprachen über die Versuche mit Mitschülerinnen und Mitschülern. Ein noch höherer Anteil erzählte zu Hause von den Versuchen.

Es ließ sich aber nicht nur ein Einfluss auf die Schülerinnen und Schüler nachweisen. Auch der Unterricht veränderte sich durch die MINIPHÄNOMENTA: Inspiriert durch die offenen Stationen probierten die Lehrerinnen und Lehrer neue Unterrichtsmethoden aus. So erzählte im Interview eine Lehrerin begeistert von einer Stunde, in der die Schülerinnen und Schüler das Lungenvolumen des Menschen experimentell bestimmen sollten. Zu diesem Zweck begann sie die Stunde mit der Frage, wie hoch das Lungenvolumen des Menschen sei. Sie forderte Schülerinnen und Schüler auf, diese Frage mit den mitgebrachten Materialien (Eimer, Schläuche usw.) zu klären. Es entwickelte sich in der Stunde eine rege Experimentiertätigkeit, die zum Teil außerhalb des Klassenraumes stattfand. Am Ende der Stunde war die Eingangsfrage beantwortet. Es hatten sich darüber hinaus weitere Fragestellungen ergeben, die in den folgenden Stunden geklärt werden sollten.

Auch die Elternbefragung zeigte die positiven Effekte von ASIP. Zu 75% berichteten die Kinder den Eltern von den Stationen. 86% der Kinder zeigen sich auch zu Hause von den Stationen begeistert. Die Eltern wurden durch die MINIPHÄNOMENTA angeregt sich mit ihren Kindern mehr über naturwissenschaftliche Fragestellungen im Alltag zu unterhalten. Drei Viertel aller Eltern können sich vorstellen, zusammen in einer Arbeitsgruppe Experimentierstationen für die Schule zu bauen.

Für den Zusammenhang dieser Untersuchung ist besonders die längsschnittliche Erhebung von Sauer interessant. Der Autor befragte die Kinder nach drei Monaten erneut, an welche Stationen sie sich noch erinnern können. Die Schülerinnen und Schüler erinnerten sich im Durchschnitt noch an 21 Stationen. Auch die Funktion und der Ablauf der Versuche war ihnen noch erstaunlich präsent. So lieferten 82% der Schülerinnen und Schüler richtige Ergebnisse bei der Darstellung der Versuchsverläufe mit Legebögen.

John (2004) untersuchte das Engagement der Eltern bei der MINIPHÄNOMENTA am Beispiel einer Grundschule im Rahmen der schriftlichen Hausarbeit zum ersten Staatsexamen für Grund- und Hauptschullehrer. Er kam dabei zu den folgenden qualitativ orientierten Ergebnissen:

In der MINIPHÄNOMENTA engagieren sich vorrangig Eltern, die an der Bildung ihrer Kinder Anteil nehmen. Sie zeigen kein überdurchschnittliches soziales oder politisches Engagement. Die Eltern haben im Durchschnitt auch kein erhöhtes Interesse an Naturwissenschaften oder Technik. Vielmehr interessieren sie sich für das konkrete Projekt. Sie sind zumeist berufstätig. Die aktive Teilnahme an der MINIPHÄNOMENTA bedeutet für sie folglich ein zeitliches Opfer. Durch die vielfältigen beruflichen Orientierungen der Eltern können sie die auftretenden Probleme allein lösen. Die Hilfe der Universität Flensburg beziehungsweise ihres Personals war nur selten erforderlich. Die MINIPHÄNOMENTA wirkt sich positiv auf das Schulklima und die Zusammenarbeit von Eltern und Lehrern aus.

Zusammenfassend kann daher gesagt werden, dass das Engagement der Eltern im Rahmen der MINIPHÄNOMENTA sehr positive Effekte hat. Die anfängliche Befürchtung, das Interesse der Kinder könnte durch das Engagement der Eltern geschwächt werden, entkräftete John. Vielmehr bildete sich durch die gemeinsame Arbeit eine Triade Eltern-Schule-Kinder, in der die Kinder ihr eigenes Interesse am Projekt beibehielten. Es wurde durch das Engagement der Eltern sogar noch verstärkt, zum Beispiel in dem die MINIPHÄNOMENTA zu einem Thema im privaten Umkreis der Familie wurde.

Bisherige überwiegend querschnittlich angelegte Evaluationen zu unterschiedlichen Bereichen bescheinigen dem Projekt große Erfolge, so dass anzunehmen ist, dass sich diese Ergebnisse auch im längsschnittlichen Design dieser Arbeit zeigen lassen.

1.4 Die Grundlagen der abhängigen Variablen der Untersuchung: Zentrale Definitionen und Theorien

1.4.1 Die abhängigen¹⁵ Variablen im Untersuchungskontext

An dieser Stelle sei nochmals das Ziel der Untersuchung in Erinnerung gerufen, das in der Erfassung der Langzeitwirksamkeit der MINIPHÄNOMENTA besteht. Folglich muss der Begriff der Langzeitwirksamkeit beziehungsweise Wirksamkeit als abhängige Variable dieser Untersuchung operationalisiert werden. Hierfür sind drei Bezugssysteme denkbar:

1. Der Lehrplan Physik der Sekundarstufe I: Physikalischer Unterricht war in Bezug auf ein bestimmtes Thema wirksam, wenn die Schülerinnen und Schüler bezugnehmend auf dieses Sach-, Selbst-, Methoden- und Sozialkompetenz entwickelt haben. Übertagen auf die hier zu verhandelnde Untersuchungsfrage würde dies bedeuten, dass die MINIPHÄNOMENTA wirksam ist, sofern die Schülerinnen und Schüler durch ihre Teilnahme am Projekt in den genannten vier Bereichen gefördert werden. Bezugnehmend auf den Abschnitt 1.2 in dieser Arbeit ist einschränkend zu bemerken, dass ASIP auf das Sozialverhalten der Schülerinnen und Schüler wenig Einfluss haben dürfte. Viele Untersuchungen belegen den Einfluss anderer - oft außerschulischer - Faktoren.
2. Das Konzept der Physikalischen Bildung als „Stärkung des Ich“: Physikalische Bildung kann als ein Teil der allgemeinen Bildung verstanden werden, die dem Individuum drei zentrale Emanzipationsmöglichkeiten bietet: Schulung allgemeiner Fertigkeiten, Entwicklung logischer Fähigkeiten und die Erhöhung der Selbstwirksamkeit. Damit ist die MINIPHÄNOMENTA wirksam, sofern die teilnehmenden Schülerinnen und Schüler durch den Umgang mit den Versuchen ihre physikalische Bildung vertiefen konnten.
3. Die Bausteine der Theorie des interaktiven Lernens: Es konnte gezeigt werden, dass es sich beim Umgang der Schülerinnen und Schüler mit den Stationen um einen hochkomplexen Lernvorgang handelt, der auf mehreren Ebenen beschrieben werden muss. Die MINIPHÄNOMENTA ist in Bezug auf die Handlungsregulation an den Stationen wirksam, wenn die Schülerinnen und Schüler durch das tätige Experimentieren elaboriertere Handlungskontrollstrategien entwickeln können.

Der beschriebene Hintergrund zeigt, dass das Merkmal der Wirksamkeit der MINIPHÄNOMENTA komplex und heterogen ist. Um zu belastbaren Ergebnissen zu kommen, muss der Begriff der Wirksamkeit durch mehrere Variablen aus unterschiedlichen Bereichen operationalisiert werden.

¹⁵ Der Begriff der abhängigen Variablen leitet sich hier aus dem quasiexperimentellen Design der Studie ab (vgl. Abschnitt 2.2 in dieser Arbeit).

Die beiden letztgenannten Aspekte (Physikalische Bildung und die Theorie des interaktiven Lernens) bringen in Bezug auf die vorzunehmende operationale Definition des Wirksamkeitsbegriffes ein zentrales Problem mit sich: Ihre Komplexität und Heterogenität. Sie sorgen zwar im theoretischen Bereich für Stimmigkeit, sind für eine exakte Operationalisierung nahezu ungeeignet. Beide Bereiche entziehen sich geradezu der empirischen Definition.

Insofern erscheint es sinnvoll sich bei der Operationalisierung des Wirksamkeitsbegriffes vorrangig auf den erstgenannten Aspekt, den Lehrplan, zu konzentrieren. Über diese forschungspraktischen Gründe hinaus hat dieses Vorgehen eine hohe face validity, bedenkt man den schulischen Kontext der MINIPHÄNOMENTA und dieser Untersuchung.

Der Bereich der **Sachkompetenz** wird in dieser Untersuchung durch die Variable [Wissen] operationalisiert. Erfasst wird dabei das Wissen zu einem in der Untersuchung durchgeführten Versuch. Die direkte Erhebung von Wissen, ähnlich einer Klassenarbeit oder einem Test, ist zum einen theoretisch in Bezug auf die Konzeption der MINIPHÄNOMENTA problematisch, zum anderen lässt sich mit Blick auf den Lehrplan des Heimat- und Sachunterrichtes der Primarstufe kein klarer Unterrichtsbezug herausarbeiten (vgl. Abschnitt 1.2.3 in dieser Arbeit). Insofern erhebe ich hier eine Mischform aus prozeduralem und deklarativem Wissen, indem die Schülerinnen und Schüler gebeten werden, auf der Basis einer Versuchsdurchführung und ihrer konkreten Beobachtungen, eine Erklärung für ein Phänomen zu finden.

In Bezug auf die **Methodenkompetenz** erhebe ich zwei Variablen: Zunächst wird die [formale Kompetenz] in einem Laborsetting an einem konkreten Versuch erhoben. Das [allgemeine Experimentierverhalten] hingegen wird in einer offenen Unterrichtssituation in Bezug auf einzelne Schülerinnen und Schüler im gesamten Klassenverband ermittelt. Beide Variablen orientieren sich dabei am Lehrplan des Faches Heimat- und Sachunterricht in der Primarstufe (vgl. Abschnitt 1.2.3 in dieser Arbeit). Zum Ende der dritten Klassenstufe sollen die Kinder in altersangemessener Art und Weise experimentieren können.

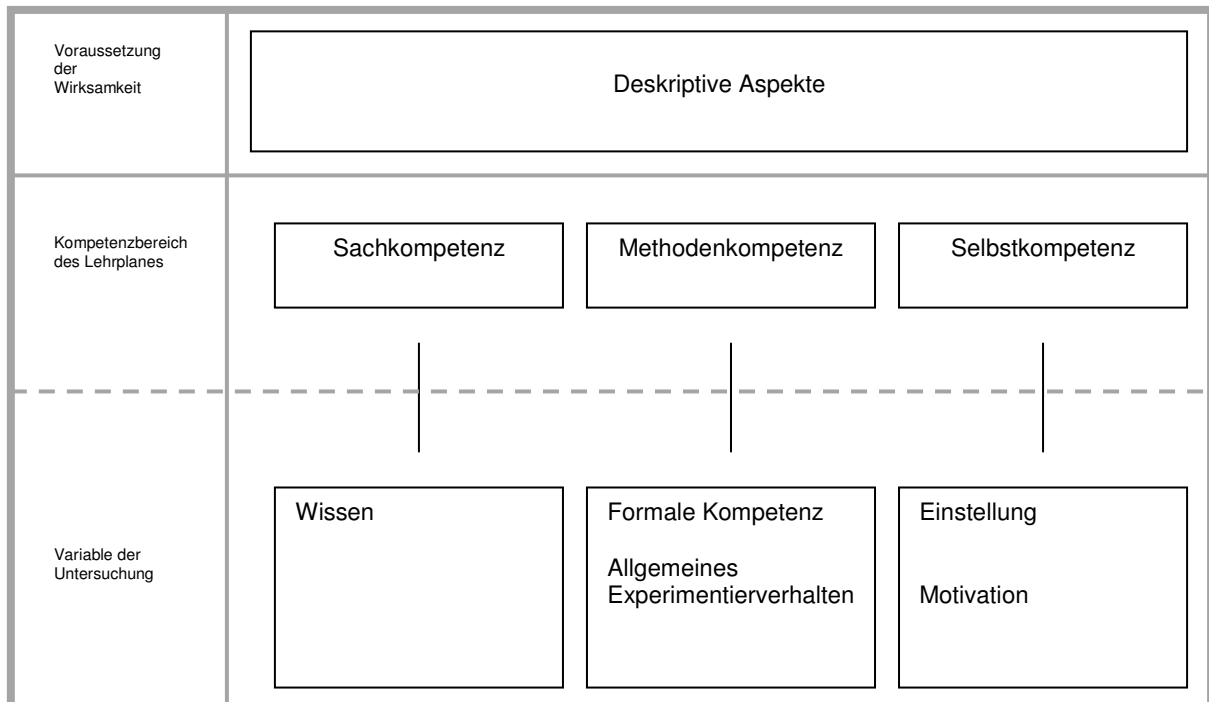
Die **Selbstkompetenz** soll durch die beiden Variablen Einstellung (Variablenname: [Einstellung gegenüber Phänomenen und ihrer Bearbeitung]) und Motivation (Variablenname: [Motivation zur tiefergehenden Bearbeitung physikalischer Fragestellungen]) operationalisiert werden. Eine positive Einstellung gegenüber dem Experimentieren und eine ausreichend hohe Motivation physikalische Probleme zu bearbeiten stellen die Grundlage für physikalische Lernprozesse dar.

Zusätzlich zu den genannten fünf Variablen soll eine sechste erfasst werden. Das durch sie verfolgte Ziel besteht in der differenzierten Erfassung der Erinnerung der Schülerinnen und Schüler an die Stationen (Variablenname: [deskriptive Aspekte]). Insbesondere aus zwei Gründen kommt dieser Variablen eine Sonderstellung zu:

1. Vorausgreifend auf den Abschnitt 2.5 in dieser Arbeit kann hier festgestellt werden, dass die Wirksamkeit einer Variablen immer nur im Vergleich mit den Prätestwerten ermittelt werden kann. Untersuchungsbedingt wird die Erinnerung der Schülerinnen und Schüler an die MINIPHÄNOMENTA im Prätest nicht erhoben. Insofern ist eine Zuweisung der Ergebnisse zu den einzelnen Wirksamkeitsstufen nicht möglich.
2. Die Erinnerung an die Stationen führt nicht notwendigerweise zu einer Veränderung im Verhalten beziehungsweise der Einstellung, der Motivation usw. Anders formuliert: Auch wenn sich die Schülerinnen und Schüler an die Stationen der MINIPHÄNOMENTA erinnern, so ist hiermit kein Indikator für die Vertiefung der Physikalischen Bildung gegeben.

Im Rahmen dieser Arbeit verstehe ich die Erinnerung der Schülerinnen und Schüler als eine **Voraussetzung der Wirksamkeit**. Nur durch die dauerhafte kognitive Repräsentation der Versuchstationen sind Veränderungen auf den übrigen Variablen möglich. Diese Interpretation wird durch die Tatsache gestützt, dass die Schülerinnen und Schüler die Stationen als eidetische Vorstellungsbilder repräsentieren (vgl. hier insbesondere die Abschnitte 3.2.2.1, 3.2.3.1 und 3.3.1). Dieses Ergebnis macht die zentrale Stellung der Versuchstationen deutlich.

Die folgende Grafik stellt zur besseren Übersicht die empirische Definition des Wirksamkeitsbegriffes anhand der sechs Variablen inklusive ihres Bezugs zum Lehrplan zusammenfassend dar:



Grafik 1.4.1.1: Die Variablen im Kontext der Untersuchung;
- die Abhängigen Variablen der Untersuchung und ihr Lehrplanbezug
Der Verfasser

In den folgenden Abschnitten werden die theoretischen Hintergründe der Variablen und, sofern vorhanden, bereits vorliegende empirische Ergebnisse erörtert. Dabei ist es wichtig zu betonen, dass es nicht das Ziel sein kann den jeweiligen Forschungsstand zu den Variablen dazustellen. Vielmehr werden die theoretischen und empirischen Ergebnisse selektiv vor dem Hintergrund der später einzuführenden Messverfahren erörtert.

1.4.2 Deskriptive Aspekte

Unter dem Variablennamen [deskriptive Aspekte] verstehe ich die Erinnerung der Schülerinnen und Schüler an die Stationen der MINIPHÄNOMENTA.

Die theoretischen Grundlagen von Fragestellungen über den Prozess der Erinnerung sind Vorstellungen über das Gedächtnis aus der allgemeinen Psychologie. Im Zusammenhang dieser Arbeit soll eine einfache, stark systematisierte Vorstellung des Gedächtnisses von Edelman (2000, S. 164 ff.) verwendet werden. Er unterscheidet im Gedächtnis drei Speicherarten:

1. Das sensorische Gedächtnis: Hier bleiben die Sinneseindrücke weniger als eine Sekunde gespeichert. Sein Inhalt ist ein detailliertes Abbild der vorausgegangenen Wahrnehmung.
2. Das Kurzzeitgedächtnis: Hier überdauern Inhalte etwa 15 Sekunden. Es wird häufig auch als Arbeitsspeicher bezeichnet. Enthalten sind hier diejenigen Informationen, die gerade verwendet werden. Im Gegensatz zum sensorischen Gedächtnis enthält das Kurzzeitgedächtnis bereits bearbeitete Informationen.
3. Das Langzeitgedächtnis: In ihm gespeicherte Inhalte können dauerhaft behalten werden. Es hat eine immense Kapazität.

Doch wie lange können wir uns Dinge merken? Ebbinghaus (1885, nach Edelman 2000, S. 166) ermittelte in diesem Zusammenhang die so genannte Vergessenskurve: Er ließ die Versuchspersonen sinnlose Silben, wie zum Beispiel TAK, PIR und GAN auswendig lernen. Das Ergebnis seiner Untersuchungen: Das Vergessen geht zu Beginn sehr schnell. So konnten bereits nach wenigen Minuten immer geringere Teile der Silben memoriert werden. Mit zunehmendem Abstand geht das Vergessen immer langsamer. Bezogen auf die Stationen der MINIPHÄNOMENTA hieße es, dass zu Beginn relativ viele Aspekte vergessen werden, mit fortschreitendem zeitlichen Abstand sich die Anzahl der erinnerten Stationen aber stabilisiert.

Sind aber nun die genannten Ergebnisse Ebbinghausens für die MINIPHÄNOMENTA zu erwarten? Aus der Sicht des Verfassers sprechen hiergegen vor allem drei Aspekte:

1. Zunächst werden in Bezug auf die interaktiven Stationen sinnvolle Inhalte gelernt. Durch das aktive Experimentieren bauen die Kinder eine Beziehung zu den Exponaten auf. Sie speichern die Exponate wahrscheinlich, was vor allem für die Lebensphase der Kindheit und Jugend typisch ist, in der Form von eidetischen Bildern, Vorstellungsbilder, die den Charakter von Wahrnehmungen haben (vgl. Edelmann 2000, S. 167 f.).
2. Die MINIPHÄNOMENTA ist in Bezug auf das übliche Umfeld der Schule hoch salient, das heißt sie ist als Stimulus völlig anders geartet, als die Umgebung, das setting Schule (vgl. Strebe, Jonas, Hewstone 2002, S. 126 ff.). Die Besonderheiten der MINIPHÄNOMENTA sind in dieser Arbeit vor allem in dem Abschnitt 1.3 herausgearbeitet worden und sollen daher an dieser Stelle nicht wiederholt werden.
3. Der intensive Umgang mit den Stationen: Das Experimentieren mit allen Sinnen, der Umgang mit dem Phänomen, das heißt das Beobachten, Sortieren der Beobachtungen das Nachdenken über Erklärungen und die Diskussion mit Mitschülerinnen und Mitschülern fördern die multiple Repräsentation der Inhalte.

Insofern ist bei der Erinnerung an die Stationen mit einer geringen Vergessensrate zu rechnen. Dies zeigen auch erste querschnittliche Ergebnisse (vgl. Abschnitt 1.3.3 in dieser Arbeit).

1.4.3 Einstellung gegenüber Phänomenen und ihrer Bearbeitung

Eine klassische Theorie des Einstellungsbegriffes ist das Drei-Komponenten-Modell von Rosenberg und Hovland (1960). Es beschreibt die innere Struktur von [Einstellungen]. Jede Einstellung weist nach diesem Modell eine ‚kognitive‘, eine ‚affektive‘ und eine ‚verhaltensbezogene‘ Facette auf:

„Eine [Einstellung] ist die Bereitschaft auf ‚kognitiver‘, ‚affektiver‘ und ‚verhaltensbezogener‘ Ebene bestimmte Reaktionen zu zeigen“
(Hannover, Bettge 1993, S.16).

Bezogen auf die Untersuchungsfrage kann die Definition wie folgt angewendet werden: Ein Schüler kann den Physikunterricht für subjektiv sinnvoll halten. Dieser habe viel mit seinem Leben gemein (‚kognitive Komponente‘). Ebenso kann er den Unterricht ‚emotional‘ bejahen. Er macht ihm Spaß. Auf der ‚Verhaltensebene‘ zeigt sich die Einstellung zum Beispiel im Engagement im Fach Physik.

Die oben genannten Autoren fassen die [Einstellung] als eine Bereitschaft auf. Ein linearer, kausaler Zusammenhang zwischen Einstellung und tatsächlichem Verhalten kann nicht angenommen werden. Diese Vermutung stützen auch andere, frühere Untersuchungen. LaPiere (1934, nach Stroebe, Jonas, Hewstone 2002, S.266 ff.) befragte in den USA Hotelbesitzer, ob in ihrem Hotel auch Chinesen übernachten dürfen¹⁶. 92% der Befragten verneinten dieses Anliegen. Im Anschluss suchte ein chinesisches Ehepaar eine Vielzahl von Hotels auf. Lediglich in einem wurden sie abgewiesen.

Dengler (1995) untersuchte die [Einstellung] gegenüber der Physik. Unter der Verwendung eines Fragebogens, genauer eines semantischen Differentials erhob er die Einstellung gegenüber der Physik vergleichend zu anderen Bereichen, zum Beispiel der Literatur, aber auch der anderen Naturwissenschaften, bei unterschiedlichen Personengruppen.

Das zentrale Ergebnis der Untersuchung: Es gibt eine hohe Akzeptanz gegenüber der Physik auf ‚kognitiver‘ Ebene. Stets wurde ihre Notwendigkeit betont, vor allem auch ihre wachsende Bedeutung für unsere Zukunft. Weniger als die Hälfte der Befragten kann sich jedoch für das Fach begeistern. Selbst Physikstudenten haben keine ‚affektive‘ Verbindung zu ihrem Studienfach.

¹⁶ In den USA gab es zum Untersuchungszeitpunkt gegenüber Mitbürgerinnen und Mitbürgern mit chinesischem Migrationshintergrund eine negative, in Teilen feindselige Einstellung.

Physikunterricht wird von den Versuchspersonen bei Dengler zuerst mit Formeln und abstrakten Erklärungen in Verbindung gebracht. Er sei im Wesentlichen mathematisch überlastet. Diese Formalisierung gilt als das stärkste Hindernis für den Zugang zur Physik. Physik und Mathematik weisen starke Übereinstimmungen in ihren Eigenschaftsprofilen auf. Die durchschnittliche Korrelation beträgt .4 und ist auf einem Alpha = 5% Niveau signifikant.

In Bezug auf die [Einstellung] gegenüber der Physik konnte Dengler deutliche Geschlechtsabhängigkeit nachweisen: Mädchen und Frauen weisen eine konsistent negativere Einstellung gegenüber der Physik auf.

Diesem negativen Bild von der Physik stehen die positiven Erfahrungen in Bezug auf das eigene Experimentieren gegenüber. Mehr als 90% der Befragten äußerten positive Erinnerungen an Demonstrations- und Schülerversuche.

1.4.4 Formale Kompetenz

[Formale Kompetenz] wird in dieser Untersuchung gefasst als die Gesamtheit der ‚logischen und experimentellen Kompetenzen‘, sowie eines ersten Einblickes in ‚spezifisch-naturwissenschaftliche Denkweisen‘.

Ich beginne mit den ‚logischen Fähigkeiten‘: „Logik ist die Lehre vom richtigen Denken, genauer von den Formen und Methoden (also nicht dem Inhalt), des richtigen Denkens. Sie kann nicht zeigen, was man Denken muss, sondern nur, wie man von irgendeinem Gegenstand ausgehend, denkend fortschreiten muss, um zu richtigen Ergebnissen zu gelangen“ (Störing 1987, S. 177).

‚Logische Kompetenzen‘ sind damit folglich solche, die die formellen Aspekte des Denkens, die Korrektheit des Schlussfolgerns ermöglichen. Der Fokus liegt hier nicht auf den Inhalten des Denkens, also zum Beispiel auf der zutreffenden Lösung für eine Aufgabe, sondern vielmehr auf den Denkbemühungen die zu diesem Ergebnis geführt haben.

In der Logik können zwei Arten des Schließens unterschieden werden (vgl. Speck 1980):

1. deduktives Schließen: Es orientiert sich am Allgemeinen und leitet aus ihm die Konstitution des Besonderen ab. Aus den Prämissen ergibt sich die Schlussfolgerung zwingend¹⁷.
2. induktives Schließen: Es ist der einzige Weg zur neuen Erkenntnis. Hier wird genau anders herum gedacht: Aus dem Speziellen wird das Allgemeine erschlossen. Formell lässt er sich wie folgt beschreiben „Wenn eine große Anzahl A unter einer großen Anzahl von Bedingungen beobachtet wird und wenn alle diese beobachteten A ohne Ausnahme die Eigenschaft B besitzen, dann besitzen alle A die Eigenschaft B (Chalmers 2001, S. 40).“

Haben Kinder diese metakognitiven Kompetenzen, das heißt sind sie in der Lage korrekt induktiv und deduktiv zu schließen? Der bisherige Forschungsstand zum deduktiven Denken von Kindern ist äußerst heterogen. Es spricht aber dennoch vieles dafür, dass Kinder ab dem Alter von 8 Jahren einfache deduktive Schlüsse ziehen können (vgl. Oerter, Montada 1995, S.593 ff.). In Bezug auf das induktive Schließen ergibt sich ein optimistischeres Bild. Kinder scheinen die diesem Schluss zugrunde liegende Fähigkeit des Vergleiches schon früher zu entwickeln (vgl. Oerter, Montada 1995, S. 599 ff.).

¹⁷ Die zahlreichen Fragen und Probleme, zum Beispiel woher die erste Aussage stammt, das so genannte Anfangsproblem, sollen hier aus Umfangsgründen nicht diskutiert werden. Gleiches gilt auch für das folgende Induktionsprinzip, dessen logische Probleme, seine Zirkularität usw. Der Verfasser verweist auf Standardlehrbücher der Wissenschaftstheorie (vgl. Speck 1980; Chalmers 2001)

Der zweite in dieser Variablen erfasste Teilbereich betrifft die ‚spezifisch naturwissenschaftlichen Fähigkeiten‘. Er lässt sich beispielhaft durch folgende Kernfragestellungen kennzeichnen: Gelingt den Kindern typisch naturwissenschaftliches Denken, nehmen sie also zum Beispiel eine Verknüpfung von Empirie und Theorie vor, erkennen und trennen sie Variablen, verstehen sie das Prinzip der kausalen Verursachung? Untersuchungen zu dieser Thematik liegen nicht vor, so dass keine Aussagen darüber getroffen werden können, ob diese Fähigkeiten bei den Schülerinnen und Schülern der Primarstufe vorhanden sind.

Der letzte Teilbereich beschreibt die ‚experimentellen Fähigkeiten‘. Besitzen die Kinder ausreichende experimentelle Kompetenzen, manipulieren sie also die Variablen, tun sie dies mit der notwendigen Gründlichkeit, gehen sie systematisch vor, benötigen sie wenig Unterstützung usw.? Wie bereits mehrfach zitiert, weisen die TIMSS und die PISA-Studie darauf hin, dass diese Kompetenzen in der siebten und achten Klasse in nur sehr geringem Umfang vorliegen (vgl. Baumert, Lehmann 1997, S. 55 f.). Insofern ist es zu erwarten, dass derartige Fähigkeiten in der Grundschule nicht vorhanden sind.

1.4.5 Allgemeines Experimentierverhalten

Unter dem Begriff des [allgemeinen Experimentierverhaltens] verstehe ich das Verhalten der Schülerinnen und Schüler in einer offenen Experimentiersituation unter der besonderen Berücksichtigung formaler Aspekte des Experimentierens, der Vertieftheit in das Experiment und dem Aspekt des sozialen Lernens an den Versuchen.

Das Konzept des [allgemeinen Experimentierverhaltens] der Schülerinnen und Schüler basiert auf einer Weiterentwicklung der Theorie der Lerntiefe (vgl. Strauss 1987, nach Schließmann 2005) die im Folgenden beschrieben werden soll.

Dabei impliziert der Begriff der Lerntiefe die unterschiedlich umfassende kognitive Bearbeitung von Inhalten. Je umfassender ein Inhalt bearbeitet wurde, desto größer ist der Lernerfolg des Individuums in Bezug auf die Tiefe des Verständnisses, den Umfang des gewonnen Wissens und die Dauerhaftigkeit des Lernerfolges.

Strauss (1987, nach Schließmann 2005, S. 30 ff.) stellte fest, dass sich die Lerntiefe durch Verhaltensbeobachtung messen lässt. Er entwickelte ein Verfahren, anhand dessen das Lernverhalten von Besuchern in Science-Centern an interaktiven Stationen klassifiziert und beurteilt werden kann. Insgesamt sieben verschiedene Lern-Verhaltensweisen konnten unterschieden werden. Diese werden in drei Kategorien klassifiziert. Die sieben Verhaltensweisen treten gemeinsam in individuell-unterschiedlichen Kombinationen auf. Eine hohe Lerntiefe ist dabei so definiert, dass möglichst viele der sieben Lern-Verhaltensweisen beim Umgang mit der interaktiven Station auftreten. Die folgende Tabelle zeigt die sieben Lern-Verhaltensweisen, klassifiziert in drei Kategorien:

Lern - Verhaltensweise	Klassifikation
1. Verbringt Zeit damit anderen beim Experimentieren zu zusehen	I. Eingangsverhalten
2. Erhält Informationen oder Hilfe durch Betreuer oder andere Besucher	
3. Wiederholt den Versuch	
4. Zeigt positive Emotion als Reaktion auf die Durchführung der Tätigkeit	II. Übergangsverhalten
5. Bezieht sich auf voraus gegangene Erfahrungen beim Experimentieren	
6. Sucht und nimmt Anteil an Informationen	
7. Engagiert und vertieft	III. Durchbruchverhalten

Tabelle 1.4.5.1: Allgemeines Experimentierverhalten;
- Klassifikation der Lern-Verhaltensweisen
Strauss (1987), zitiert nach Schließmann 2005, S. 32

Schließmann (2005, S. 36 ff.) erweiterte das von Strauss entwickelte Verfahren um eine Kategorie, mit deren Hilfe negative Reaktionen auf die interaktive Station erfasst werden können:

Lern – Verhaltensweise	Klassifikation
8. Geht nach kurzem Verweilen an der Station weiter	IV. negative Reaktion
9. Bricht das Experiment im frühen Stadium ab	

Abbildung 1.4.5.2: Allgemeines Experimentierverhalten;
- Klassifikation negativer Reaktionen auf einen Versuch
Schließmann 2005, S. 36 ff.

Seine Untersuchung zeigte weiterhin, dass die Lerntiefe maßgeblich von der Anzahl der an dem Versuch beteiligten Personen abhängt. So wurden die Lerntiefen „Übergangsverhalten“ und „Durchbruchverhalten“ sehr viel häufiger in Kleingruppen von drei bis sechs Personen erreicht, als in Einzel- oder Partnerarbeit. Ferner zeigte sich, dass schwerere Aufgaben eher mit einer höheren Lerntiefe bearbeitet werden als leichtere.

Das von Schließmann entwickelte Verfahren soll im Rahmen dieser Untersuchung nochmals modifiziert und erweitert werden. Ausschlaggebend für diese Entscheidung war dabei, dass das Verfahren für Untersuchungen in Science-Centern entwickelt wurde, folglich nicht auf die Spezifika des settings Schule zugeschnitten ist. Vor diesem Hintergrund erscheinen besonders die Kategorien: „fordert weitere Informationen“ und „sucht und nimmt Anteil an Informationen“ ungeeignet. Gibt es doch bei der Untersuchung keine Betreuer, wie in einem Science Center, bei denen man Informationen und Hintergrundwissen erfragen könnte. Aus diesem Grund sollen diese Kategorien entfallen.

Weiterhin stellt sich die Frage nach der sicheren Beobachtbarkeit aller Lern - Verhaltensweisen. Dies betrifft vor allem die Kategorien „bezieht sich auf voraus gegangene Erfahrungen beim Experimentieren“ und „verwendet frühere Erkenntnisse“. In einigen Fällen sind diese Prozesse durch Verhaltensbeobachtung sicher zu dokumentieren. Keinesfalls ist dies jedoch zwangsläufig gegeben.

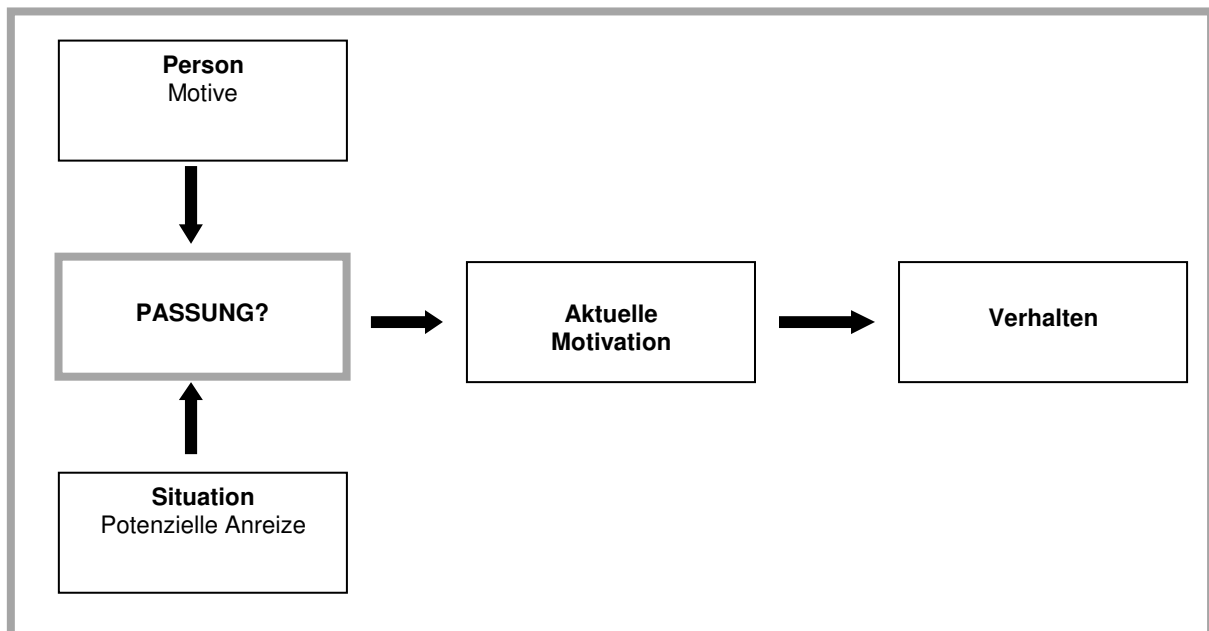
Eine letzte Erweiterung ist zum Verständnis des hier vorgeschlagenen Verfahrens erforderlich. Diese betrifft den Themenkomplex des sozialen Lernens. Integraler Bestandteil der Frage nach dem allgemeinen Experimentierverhalten, ist der Aspekt des sozialen Lernens. Es soll also der Frage nachgegangen werden, ob und wie weit gehend die Schülerinnen und Schüler in der Lage sind, an den Experimenten produktiv gemeinsam zu arbeiten.

Empirische Ergebnisse liegen zu dieser Variablen noch nicht vor. Es ist allerdings anzunehmen, dass die Schülerinnen und Schüler ein wenig elaboriertes [allgemeines Experimentierverhalten] zeigen, da das Experimentieren in der Grundschule in der Regel nicht geübt wird (vgl. Abschnitt 1.2.3 in dieser Arbeit).

1.4.6 Aktuelle Motivation zur tiefergehenden Bearbeitung physikalischer Fragestellungen

„[Motivation] [...] ist die allgemeine Bezeichnung für alle Prozesse, die körperliche und psychische Vorgänge auslösen, steuern oder aufrechterhalten.“ (Zimbardo, Gerrig 1999, S. 319).

Der Feldtheorie von Lewin folgend wird angenommen, dass sich das konkrete Verhalten im Feld immer durch die Kräfte des Innen und Außen ergibt¹⁸, das heißt durch die psychische Konstitution der Person und durch das Setting. Ein Teil der psychischen Konstitution der Person sind die Motive. Hierbei handelt es sich um hochgeneralisierte, dauerhafte Präferenzen der Person, bestimmte Klassen von Anreizen zu bevorzugen. Passen Motive und Merkmale des Settings, d.h. potenzielle Anreize zusammen, so entsteht die aktuelle Motivation, die Verhalten auslöst, steuert und aufrechterhält:



Grafik 1.4.6.1 : Aktuelle Motivation;
- klassisches Modell der Motivationspsychologie:
Rheinberg, Vollmeyer, Burns, Diagnostica, Nr. 47, S.3

¹⁸ Die Sichtweise drückt sich in formalisiert in der fundamentalen Verhaltensgleichung aus:
 $V = f(P, U)$: Das Verhalten ist eine Funktion von Person- und Umweltmerkmalen.

Die aktuelle Motivation wird bei den genannten Autoren durch den FAM (Fragebogen zur Erfassung der aktuellen Motivation) empirisch definiert. Der FAM operationalisiert den Begriff der aktuellen Motivation auf vier Skalen:

1. Interesse
2. Erfolgswahrscheinlichkeit
3. Misserfolgsbefürchtungen
4. Herausforderung

Der Versuchsperson wird bei diesem Design eine konkrete Aufgabe gestellt. Zum Beispiel, die Lösung der Aufgabe „Turm von Hanoi“ oder das Spielen von „Flottenmanöver“. Nach der Aufgabenstellung, aber noch bevor mit der Bearbeitung des Problems begonnen wird, sollen sich die Versuchspersonen in einem Fragebogen auf einer jeweils siebenstufigen Skala selbst einschätzen.

Die zu lösenden Aufgaben variierten erheblich: Rheinberg, Vollmeyer und Burns (2001, S. 7) nennen insgesamt sieben unterschiedliche Aufgabentypen, bei denen der FAM angewendet wurde. Der FAM erfasst die aktuelle Lern- und Leistungsmotivation. Es ist daher davon auszugehen, dass der Aufgabentyp einen entscheidenden Einfluss auf die Ergebnisse der Skalen und damit auch auf den Gesamtwert hat. Dies konnte in einem Vergleich von zwei Aufgabentypen („Turm von Hanoi“ und „Flottenmanöver“) anhand einer gleichen Stichprobe eindeutig festgestellt werden. Bei der Aufgabe „Flottenmanöver“ waren die subjektiven Misserfolgsbefürchtungen geringer, die Erfolgswahrscheinlichkeit höher und das Ausmaß der Herausforderung niedriger als beim „Turm von Hanoi“ (t-Test, Signifikanz auf einem Alpha = 1%-Niveau).

Doch wie steht es um die Motivation von Schülerinnen und Schülern physikalische Fragestellungen zu bearbeiten? Dieser Frage wurde in der TIMSS für Jugendliche in der siebten und achten Klassenstufe nachgegangen.

Die Autoren operationalisieren Motivation ähnlich der Konzeption des FAM, dem die obigen Ausführungen zugrunde liegen. Lediglich bei Punkt vier der Tabelle ergibt sich ein Unterschied zwischen der TIMSS und dem FAM. Insgesamt lässt sich aber eine starke Ähnlichkeit der Konzeptionen erkennen, so dass das Konzept der TIMSS als eine Anpassung des FAM an die Erfordernisse der Schule aufgefasst werden kann:

FAM (Motive)	TIMSS (Motive)
1. Interesse	1. Sachinteresse
2. Erfolgswahrscheinlichkeit	2. Vertrauen in die eigenen Fähigkeiten
3. Misserfolgsbefürchtungen	3. Prüfungsangst
4. Herausforderung	4. Schulunlust/Schulfreude

Tabelle 1.4.6.1: Aktuelle Motivation;
- Motive des FAM versus der TIMSS
Der Verfasser

Im Folgenden sollen die Ergebnisse zu den einzelnen Motiven zusammengefasst und interpretiert werden:

Punkt eins (Interesse / Sachinteresse) wurde bereits im Abschnitt 1.2.2.2 in dieser Arbeit ausführlich erörtert und soll daher an dieser Stelle nicht weiter diskutiert werden.

Betrachtet man das Vertrauen in die eigenen schulischen Fähigkeiten so erkennt man im Durchschnitt aller naturwissenschaftlichen Fächer eine ausgeprägte Stabilität. Es gibt kaum Unterschiede zwischen den Schulformen. Auch lassen sich keine statistisch bedeutsamen Unterschiede zwischen Jungen und Mädchen nachweisen.

Differenziert man nach Fächern, lassen sich jedoch deutliche Unterschiede herausarbeiten: Im Fach Biologie ist das Vertrauen in die eigenen Fähigkeiten am höchsten. Geschlechtsspezifische Unterschiede sind praktisch nicht vorhanden. Im Fach Physik zeichnet sich ein völlig anderes Bild ab. Die Werte liegen deutlich tiefer, zum größten Teil unterhalb des Neutralitätspunktes der Skala. Es gibt statistisch signifikante Unterschiede zwischen den Schulformen. Das Vertrauen in die eigenen Fähigkeiten ist in der Hauptschule am geringsten, in der Schulform Gymnasium am höchsten. Die Realschule nimmt einen mittleren Wert ein.

Mädchen schätzen ihre eigenen Fähigkeiten systematisch zu gering ein. Sie schreiben sich, in Relation zu den Jungen gleicher Leistung, geringere Fähigkeiten zu. Jungen überschätzen hingegen ihre Fähigkeiten systematisch. Korreliert man den Durchschnitt der Selbsteinschätzung mit dem Leistungsdurchschnitt, so ergibt sich ein bizarres Bild: „Feeling good and doing bad“ (Baumert, Lehmann 1997, S.173).

Nimmt man Prüfungsangst als Indikator für schulische Misserfolgsbefürchtungen, so ergibt sich ein positives Bild. Die empfundene Prüfungsangst liegt in den naturwissenschaftlichen Fächern deutlich unterhalb des Neutralitätspunktes der Skala. Im Fach Physik ist in der achten Klasse ein gewisser Anstieg der Prüfungsangst zu verzeichnen. Dieser ist allerdings statistisch nicht bedeutsam. Diese Befunde sind konsistent mit den obigen Annahmen bezüglich der durchschnittlichen Überschätzung der schulischen Kompetenz in einem Fach. Ein hohes Vertrauen in die eigenen Fähigkeiten in einem Fach geht offensichtlich einher mit einem geringen Maß an Prüfungsangst.

Die Ausprägung der Schulunlust ist unabhängig vom Geschlecht und der besuchten Schulform. Die Messwerte liegen nahe dem Neutralitätspunkt der Skala.

Zusammenfassend kann damit das Bild der Pubertät als einer „Sturm und Drang-Phase“, eines stürmischen Lebensabschnittes, in dem das Leistungsvertrauen der Jugendlichen sinkt, die Schulunlust rapide ansteigt, und sich die Jugendlichen vorrangig außerschulischen Betätigungen zuwenden, nicht aufrecht erhalten werden. Vielmehr legt die Studie die relative Stabilität der motivationalen Lage Jugendlicher nahe.

Betrachtet man diese eher positiven Ergebnisse für die Lebensphase „Jugend“ und die Tatsache des Vorliegens einer „sensiblen Periode“ für physikalische Fragestellungen in der Kindheit (vgl. Abschnitt 1.2.3 in dieser Arbeit), so ist von einer erhöhten Motivation der Schülerinnen und Schüler, physikalische Fragestellungen zu bearbeiten, auszugehen.

1.4.7 Wissen

Unter [Wissen] über einen bestimmten Inhalt verstehe ich im Rahmen dieser Arbeit die Summe der im Gedächtnis verfügbaren Propositionen und Verknüpfungen, die in der Form eines semantischen Netzwerkes gespeichert sind. Es kann zwischen ‚deklarativem‘ und ‚prozeduralem‘ Wissen unterschieden werden.

Die theoretisch Grundlage dieser Wissensdefinition ist das „ACT-Modell“ (vgl. Asendorpf 1999, S. 65 ff.): Es unterscheidet grundsätzlich zwischen zwei Arten des Wissens: prozedurales (Regeln, Strategien: „Wie-Wissen“) und deklaratives Wissen (Aussagen über Dinge, Prozesse und Ereignisse: „Was-Wissen“). Beide Formen des Wissens kommen sowohl im Langzeit- als auch im Kurzzeitgedächtnis vor. Das Wissen ist als propositionales Netzwerk organisiert, wobei die Knoten das deklarative und die Verbindungen das prozedurale Wissen repräsentieren. Das Kurzzeitgedächtnis besteht aus den momentan aktiven Knoten, das Langzeitgedächtnis besteht aus der Summe der verfügbaren Knoten und Verknüpfungen. Informationsverarbeitung geschieht nach der Vorstellung dieses Modells, in Analogie zu neuronalen Prozessen, durch eine Erregung von Knoten durch externe Wahrnehmungen oder interne Verarbeitungsschritte. Diese Erregung breitet sich dann in verwandte Knoten aus, wobei mit zunehmender Distanz und verstreichender Zeit die Erregung immer schwächer wird. Die „Stärke“ eines Knotens variiert in Abhängigkeit seiner Abrufhäufigkeit. Je öfter ein Knoten in Gebrauch ist, desto leichter ist er aktivierbar.

Welche Art von Wissen kann im Rahmen dieser Arbeit erhoben werden? Hierzu betrachte man zunächst den bereits mehrfach genannten Lehrplan des Faches Heimat- und Sachunterricht für die Primarstufe. Hier wird definiert, dass die Schülerinnen und Schüler erst zum Ende der vierten Klassenstufe einfache naturwissenschaftliche Gesetzmäßigkeiten beschreiben können müssen. Weiterhin muss das Ziel von ASIP bedacht werden. Es besteht keinesfalls, dass wurde insbesondere in den Abschnitten 1.2.1.2 und 1.3.1 deutlich, im vorrangigen Aufbau physikalischen Wissens. Vielmehr sollen die Schülerinnen und Schüler, Bezug nehmend auf den erweiterten Begriff der Physikalischen Bildung eine Vielzahl von Erfahrungen machen, die weit über die bloße Akkumulation von Wissen hinausgeht. Aus beiden Gründen ist die Erfassung der Wirksamkeit der MINIPHÄNOMENTA zum Beispiel durch einen Test oder eine Arbeit im traditionellen Verständnis nicht möglich.

Eine Alternative bietet das Verfahren der Concept-Maps: Ihm liegt der Gedanke zugrunde, dass sich die kognitive Struktur über einen bestimmten Gegenstand bildlich darstellen lässt. Das semantische Netzwerk, bestehend aus ‚deklarativem‘ und ‚prozeduralem‘ Wissen, wird in Form eines Schaubildes mit Knoten und Verknüpfungen dargestellt (vgl. Starcke 2004, S. 17ff.). Concept-Maps können in drei unterschiedlichen Varianten durchgeführt werden: Computerbasierte Versionen, „paper and pencil-Versionen“ und Mischformen aus beidem (vgl. Anderson-Inmann, Horney 1996, S. 6ff.).

Concept-Maps können weiterhin in Bezug auf die Offenheit beziehungsweise Geschlossenheit des Verfahrens und die angewandte Methode der Auswertung unterschieden werden. Zunächst zum ersten Kriterium: Hier können prinzipiell drei Möglichkeiten unterschieden werden:

1. Völlig standardisierte Verfahren: Sämtliche Begriffe und Relationen sind vorgegeben und müssen lediglich in die richtige Reihenfolge gebracht werden.
2. Offene Verfahren: Knoten und Relationen werden selbst definiert und in die passende Reihenfolge gebracht.
3. Halbstandardisierte Verfahren: Mischform der beiden erstgenannten Varianten

Auch die Auswertung von Concept-Maps kann auf unterschiedliche Weise geschehen. Es existieren quantitative und qualitative Verfahren, sowie Mischformen aus beiden. Am bekanntesten ist hier sicherlich die quantitative Auswertungsmöglichkeit des Computerprogramms CCmap (vgl. Reiska 2002) anhand von acht Variablen.

Das eingeführte Verfahren muss für den Kontext dieser Arbeit modifiziert werden. Für Schülerinnen und Schüler der Primarstufe scheint das Verfahren zu komplex zu sein. Darauf deuten die Ergebnisse von Holst (2005) hin, der herausfand, dass die Concept-Maps der Schülerinnen und Schüler im Durchschnitt nur sehr klein sind, folglich aus wenigen Begriffen bestehen. Insofern wäre es möglich, dass bedingt durch die Komplexität des Verfahrens die Vorstellungen der Schülerinnen und Schüler über den Gegenstand unterschätzt werden. Insofern entscheide ich mich im Kontext dieser Arbeit für ein Verfahren in dem das Wissen anhand von Knoten und Verknüpfungen beschrieben wird, diese jedoch lediglich in einer unidirektionalen Kausalitätskette und nicht als Concept-Map zu ordnen sind. Die Fähigkeit zur Bildung einer Rangreihe ist bei Schülerinnen und Schülern der dritten und vierten Klassenstufe, dies belegen die Ergebnisse der Entwicklungspsychologie (vgl. Oerter, Montada 1995, S. 519 ff.), vorhanden.

Durch dieses vereinfachte Verfahren der Concept-Maps, das sich an dem eigenständigen Nachdenken über einen vorher durchgeführten Versuch orientiert, ist es möglich, das [Wissen] theorieadäquat zu erfassen. Erhoben wird zum einen ‚prozedurales Wissen‘, zum anderen auch ‚deklaratives‘, das jedoch nicht vor der Versuchsdurchführung vorhanden sein muss, sondern vielmehr in der konkreten Situation des Experimentierens auf der Basis von Beobachtungen generiert wird.

Leider liegen für diese Variable keine Forschungsergebnisse vor. Die bisher genannten Ergebnisse legen jedoch die Vermutung nahe, dass die Schülerinnen und Schüler in Bezug auf diese Variable Defizite aufweisen werden.

2 Methodische Aspekte

2.1 Der Begriff der Wirksamkeit und seine methodischen Implikationen

Betrachtet man den Titel dieser Arbeit, so wird die zentrale Stellung des Begriffes der Langzeitwirksamkeit für den Kontext der vorliegenden Untersuchung deutlich. Das herausragende Definitionsmerkmal des Terminus der Langzeitwirksamkeit ist der in ihm enthaltene Begriff der Wirksamkeit, der im Folgenden erläutert werden soll.

Nach der Durchsicht gängiger pädagogischer Fachwörterbücher kann festgestellt werden, dass es sich beim Begriff der Wirksamkeit, im Gegensatz zu anderen Wissenschaftsdisziplinen - man denke hier nur an die Pharmakologie oder die Medizin - nicht um einen „einheimischen Begriff“ der Erziehungswissenschaft handelt (vgl. Schaub, Zinke 2000; Lenzen 1998; Benner, Oelkers (Hrsg.) 2004). Lediglich bei Lenzen (1989) findet sich unter dem Stichwort „Wirksamkeitsüberprüfung“ ein Querverweis mit Bezug auf den Begriff der Evaluation. Gestützt wird diese Annahme auch durch die Tatsache, dass selbst in einschlägigen Handbüchern der Schulpädagogik (vgl. Helsper, Böhme 2004) der Begriff der Wirksamkeit, zum Beispiel in Bezug auf die Wirksamkeit von Unterricht, nicht explizit unter diesem Terminus diskutiert wird. Der verwandte aber nicht deckungsgleiche Begriff der Effektivität wird hingegen in mehreren Zusammenhängen ausgiebig erörtert (vgl. Helsper, Böhme 2004, S. 707). Auch in allgemein-erziehungswissenschaftlichen Einführungen in die Hauptbegriffe der Pädagogik¹⁹ ist der Begriff der Wirksamkeit nicht enthalten (vgl. Krüger, Helsper 1995).

Woher stammt die Zurückhaltung gegenüber diesem Begriff? Immerhin wird im Zuge neuer Steuerungsmodelle für die Schule der Nachweis von Wirksamkeit auf empirischem Wege gefordert²⁰ (vgl. Galuske 2005, S. 314ff.) und zum Beispiel in der PISA-Studie und „third international mathematics and science study“ praktisch durchgeführt.

¹⁹ Die Begriffe Erziehungswissenschaft und Pädagogik werden im Kontext dieser Arbeit synonym verwendet. Der Verfasser verweist auf die begriffliche Diskussion zum Beispiel bei Benner, Oelkers (Hrsg.) (2004), die beide Begriffe streng trennen.

²⁰ Kernelement des neuen Steuerungsmodells ist das so genannte Kontraktmanagement: Verwaltungen werden damit als autonome business-units konstruiert, sie operieren auf der Basis eines eignen Betriebsplanes, die Fach- und Ressourcenverwaltung wird dezentralisiert usw. (vgl. Böttcher, Holtappels, Brohm 2006, S. 51ff.).

Noch unverständlicher wird die Tatsache dadurch, dass Bernfeld bereits 1925 in „Sisyphus oder die Grenzen der Erziehung“ explizit eine Wirksamkeitsüberprüfung der Konstrukte der „Pädagogiker“ forderte. Die „Pädagogiker“, er meint vor allem Rousseau, Locke, Kant, jene „[...] wunderlichen Astronomen, die nachts fest schlafen und sich morgens von den Sternen erzählen lassen, um nach Tische über sie zu denken und zu schreiben“ (ebd. 1925, S.31). Sie sind nicht nur der Praxis fern sondern entwerfen umfangreiche theoretische Konzeptionen deren pädagogische Wirksamkeit durch sie nicht geprüft wird. Genau dies müsse aber die Pädagogik als eine Wissenschaft leisten. Sie solle wirksame erzieherische Mittel entwickeln und überprüfen, um auf Kinder sittlich einwirken zu können.

Die Forderung einer Wirksamkeitsüberprüfung ist folglich nicht neu. Auch wird sie im Zuge der Expansion der Häufigkeit von Evaluationsvorhaben und internationalen Schulvergleichsstudien praktisch umgesetzt, ohne dass der Begriff der Wirksamkeit Einzug in den erziehungswissenschaftlichen Begriffsraum gefunden hätte.

Eine mögliche Begründung für dieses paradoxe Phänomen ist die Konstitution der universitären Erziehungswissenschaft, die in ihrer Tradition, trotz der heute vorherrschenden Pluralität der Paradigmen, häufig geisteswissenschaftlich geprägt ist. Das Erkenntnisparadigma passt nicht zur folgenden Definition des Wirksamkeitsbegriffes.

Wirksamkeit wird in Meyers enzyklopädischem Lexikon definiert als „ein Geschehen (im Sinne von Prozeß) oder Handeln bzw. deren Ergebnis (ein Ereignis oder Sachverhalt), das nach einem Kausalurteil in einem [...] Kausalzusammenhang mit einem anderen (vorausgehenden) Geschehen oder Ereignis bzw. Sachverhalt steht oder stehen soll [...].“ (ebd. 1979, S. 420). Der Begriff der Wirkung impliziert damit den Begriff der Kausalität (A ist die Ursache einer Wirkung B, wenn A B herbeiführt²¹).

Zentrales Definitionsmerkmal des Wirksamkeitsbegriffes ist die Kausalität.

Nach Dilthey kann man in den Geisteswissenschaften nur „verstehen“: Etwas als etwas Menschliches erkennen, nicht erklären, das heißt einen Vorgang oder ein Ereignis auf eine oder mehrere Kausalursachen zurückführen. Erziehung ist damit immer auf einen Verstehensprozess angewiesen, indem der Erzieher den Zögling besser versteht als dieser sich selbst. Sie ist zutiefst in geschichtliche Zusammenhänge eingebunden. Allgemeingültige Kausalitäten sind nach der Ansicht dieses Paradigmas nicht möglich, der Begriff der Wirksamkeit ist folglich nicht anwendbar.

²¹ Diese Form der Definition ist zwar plastisch, wissenschaftlich jedoch unbrauchbar aufgrund der Zirkularität der Argumentation: Es ist lediglich die Abfolge von Phänomenen feststellbar, „[...] niemals jedoch, dass Eines ein Anderes hervorbringt“ (Speck 1980, S. 319). Die Ursache oder besser ein funktionaler Zusammenhang wird als eine Vorbedingung gefasst auf die die Wirkung jederzeit und unbedingt folgt. Diese Aussagen führen zwangsläufig zum Ergebnis der Modellhaftigkeit von Erklärungen, denn letztendliche Gültigkeit (Verifikationsprinzip) kann nie bewiesen werden (vgl. Speck 1980).

Nach der Diskussion des Begriffes der Wirksamkeit kann nun zu seiner Erweiterung, dem Begriff der Langzeitwirksamkeit übergegangen werden. Hierunter verstehe ich lediglich die Wirksamkeit über einen längeren Zeitraum, wobei der Umfang dieses Zeitraumes nicht näher bestimmt wird. Auch besteht das zentrale Definitionsmerkmal in dem beschriebenen Kausalitätsverhältnis. Über die genannte Wirksamkeitsdefinition hinaus impliziert der Begriff der Langzeitwirksamkeit jedoch, dass die Wirkung, hier die der MINIPHÄNOMENTA, über einen längeren Zeitraum anhält.

Nach dieser grundsätzlichen Klärung des Begriffes der Langzeitwirksamkeit, kann dieser nun auf die Methodik der vorliegenden Arbeit angewandt werden. Dies geschieht als erste Orientierung anhand der folgenden Tabelle:

Nummer	Implikationen aus dem Wirksamkeitsbegriff
1	Die Notwendigkeit eines experimentellen Designs
2	Entwicklung von Verfahren um Wirksamkeit zu messen
3	Ebenen und Stufen der Wirksamkeitsmessung
4	Stufen der Wirksamkeit

Tabelle 2.1.1: Der Begriff der Wirksamkeit;
 - der Begriff der Wirksamkeit als methodischer Zentralbegriff der Arbeit
 Der Verfasser

Das in der Definition des Wirksamkeitsbegriffes geforderte Kausalitätsverhältnis kann im Rahmen der bestehenden sozialwissenschaftlichen Methoden nur durch die Verwendung eines Experimentes eingelöst werden. Die Zuweisung zur Experimental- und Kontrollgruppe ist dabei nicht zufällig, sondern durch das Kriterium der Teilnahme beziehungsweise Nichteintragung an der MINIPHÄNOMENTA bestimmt. Damit handelt es sich im streng methodologischen Sinne ein Quasiexperiment. Dies geschieht in zwei Untersuchungsstufen: Einem Ein-Gruppen-Prä-Posttest-Quasiexperiment und ein Zwei-Gruppen-Prä-Posttest-Quasiexperiment mit zwei Posttestmessungen (vgl. Bortz, Döring 1995, S. 520 ff.). Durch die Kombination beider ist es möglich die Wirksamkeit des Projektes umfassend zu beschreiben. Der Schwerpunkt der vorliegenden Studie liegt dabei auf der Verwendung des zweiten Designs, da hier die interne Validität deutlich höher ist und es weniger weitere methodologische Probleme gibt, als im Falle des Ein-Gruppen-Prä-Posttest-Quasiexperimentes. Dieses dient, aufgrund der mit diesem Vorgehen verbunden methodischen Probleme, in der vorliegenden Untersuchung lediglich als Ausblick.

Auf der Basis der operationalen Definition des Wirksamkeitsbegriffes werden im Methodenteil dieser Arbeit weiterhin Untersuchungsverfahren konzipiert mit denen das Konstrukt der Wirksamkeit der MINIPHÄNOMENTA zu unterschiedlichen Zeitpunkten bei Schülerinnen und Schülern der dritten und vierten Klassenstufe erhoben wird.

Abschließend werden im Methodenteil Stufen und Ebenen der Wirksamkeit eingeführt:

- Wirksamkeitsstufen: Diesem Begriff liegt die Einsicht zu Grunde, dass Wirksamkeit kein dichotomes Merkmal ist. Vielmehr sehe ich Wirksamkeit als ein komplexes und polytomes, dessen Extrema durch die beiden Pole HOCHWIRKSAM und unwirksam beschrieben werden. Die Zuweisung zu den einzelnen vier zu definierenden Wirksamkeitsstufen geschieht auf der Basis mehrerer Merkmale, wobei hier insbesondere der Effektgröße eine zentrale Rolle zukommt.
- Wirksamkeitsebenen: Die unterschiedlichen Wirksamkeitsebenen beziehen sich auf die interne Validität des zentralen Kausalitätsschlusses von der unabhängigen Variablen (treatment: Teilnahme beziehungsweise Nichtteilnahme an der MINIPHÄNOMENTA) auf die abhängige Variable. Auf der Ebene I wird Wirksamkeit als Prä-Posttest-Differenz ermittelt. Auf der Ebene II wird die Prä-Posttest-Differenz zusätzlich an den Ergebnissen einer Kontrollgruppe relativiert.

2.2 Das Untersuchungsdesign

Das Untersuchungsdesign meiner Arbeit ist zweistufig²² angelegt. Beide Stufen müssen aufgrund der Definition des Wirksamkeitsbegriffes zwingend experimentell organisiert werden. Nur durch ein solches Design lässt sich in den Sozialwissenschaften Kausalität exakt nachweisen²³. Dies geschieht, indem der Einfluss der unabhängigen Variablen (treatment: MINIPHÄNOMENTA) auf die abhängigen Variablen ([Wissen], [Einstellung] usw.), im Vergleich zu einer Kontrollgruppe ermittelt wird.

Die Untersuchungsstufe I stellt in dieser Arbeit das Zentrum der Analyse dar. Sie ist methodisch aufwendig gestaltet, um eine möglichst hohe Anzahl von Fehlern kontrollieren zu können. Bei der Untersuchungsstufe II handelt es sich um ein vergleichsweise einfaches Design, das die Validität der Ergebnisse deutlich einschränkt. Diese Untersuchungsstufe ist lediglich als ein Ergebnisausblick zu sehen, dessen Überprüfung ich anderen Vorhaben überantworte.

Bei dem Design der Untersuchungsstufe I handelt es sich nach der Definition von Bortz und Döring (1995, S. 521 f.) um einen Zwei-Gruppen-Prätest-Posttest-Plan mit zwei Messwiederholungen. Die folgende Tabelle stellt den Plan dar:

		t ₋₁ Monat	t ₀	t ₊₃ Monate	t ₊₆ Monate
3 Klasse	Experimentalgruppe I (3)	Prätest	treatment	Posttest I	Posttest II
	Kontrollgruppe (3)	Prätest			Posttest II'
4 Klasse	Experimentalgruppe I (4)	Prätest	treatment	Posttest I	Posttest II
	Kontrollgruppe (3)	Prätest			Posttest II'

Tabelle 2.2.1: Das Untersuchungsdesign;
- das quasiexperimentelle Design der Untersuchungsstufe I
Der Verfasser

²² Der Begriff der Untersuchungsstufe darf hier nicht mit dem der Wirksamkeitsebene beziehungsweise Wirksamkeitsstufe verwechselt werden:

1. Untersuchungsstufe: Die Untersuchung basiert auf zwei quasiexperimentellen Plänen. Die Untersuchungsstufe I ist ein Zwei - Gruppen - Prätest - Posttest - Plan mit dem die Wirksamkeit der MINIPHÄNOMENTA über einen Zeitraum von sechs Monaten untersucht wird. Die Untersuchungsstufe II dient der Erfassung der Wirksamkeit über einen Zeitraum von 27 Monaten. Hier wird ein Ein - Gruppen - Posttest - Plan verwandt.
2. Wirksamkeitsebene. Es werden zwei Ebenen eingeführt, die die Wirksamkeit des treatments auf zwei methodisch unterschiedlich kontrollierten Ebenen beschreiben.
3. Wirksamkeitsstufe: Wirksamkeit wird in dieser Arbeit als komplexes polytomes Merkmal begriffen. Insgesamt werden vier Wirksamkeitsstufen unterschieden. Die Zuweisung Ergebnisse zu diesen Stufen geschieht auf der Basis operationaler Kriterien.

²³ Die einzige Ausnahme bilden hier komplexe korrelationsstatistische Verfahren, mit denen sich ebenfalls, wenn auch nicht mit zwingender Evidenz Kausalität nachweisen lässt (vgl. Borz, S. 456 ff.).

Die Untersuchung findet in der dritten und vierten Klassenstufe statt. Für beide Klassen werden jeweils eine Experimental- und eine Kontrollgruppe gebildet. Die Zuweisung der Teilnehmerinnen und Teilnehmer geschieht durch ein Kriterium: Die Teilnahme beziehungsweise Nichtteilnahme an der MINIPHÄNOMENTA: Es handelt sich folglich um ein quasiexperimentelles Design. Aus methodischer Sicht muss zur Erhöhung der internen Validität ein Prätest durchgeführt und die Vergleichbarkeit von Kontroll- und Experimentalgruppe nachgewiesen werden. Es sind drei Messzeitpunkte vorgesehen:

1. Prätest: Ein Monat vor der Teilnahme der Experimentalgruppe I an der MINIPHÄNOMENTA
2. Posttest I: Drei Monate nach der Teilnahme an der MINIPHÄNOMENTA
3. Posttest II: Sechs Monate nach der Teilnahme an der MINIPHÄNOMENTA

Mit dem dargestellten Design lassen sich die Langzeitwirkungen der MINIPHÄNOMENTA über einen Zeitraum von insgesamt sechs Monaten umfassend messen.

Betrachtet man den obigen Plan, so fällt eine Besonderheit die Auslassung der Erhebung der Kontrollgruppe beim Posttest I auf: Die wahrscheinlichsten Quellen für Veränderungen, unabhängig vom Einfluss der MINIPHÄNOMENTA, liegen in Reifungseinflüssen und in durch externe Umstände zustande gekommenen systematischen Wissenszuwächsen (zum Beispiel Einflüsse des Unterrichtes). Beide Veränderungen lassen sich durch das dargestellte Design problemlos kontrollieren, so dass eine Reduktion der Datenmenge hier angebracht erscheint.

Eine weitere Fehlerquelle kann durch dieses Design jedoch nicht kontrolliert werden. Die MINIPHÄNOMENTA, dieser Umstand wurde im Theorieteil deutlich hervor gearbeitet, zeichnet sich wesentlich durch ihren Angebotscharakter aus. Hierdurch werden Effekte der Selbstselektion wahrscheinlich. Dieser Effekt tritt auf zwei Ebenen auf: Zunächst auf der Ebene der Schülerinnen und Schüler darüberhinaus auch auf der Ebene der teilnehmenden Schulen. Relativierend kann angeführt werden, dass die Schülerinnen und Schüler überaus intensiv an den Stationen der MINIPHÄNOMENTA arbeiten (vgl. Holst 2005). Der Aufforderungscharakter eben dieser Stationen ist so groß, dass diese Effekte insgesamt gering ausfallen dürften. Auch auf der Ebene der Schulen ist nur ein geringer Selbstselektionseffekt zu erwarten, eben diese anhand einer Zufallsliste gezogen wurden, und da sich die Ergebnisse von Kontroll- und Experimentalgruppe weitgehend ähneln.

Bei dem Design der Untersuchungsstufe II handelt es sich um einen Ein-Gruppen-Prätest-Posttestplan (vgl. Bortz, Döring 1995, S. 520 f.), dessen Design der folgenden Tabelle zu entnehmen ist:

		Prätest	t_0	$t_{+ 2 \frac{1}{4} \text{ Jahre}}$ (Posttest)
3. Klasse	Experimentalgruppe II	Prätestergebnisse der <i>Untersuchungsstufe I</i> <i>siehe Tabelle 2.2.1</i>	treatment	Posttest
4. Klasse	Experimentalgruppe II	Prätestergebnisse der <i>Untersuchungsstufe I</i> <i>siehe Tabelle 2.2.1</i>	treatment	Posttest

Tabelle 2.2.2: Das Untersuchungsdesign;
- das quasiexperimentelle Design der Untersuchungsstufe II
Der Verfasser

Das Design wechselt an dieser Stelle von einem längsschnittlichen in ein querschnittliches Vorgehen. 27 Monate nach der Teilnahme an der MINIPHÄNOMENTA werden Schülerinnen und Schüler der dritten und vierten Klassenstufe untersucht. Die Ergebnisse werden mit denen des Prätests der Untersuchungsstufe I verglichen. Der Leserin oder dem Leser fallen die zahlreichen methodischen Probleme dieses Designs unmittelbar ins Auge. Hier die wichtigsten Beispiele:

1. Da kein eigener Prätest durchgeführt wird und auch keine Kontrollgruppe existiert, kommt dem Design eine geringe interne Validität zu.
2. Deutliche Probleme beim zentralen Kausalitätsschluss, da die Vielzahl von konfundierenden Variablen in einem Zeitraum von über zwei Jahren nicht zu kontrollieren ist.
3. Der Wechsel von einem längsschnittlichen in ein querschnittliches Design führt zu unterschiedlichen Zeitpunkten der Teilnahme am treatment in der Experimentalgruppe II und dem Prätest der Untersuchungsstufe I:
 - Experimentalgruppe II: Die Schülerinnen und Schüler sind zum Zeitpunkt der Untersuchung in den Klassenstufen drei und vier. Bedenkt man den genannten zeitlichen Abstand zum treatment, waren sie bei ihrer Teilnahme an dem Projekt in der ersten beziehungsweise zweiten Klasse.
 - Prätest der Untersuchungsstufe I: Die Schülerinnen und Schüler dieser Untersuchungsgruppe sind zum Zeitpunkt ihrer Teilnahme an der MINIPHÄNOMENTA bereits in der dritten und vierten Klassenstufe.

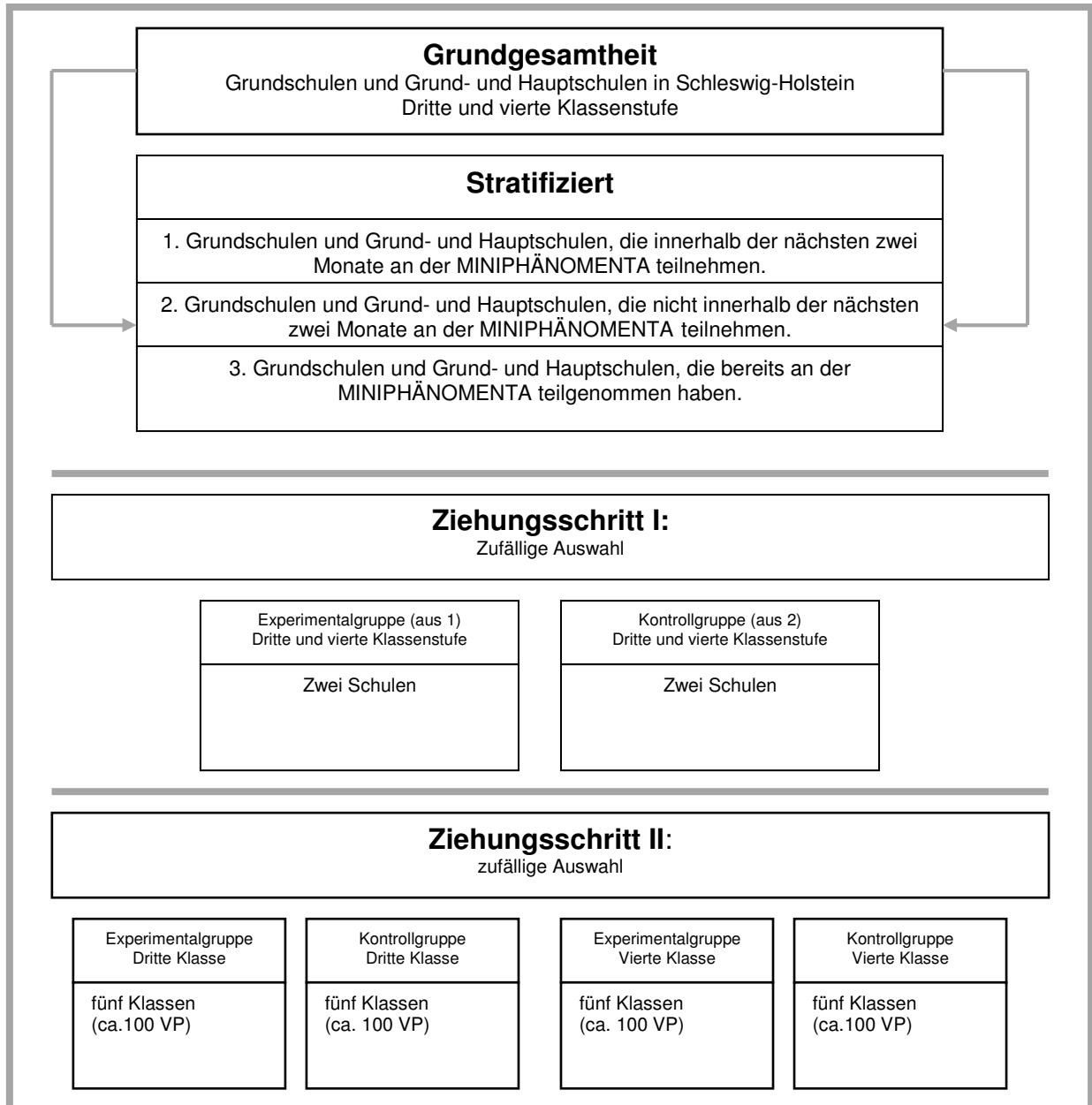
Die unterschiedlichen Zeitpunkte der Teilnahme an der MINIPHÄNOMENTA resultieren aus dem Design der Untersuchungsstufe II, schränken aber die Validität der erzielten Ergebnisse ein.

Dennoch habe ich mich zur Aufnahme dieser Untersuchungsstufe in die Arbeit entschieden. Auch wenn die Ergebnisse nicht die gleiche Validität wie die der Untersuchungsstufe I besitzen, so lassen sich doch erste Ansätze zur Langzeitwirksamkeit über sehr lange Zeiträume gewinnen, die in dieser Studie als Ausblick verstanden werden sollen.

2.3 Die Stichprobe

2.3.1 Stichprobenziehung in der Untersuchungsstufe I

Die folgende Grafik veranschaulicht den Prozess der Stichprobenziehung für die Untersuchungsstufe I:



Grafik 2.3.1.1: Die Stichprobe;
- Ziehung der Stichprobe in der Untersuchungsstufe I
Der Verfasser

Es handelt sich um eine komplexe Stichprobe deren wesentliche Kennzeichen im folgenden Textverlauf erörtert werden sollen: Zunächst ist die Grundgesamtheit in Bezug auf das Merkmal Teilnahme an der MINIPHÄNOMENTA in drei Schichten stratifiziert (vgl. Lohr 1999, S. 95 ff.). Aus den Schichten wird in zwei zufallsorientierten Ziehungsschritten eine Klumpenstichprobe gezogen. Nach Bortz und Döring (1995, S. 376 ff.) ist sie dadurch definiert, dass die Gesamtpopulation aus verschiedenen Klumpen, hier Schulen oder Schulklassen, besteht. Gezogen werden damit keine einzelnen Versuchspersonen sondern vielmehr ganze Schulen beziehungsweise Klassen.

Ziehungsschritt I: Aus den Schichten eins und zwei werden zufallsorientiert vier Schulen gezogen: Dabei werden zwei Schulen aus der ersten Schicht ermittelt. Sie stellen die Basis der Experimentalgruppe dar. Zwei weitere Schulen werden aus der zweiten Schicht für die Kontrollgruppe ermittelt

Ziehungsschritt II: Aus den Schulen werden in einem zweiten erneut zufallsorientierten Ziehungsschritt Klassen gezogen. Es ergeben sich für die dritte und vierte Klassenstufe eine Kontroll- und eine Experimentalgruppe mit jeweils fünf Klassen und damit etwa 100 Untersuchungsteilnehmern je Klasse und Gruppe.

Aus Gründen der Anonymität ist bei allen Stichproben immer nur der Durchschnitt der Versuchspersonen von Bedeutung. Die Zuordnung von Individualergebnissen über unterschiedliche Messzeitpunkte hinweg, zum Beispiel über die Vergabe von Kennziffern, ist damit nicht möglich. Es handelt sich damit in Bezug auf die unterschiedlichen Messzeitpunkte um unabhängige Stichproben (vgl. Bortz 1995, S. 86 ff.)

Es handelt sich um eine komplexe Stichprobe. Sie basiert auf einer stratifizierten Grundgesamtheit aus der anhand von Zufallskriterien eine zweistufige Klumpenstichprobe ermittelt wird.

Für die folgenden Variablen müssen in Bezug auf den Stichprobenumfang Einschränkungen gemacht werden:

1. [Formale Kompetenz]
2. [Allgemeines Experimentierverhalten]
3. [Wissen]

Die beiden Erstgenannten werden durch eine Kombination von Beobachtungs- und Interviewverfahren erhoben. Im Rahmen einer Schulstunde, die für die Erhebung beider Variablen zur Verfügung steht, ist es nicht möglich, alle Schülerinnen und Schüler einer Klasse zu erfassen. Es konnten folgende Durchschnittswerte erzielt werden:

1. [Formale Kompetenz]: 7 Erhebungen je Schulklasse
2. [Allgemeines Experimentierverhalten]: 8 Erhebungen je Schulklasse

Die untersuchten Schülerinnen und Schüler werden zufällig bestimmt. In Bezug auf diese beiden Variablen ergibt sich damit ein dritter, zufallsorientierter Ziehungsschritt.

Für die Variablen [formale Kompetenz] und [allgemeines Experimentierverhalten] ist aus forschungspraktischen Gründen ein dritter zufallsorientierter Ziehungsschritt erforderlich.

Die Erhebung des [Wissens] erfolgt nicht, wie die Bearbeitung der Fragebögen in Einzelarbeit, sondern in Gruppen von jeweils zwei bis drei Personen. Dabei findet die Gruppe eine gemeinsame Lösung für die Aufgabe. Die Gruppen werden durch ein Losverfahren gebildet. Daher liegen auch für diese Variable weniger Ergebnisse als Versuchspersonen vor.

Die Bearbeitung der Variable [Wissen] erfolgt in Kleingruppenarbeit. Die Gruppen werden anhand von Zufallskriterien gebildet.

2.3.2 Kennwerte der Experimentalgruppe I und der Kontrollgruppe

Entsprechend dem Stichprobenplan sind diesen Gruppen für die dritte und vierte Klassenstufe jeweils fünf Klassen zugeordnet. Die Untersuchung fand an Schulen in Schleswig-Holstein statt, deren Nennung Gründe der Anonymität verbieten. Der folgenden Tabelle sind die zentralen Werte der Stichprobe der Experimentalgruppe I²⁴ zu entnehmen:

	Dritte Klasse		Vierte Klasse	
Gesamtzahl der Versuchspersonen	107		110	
Geschlechterverteilung	weiblich: 46	männlich: 61	weiblich: 61	männlich: 49
Durchschnittliches Alter	8,3		9,7	
Formale Kompetenz (Anzahl der Erhebungen)	36		39	
Allgemeines Experimentierverhalten (Anzahl der Erhebungen)	40		37	
Wissen (Anzahl der Erhebungen)	52		51	

Tabelle 2.3.2.1: Die Stichprobe;
- Stichprobenwerte der Experimentalgruppe I
Der Verfasser

Es folgt die Auflistung der Werte aus der Kontrollgruppe:

	Dritte Klasse		Vierte Klasse	
Gesamtzahl der Versuchspersonen	104		115	
Geschlechterverteilung	weiblich: 55	männlich: 49	weiblich: 59	männlich: 56
Durchschnittliches Alter	8,5		9,9	
Formale Kompetenz (Anzahl der Erhebungen)	38		42	
Allgemeines Experimentierverhalten (Anzahl der Erhebungen)	62		69	
Wissen (Anzahl der Erhebungen)	53		54	

Tabelle 2.3.2.2: Die Stichprobe;
- Stichprobenwerte der Kontrollgruppe
Der Verfasser

²⁴ Die Werte basieren auf der Erhebung des Prätests. Aufgrund von Klassenwechselln, Krankheit usw. ist es in den folgenden Tests zu kleineren Abweichungen gekommen. Diese Einschränkung gilt ebenfalls für die Kontroll- und die Experimentalgruppe II.

Betrachten wir hier zunächst die Vergleichbarkeit der Geschlechts- und Altersverteilungen in der Experimentalgruppe I und der Kontrollgruppe:

	Geschlecht		Durchschnittsalter
	weiblich	männlich	
Experimentalgruppe I dritte und vierte Klassenstufe insgesamt	107 (49%)	110 (51%)	9,0
Kontrollgruppe dritte und vierte Klassenstufe insgesamt	113 (52%)	104 (48%)	9,2

Tabelle 2.3.2.3: Die Stichprobe;
- zentrale Stichprobenparameter der beiden Gruppen im Vergleich
Der Verfasser

Die Differenzen zwischen den unterschiedlichen Werten sind gering. Folglich ist die Vergleichbarkeit der Experimentalgruppe I und der Kontrollgruppe im Hinblick auf die Untersuchungsmerkmale gegeben.

Die Kontrollgruppe und die Experimentalgruppe I sind in Bezug auf die Basiswerte Alter und Geschlecht vergleichbar.

In Bezug auf die Variable [allgemeines Experimentierverhalten] muss der sprunghafte Anstieg der gemachten Erhebungen in der Kontrollgruppe erklärt werden: Er kommt dadurch zustande, dass in dieser Gruppe Studentinnen und Studenten im Rahmen eines Seminars mit in die Erhebung eingebunden werden konnten. Mit der Zahl der Erhebenden steigt hier zwangsläufig auch die Anzahl der erhobenen Daten.

2.3.3 Stichprobenziehung in der Untersuchungsstufe II

Die Stichprobenziehung ist in dieser Untersuchungsstufe mit der Untersuchungsstufe I identisch. Auch hier liegt eine stratifizierte Grundgesamtheit vor, aus der in zwei Ziehungsschritten eine Klumpenstichprobe ermittelt wird. Der Unterschied besteht lediglich im Umfang der Stichprobe. Hier wird eine Grund- oder Grund- und Hauptschule ermittelt und es werden aus ihr nur zwei dritte und vierte Klassen (jeweils etwa 50 Versuchspersonen) untersucht.

Die Ziehung der Stichprobe der Untersuchungsstufe II ist der der Untersuchungsstufe I identisch.

Hier sei nochmals auf den Abschnitt 2.2 in dieser Arbeit verwiesen. Im Falle der Untersuchungsstufe II wird keine separate Kontrollgruppe ermittelt. Vielmehr werden die hier erhobenen Daten mit denen des Prätests der Untersuchungsstufe I verglichen.

2.3.4 Die Experimentalgruppe II

Die Experimentalgruppe II besteht aus deutlich weniger Versuchspersonen als die Gruppen der Untersuchungsstufe I. Der Stichprobenplan sieht die Ziehung von jeweils zwei dritten und vierten Klassen vor. Die geringere Anzahl von Versuchspersonen ist dabei vor allem dem Umstand geschuldet, dass die Untersuchungsstufe II einen Ergebnisausblick in der vorliegenden Arbeit darstellt. Der eigentliche Hauptteil der Untersuchung besteht aus dem längsschnittlichen Design der Untersuchungsstufe I.

Die Tabelle gibt erneut die wichtigsten Stichprobendaten wieder:

	Dritte Klasse		Vierte Klasse	
Gesamtzahl der Versuchspersonen	44		46	
Geschlechterverteilung	weiblich:22	männlich: 22	weiblich: 24	männlich:22
Durchschnittliches Alter	8,4		9,4	
Formale Kompetenz (Anzahl der Erhebungen)	15		13	
Lerntiefe (Anzahl der Erhebungen)	21		23	
Wissen (Anzahl der Erhebungen)	19		20	

Tabelle 2.3.4.1: Die Stichprobe;
- zentrale Kennwerte der Stichprobe
Der Verfasser

Die zentralen Werte werden an dieser Stelle, wie in Abschnitt 2.3.2 in dieser Arbeit, mit denen der Experimentalgruppe I und der Kontrollgruppe verglichen. Dies erfolgt anhand der folgenden Tabelle:

	Geschlecht		Durchschnittsalter
	weiblich	männlich	
Experimentalgruppe I dritte und vierte Klassenstufe insgesamt	107 (49%)	110 (51%)	9,0
Kontrollgruppe dritte und vierte Klassenstufe insgesamt	113 (52%)	104 (48%)	9,2
Experimentalgruppe II dritte und vierte Klassenstufe insgesamt	46 (51%)	44 (49%)	8,9

Tabelle 2.3.4.2: Die Stichprobe;
- Vergleich der Stichprobenwerte der Untersuchungsstufe I und II
Der Verfasser

Die Differenzen sind auch hier gering. Die Gruppen können in Bezug auf die untersuchten Merkmale verglichen werden.

Die drei Untersuchungsgruppen sind in Bezug auf die Basiswerte Alter und Geschlecht vergleichbar.

2.3.5 Repräsentativitätsprobleme

Nach der Einschätzung von Wottawa und Thierau (2003, S. 36 ff.) unterscheiden sich „sozialwissenschaftliche Grundlagenforschung“ und evaluative Vorhaben unter anderem durch das Kriterium der Praktikabilität. Orientiert sich der erstgenannte Forschungszweig möglichst eng am Ziel der Wissenschaftlichkeit, kommt praktischen Gesichtspunkten bei Evaluationen ein deutlich größeres Gewicht zu.

Von dieser Einschätzung ausgehend, habe ich mich zur Ziehung einer Klumpenstichprobe entschieden. Diese ist im schulischen Kontext überaus ökonomisch: Hierdurch wird es möglich ganze Schulen beziehungsweise Klassen anstatt einzelner Versuchspersonen zu ziehen. Bedingt durch unterschiedliche Ziehungswahrscheinlichkeit vergrößert sich bei dieser Stichprobenart in Relation zur einfachen Zufallsstichprobe die Dispersion der Messwerte. Lohr (1999, S.145 f.) beziehungsweise Bortz und Döring (1995, S. 415 ff.) geben Berechnungsvorschriften zur exakten Ermittlung der Standardabweichung bei dieser Stichprobenart an.

Die Stratifizierung der Grundgesamtheit, die sofern sie in eine Gesamtstichprobe übernommen wird, zu einer Verkleinerung der Dispersionsmaße führt (vgl. Bortz 1995, S. 92), ist hier ohne Bedeutung. Dies ist darin begründet, dass die Mitglieder der einzelnen Strata auch unterschiedlichen Stichprobensubgruppen (Experimental- und Kontrollgruppe) zugewiesen werden.

Im Rahmen dieser Arbeit werde ich die zentralen deskriptiven Kennwerte mit den üblichen elementarstatistischen Verfahren bestimmen (vgl. Bortz 1995, S. 17 ff.). In Bezug auf die Dispersionsmaße ergibt sich durch dieses Vorgehen eine Ungenauigkeit, die sich letztendlich als Repräsentativitätsproblem der Stichprobe auswirkt.

Betrachtet man das Ziel der Untersuchung, so stellt man fest, dass die Unterschätzung der Standardabweichung vor allem aus zwei Gründen nicht zu größeren Validitätsverlusten der Ergebnisse der Studie führt:

1. Bei der vorliegenden Studie handelt es sich, mit Ausnahme der Schilderung der Ergebnisse des Prätests, weniger um populationsbeschreibende Untersuchung, bei der die Schätzung des Standardfehlers des Mittelwertes in der Population von zentraler Bedeutung wäre. Die ausführliche Analyse der Prätestdaten stellt hier lediglich einen Exkurs dar. Vielmehr besteht das zentrale Ziel des Prätests in der Gewinnung von Referenzdaten, um Aussagen über die Wirksamkeit des treatments, die MINIPHÄNOMENTA, machen zu können. Werden die Ergebnisse des Prätests inhaltlich als ein Indikator für eine geringe Physikalische Bildung der Schülerinnen und Schüler interpretiert, muss allerdings die Ungenauigkeit des Schätzers der Standardabweichung und des Konfidenzintervalls bedacht werden.

2. Aufgrund der besonderen Beschaffenheit der einzelnen Cluster treten Klumpungseffekte (vgl. Bortz 1995, S. 86 f.) hier, so die vorläufige These, nicht auf. Die Begründung hierfür liegt in der Zusammensetzung der unterschiedlichen Klassen (Klumpen) durch die einzelnen Schulleiterinnen und Schulleiter. Sie sind angehalten in Bezug auf eine Vielzahl von Merkmalen insbesondere das Alter, das Geschlecht und die Leistungsfähigkeit der Schülerinnen und Schüler möglichst heterogene Klassen zu bilden. Diese Hypothese wurde durch die beispielhafte Bestimmung einiger Intra-Class-Coeffizienten (ICC's) (vgl. Bortz, Döring 1995, S. 416 ff.) überprüft. Dieses Verfahren basiert auf einem varianzanalytischen Ansatz. Der ICC ist dabei definiert als der Quotient aus systematischer Varianz und der Gesamtvarianz, wobei die systematische Varianz die Differenz zwischen der Varianz innerhalb der Klumpen und der Varianz zwischen den Klumpen beschreibt. Mit Standardmethoden kann gerechnet werden, sofern die errechneten Alpha-Fehlerwahrscheinlichkeiten nicht zu nahe am Signifikanzkriterium liegen. Die p-values lagen in allen Fällen ausreichend weit von $\alpha = 5\%$ entfernt, sodass trotz der Cluster-Stichprobe mit Standardmethoden gerechnet werden kann.

2.4 Die Methoden und Untersuchungsinstrumente

2.4.1 Einleitung

Nach der Darstellung der theoretischen Aspekte und, soweit vorhanden, einer Einführung in den empirischen Forschungsstand der in dieser Arbeit erfassten Variablen, sollen in diesem Abschnitt die Untersuchungsinstrumente entwickelt und diskutiert werden. Jeder Abschnitt zu 2.4 enthält dabei Informationen zu folgenden Aspekten:

- Entwicklung des Messverfahrens
- Konkrete Anwendung des Instrumentes in der Untersuchungssituation
- Itemrevision
- Weitere quantitative und qualitative Untersuchungsmethoden, vor allem zur Überprüfung der den Messverfahren zu Grunde liegenden theoretischen Konstrukte
- Ermittlung der Gütekriterien

Bei der Bestimmung der Gütekriterien wurde auf eine empirische Erhebung der Validität verzichtet. Dies wurde vor allem unterlassen, um den Untersuchungsaufwand in einem angemessenen Rahmen halten zu können: Einfache empirische Validierungsverfahren ermitteln die Validität zumeist als Kriteriumsvalidität (vgl. Bortz, Döring 1995, S. 185ff.), mit deren Hilfe, durch die Korrelation mit einem Außenkriterium, festgestellt wird, ob ein Messverfahren tatsächlich das misst, was es zu messen vorgibt. Im Falle der vorliegenden Untersuchung könnte eine solche Form der Validierung lediglich durch ein weiteres zu entwickelndes Verfahren nachgewiesen werden. Dies würde zum einen den Untersuchungsaufwand enorm steigern, zum anderen ist nach dem Aussagegehalt der resultierenden Ergebnisse zu fragen.

Was sagt also eine hohe positive Korrelation zwischen Kriteriumswert und dem durch das Instrument ermittelten Wert letztendlich aus? Lässt es sich so zweifelsfrei nachweisen, dass das entwickelte Instrument tatsächlich das misst was es zu messen vorgibt? Aus logischer Sicht ist dieser Gedankengang keineswegs zwingend. Es ist schließlich auch denkbar, dass beide Instrumente übereinstimmend etwas anderes als das Zielkonstrukt messen. Auch ist es möglich, dass beide Verfahren etwas Unterschiedliches messen, die Ausprägungen auf den Variablen aber ähnlich sind.

Die sichere empirische Bestimmung der Validität ist forschungsmethodisch nur äußerst schwer umzusetzen. Ein Beispiel wäre das sehr sichere aber auch hochkomplexe Verfahren der multi-trait-multi-method-Validierung (vgl. Bortz, Döring 1995, S. 187 ff.). Eine derartige Methode der empirischen Validierung kann hier aus Umfangsgründen nicht eingesetzt werden. Um die logische Stringenz der Arbeit zu wahren, sollen die oben angesprochenen einfachen Methoden der empirischen Validierung nicht zum Einsatz kommen. Sie erzeugen vielmehr einen Eindruck von Validität, der bei genauerer Betrachtung nicht aufrechterhalten werden kann.

Der Nachweis der Validität der verwendeten Verfahren geschieht in dieser Arbeit nicht durch empirische Schritte, sondern argumentativ durch den Ansatz der logischen Validität (vgl. Bortz, Döring 1995, S. 185). Diese ist gegeben wenn das Messinstrument die zugrunde liegende Theorie umfassend und exakt operational definiert.

Um die Übersichtlichkeit der folgenden Textabschnitte zu steigern, soll an dieser Stelle zusammenfassend das Vorgehen der Untersuchung dargestellt werden. Die unten stehende Tabelle zeigt in detaillierter Form den Ablauf einer einzelnen Untersuchung. Ihr folgend fasse ich die einzelnen Schwerpunkte der Untersuchung im weiteren Textverlauf zusammen:

Stunde	Zeitbedarf	Gegenstand
I	5 Minuten	Vorstellung und Begrüßung
I	5 Minuten	Einleitung in die Thematik
I	5 Minuten	Erläuterung des Fragebogens zur Erfassung der Einstellung anhand einer Folie im Unterrichtsgespräch
I	15 Minuten	Die Schülerinnen und Schüler bearbeiten den Fragebogen in Stillarbeit
I	15 Minuten	Einführung in das Experimentieren: Forscherregeln (Vorstellung anhand einer Folie)
II	45 Minuten	freies Experimentieren: Ermittlung der formalen Kompetenz, sowie des allgemeinen Experimentierverhaltens anhand von elf im Klassenraum aufgebauten Versuchen
III	5 Minuten	freies Experimentieren mit dem Experiment zur Ermittlung der Motivation und des Wissens
III	5 Minuten	Erläuterung des Fragebogens zur Erfassung der aktuellen Motivation anhand einer Folie im Unterrichtsgespräch
III	10 Minuten	Die Schülerinnen und Schüler bearbeiten den Fragebogen in Stillarbeit
III	10 Minuten	Vorstellung des Prinzip des Mind-Maps im Unterrichtsgespräch an einem Beispiel
III	15 Minuten	Erstellung der Mind-Maps durch die Schülerinnen und Schüler in Kleingruppenarbeit

Tabelle 2.4.1.1: Einleitung in die Methoden und Untersuchungsinstrumente;
- Ablauf der Untersuchung in den Schulen²⁵
Der Verfasser

²⁵ Hierbei handelt es sich um den Ablauf der Prätestuntersuchung. Im Falle der Experimentalgruppe II der Untersuchungsstufe II und der Posttestuntersuchungen der Untersuchungsstufe I kommt in der ersten Stunde der Fragebogen zur Erfassung der [deskriptiven Daten] hinzu. Weiter in der ersten Stunde zu behandelnde Punkte werden entsprechend gekürzt.

Der Schwerpunkt der ersten Stunde besteht in der Einführung der Schülerinnen und Schüler in die Untersuchungssituation und in der Erhebung der [Einstellung]. In der folgenden Stunde sollen die Schülerinnen und Schüler mit elf in der Klasse aufgestellten Versuchen experimentieren. In diesem Zusammenhang werden das [allgemeine Experimentierverhalten]²⁶ und die [formale Kompetenz] erhoben. Die dritte Untersuchungsstunde dient der Erfassung des [Wissens] und der Ermittlung der [Motivation zur tiefergehenden kognitiven Bearbeitung physikalischer Fragestellungen].

Alle für die folgende Itemrevision notwendigen statistischen Berechnungen basieren, sofern nicht anders ausgewiesen, auf der Stichprobe der Voruntersuchung, die im Vorfeld der Hauptuntersuchung gezogen und untersucht wurde.

²⁶ Für die praktische Untersuchungsdurchführung und insbesondere für die notwendigen Beobachtungen in Bezug auf die Variablen [allgemeines Experimentierverhalten] und [formale Kompetenz] konnten in großem Umfang Studierende der Universität Flensburg gewonnen werden. Diese wurden in einer gründlichen Schulung in den Untersuchungsablauf und vor allem in die Methode des Beobachtens eingeführt. Neben den Beobachtungsleitfäden standen hier mögliche Verzerrungstendenzen im Vordergrund.

2.4.2 Fragebogen zur Erfassung der Einstellung gegenüber Phänomenen und ihrer Bearbeitung

In Anlehnung an die Untersuchung der Einstellung gegenüber der Physik von Dengler (1995, S. 27ff.) ließe sich auch für den hier zu verhandelnden Gegenstandsbereich ein semantisches Differential zur [Erfassung der Einstellung gegenüber Phänomenen und ihrer Bearbeitung] entwickeln.

Das Verfahren bringt gerade für den zu untersuchenden Stichprobenraum, die dritte und vierte Klassenstufe, entscheidende Vorteile mit sich: Bei der Bearbeitung des Fragebogens müssen die Schülerinnen und Schüler nur sehr wenig lesen. Gerade bei den genannten Klassenstufen ist dies ein deutlicher Vorteil, da sie im Lesen im Vergleich zu Erwachsenen noch ungeübt sind. Durch die polare Konstruktion können weiterhin umfangreiche Themenbereiche mit sehr wenigen Items beschrieben werden. Dies setzt erneut den Umfang herab, was im Hinblick auf die Stichprobe ebenfalls sinnvoll erscheint.

Bei genauerer Betrachtung weist das Verfahren jedoch eine Vielzahl von Nachteilen auf, die letztendlich gegen die Verwendung der Methode im klassischen Sinn in dieser Arbeit sprechen.

Bortz und Döring definieren den Begriff des semantischen Differentials als eine „[...] Datenerhebungsmethode, die die *konnotative* Bedeutung von Begriffen oder Objekten mit Hilfe eines Satzes von 20 bis 30 bipolaren Adjektivpaaren erfasst [...]“ (ebd. 1995, S. 173, Hervorhebung nicht im Original). Geht man von dieser Definition aus, so lässt sich der ‚kognitive und verhaltensbezogene Anteil‘ einer Einstellung, der sich aus dem Drei-Komponenten-Modell neben dem ‚affektiven Aspekt‘ ergibt, mit dieser Methode nicht adäquat messen. Daher erscheint das Verfahren, im Hinblick auf die zugrunde liegende Theorie, unpassend.

Weiterhin lassen sich, besonders für den ‚verhaltensbezogenen Anteil der Einstellung‘, die Items häufig nicht mit Adjektiven beschreiben. Es müsste an dieser Stelle auf kurze andere Formulierungen, wie zum Beispiel „einfach drauf los“ versus „geplantes Vorgehen“ zurückgegriffen werden. Die Wahl von Adjektiven wird zusätzlich durch den geringeren Wortschatz von Kindern in der dritten und vierten Klassenstufe in Relation zu dem von Erwachsenen erschwert. So ließe sich zum Beispiel die Formulierung des geplanten Vorgehens durch das Adjektiv „hypothesenprüfend“ ersetzen. Diese Wortwahl würde die Schülerinnen und Schüler selbstverständlich überfordern und muss daher unterlassen unterbleiben.

Oft lassen sich des Weiteren keine klaren Gegensatzpaare aufbauen. Wollte man zum Beispiel der Frage nachgehen, ob die Schülerinnen und Schüler den Umgang mit den interaktiven Stationen als Lernvorgang oder als Spiel sehen, so brächte man die Begriffe des Spielens und des Lernens in Widerspruch: Ein Gedanke der gerade den Erörterungen im Abschnitt 1.6 widersprechen würde.

Weiterhin muss an dieser Stelle auf Probleme bei der Polung eines semantischen Differentials hingewiesen werden: Greifen wir hierzu wieder das im obigen Absatz genannte Beispiel vom „Spielen versus Lernen“ auf. Was zeigt in diesem Umgang eine positive [Einstellung gegenüber den Phänomenen und ihrer Bearbeitung]? Ist es der spielerische Umgang oder das Lernen am Phänomen? Auch hier brächte man die Begriffe in einen unangemessenen Widerspruch.

Die vier Aspekte zeigen deutlich, dass das semantische Differential in seiner klassischen Form zwar den entscheidenden Vorteil der Ökonomie besitzt, für diesen Zusammenhang der Einstellungsmessung jedoch nicht verwendet werden kann. Ein klassischer Einstellungsfragebogen (vgl. Mummendey 1995, S. 27 ff.) mit Behauptungen die auf einer Skala eingeschätzt werden können, löst zwar viele der geschilderten Probleme, bringt aber die Schwierigkeit mit sich, dass die Schülerinnen und Schüler sehr viel lesen müssten, was wie bereits erwähnt, von Kindern der dritten und vierten Klassenstufe nicht erwartet werden kann.

In dieser Arbeit sollen daher die spezifischen Vorteile beider Instrumente zusammengeführt werden, um ein ökonomisches und theoretisch tragfähiges Forschungsinstrument zu erhalten: Der Fragebogen ist als Tabelle konstruiert. Über der Tabelle befindet sich der Satzanfang „Experimentieren ist für mich...“. In den einzelnen Zeilen der Tabelle befinden sich die Items. Sie stellen jeweils eine Vervollständigung des genannten Satzes dar und bestehen nur aus sehr wenigen Worten. Die Schülerinnen und Schüler können ihre Zustimmung beziehungsweise Ablehnung auf einer fünfstufigen bildgestützten Skala angeben.

Der Fragebogen besteht in der Prätestversion aus insgesamt 20 Items²⁷. Diese enthalten nur wenige Stichwörter. Die Schülerinnen und Schüler sollen ihre Einstellung zum jeweiligen Item auf einer fünfstufigen bildgestützten Skala angeben. Die Stichwörter sind wie folgt den Einstellungskomponenten zugeordnet:

²⁷ Der Prätestfragebogen ist dem Abschnitt 4.2 in dieser Arbeit zu entnehmen.

Experimentieren ist für mich...

Affektiv	Kognitiv	Handlungsbezogen
selber machen	wichtig	auch eine Freizeitbeschäftigung
spannend	in der Gruppe arbeiten	leicht
interessant	selber denken	spielen
Spaß	Ideen ausprobieren	vorsichtig sein
genau beobachten	etwas für Jungen und für Mädchen	alles anfassen
abwechslungsreich		einfach so loslegen
Begeisterung		eine Herausforderung
lernen		

Tabelle 2.4.2.1: Einstellung gegenüber Phänomenen und ihrer Bearbeitung;
- vorläufige Zuordnung der Items zu den Dimensionen des Fragebogens
Der Verfasser

Die Zuweisung der Items zu den Kategorien erfolgt durch die mündliche Befragung von 30 Schülerinnen und Schülern der dritten und vierten Klassenstufe. Auf den ersten Blick erscheinen einige Zuordnungen nicht zwingend evident. Die Argumentation der Kinder war in diesen Fällen jedoch immer schlüssig. Dies sei am Beispiel der Zuordnung des Items „genau beobachten“ zur ‚affektiven Dimension‘ gezeigt: Sie empfinden es als anregend, das Experiment zu betrachten, wollen sehen was als nächstes passiert. Es baut sich im Prozess des Experimentierens ein Spannungsbogen auf, der durch das genaue Beobachten moderiert wird. Insofern ist diese Zuordnung aus dieser Perspektive zur ‚affektiven Dimension‘ durchaus schlüssig.

Nach einer Einführung in den Fragebogen anhand eines Unterrichtsgespräches sollen die Schülerinnen und Schüler ihre Zustimmung oder Ablehnung zu den Aussagen angeben. Alle Items sind, um die Auswertung zu erleichtern und den Fragebogen für die Kinder möglichst übersichtlich zu gestalten, positiv gepolt.

Im Zuge der Itemrevision des Fragebogens wurden für jedes Item die Schwierigkeit und die Trennschärfe²⁸ berechnet. In Bezug auf die cut-off-scores orientiere ich mich an den Vorgaben von Bortz und Döring (1995, S. 199 ff.). Dabei sollte die Schwierigkeit im Wertebereich zwischen 0,3 und 0,8 liegen, die Trennschärfe größer gleich 0,3 sein. Die folgende Tabelle stellt die Ergebnisse der Itemrevision zusammen:

Nummer	Item	Codierung	Schwierigkeit	Trennschärfe
1	Spannend	A ₁	0,9	0,5
2	Interessant	A ₂	0,9	0,5
3	Selber machen	A ₃	0,5	0,5
4	Wichtig	K ₁	0,7	0,6
5	Auch eine Freizeitbeschäftigung	H ₁	0,6	0,7
6	Leicht	H ₂	0,7	0,1
7	In der Gruppe arbeiten	K ₂	0,7	0,3
8	Spielen	H ₃	0,7	0,3
9	Spaß	A ₄	0,9	0,4
10	Genau beobachten	A ₅	0,8	0,6
11	Lernen	A ₆	0,7	0,6
12	Vorsichtig sein	H ₄	0,9	0,4
13	Abwechslungsreich	A ₇	0,7	0,4
14	Alles anfassen	H ₅	0,6	0,2
15	Selber denken	K ₃	0,7	0,7
16	Einfach so loslegen	H ₆	0,6	0,1
17	Begeisterung	A ₈	0,8	0,6
18	Etwas für Jungen und für Mädchen	K ₄	0,8	0,3
19	Eine Herausforderung	H ₇	0,6	0,5
20	Ideen ausprobieren	K ₅	0,9	0,6

Tabelle 2.4.2.2: Einstellung gegenüber Phänomenen und ihrer Bearbeitung;
- Ergebnisse der Itemrevision
Der Verfasser

Entsprechend der genannten Kriterien habe ich mich zur Aussonderung von acht Items entschieden. Die durchschnittliche Schwierigkeit des Fragebogens liegt im mittleren Bereich bei 0,6 Punkten. Die Trennschärfe der einzelnen Items liegt mit 0,5 Punkten im guten bis sehr guten Bereich.

Die theoretische Grundlage des Fragebogens, das Drei-Komponenten-Modell und die Zuordnung der einzelnen Items zu den Einstellungskomponenten wurde durch eine konfirmatorische Faktoranalyse überprüft:

²⁸ Erläuterungen zu den Begriffen:

Die **Schwierigkeit** ist ein Index, der den Anteil der Versuchspersonen angibt, die das Item bejahen. Für polytome Items errechnet sich dieser Kennwert aus der Summe der Einzelpunktwerte dividiert durch das Produkt aus der maximal möglichen Punktzahl auf dem Item und der Anzahl der Versuchspersonen.

Die **Trennschärfe** gibt an wie gut das Einzelergebnis das Gesamtergebnis repräsentiert. Sie ist definiert als Produkt-Moment-Korrelation vom Punktwert des Items mit dem Gesamtpunktwert des Fragebogens.

Item	Codierung	Affektiv	Kognitiv	Verhaltensbezogen
Begeisterung	A ₁	0,7		
Genau Beobachten	A ₂	0,7		
Abwechslungsreich	A ₃	0,7		
Selber machen	A ₄	0,4		
Lernen	A ₅	0,7		
Etwas für Jungen und für Mädchen	K ₁		0,7	
Selber denken	K ₂		0,6	
Wichtig	K ₃		0,5	
In der Gruppe arbeiten	K ₄		0,5	
Auch eine Freizeitbeschäftigung	H ₁			0,6
Spielen	H ₂			0,8
Eine Herausforderung	H ₃			0,6

Tabelle 2.4.2.3: Einstellung gegenüber Phänomen und ihrer Bearbeitung;
- Ergebnisse der konfirmatorischen Faktoranalyse
Der Verfasser

Durch die Faktoranalyse wird mit 58% ein zufrieden stellender Teil der Gesamtvarianz aufgeklärt. Insgesamt kann damit festgestellt werden, dass das Drei-Komponenten-Modell der Einstellung und die Klassifizierung der Items durch die Schülerinnen und Schüler bestätigt wird.

Die Genauigkeit der Messung ist mit einer nach der „Spearman-Brown-Formula“ (vgl. Lienert 1969, S. 210) korrigierten split-half-Reliabilität nach den Kriterien von Bortz und Döring²⁹ (1995, S. 183 f.) mit 0,9 als hoch einzuschätzen.

Der Einstellungsfragebogen wurde auf der Basis des Drei-Komponenten-Modells entwickelt. Das theoretische Modell und die Einteilung der Schülerinnen und Schüler wurden durch eine konfirmatorische Faktoranalyse untersucht. Die Items laden erwartungsgemäß entsprechend der postulierten Struktur. Ferner liegen für jede Einstellungsdimension ausreichend geeignete Items vor. Insofern kann von einer ausreichenden logischen Validität ausgegangen werden.

Der überarbeitete Fragebogen ist dem Abschnitt 4.3 dieser Arbeit zu entnehmen.

Damit erfüllt der überarbeitete Fragebogen die gängigen Kriterien der empirischen Sozialforschung. Er kann folglich im weiteren Verlauf der Untersuchung in der dritten und vierten Klassenstufe eingesetzt werden.

²⁹Ein guter Test sollte demnach eine Reliabilität von etwa 0,8 aufweisen. Werte zwischen 0,8 und 0,9 gelten als durchschnittlich und Werte über 0,9 als hoch (vgl. Bortz, Döring 1995, S. 184 f.).

2.4.3 Fragebogen zur Erfassung deskriptiver Aspekte

An dieser Stelle soll nun die Untersuchungsmethodik zur Erhebung der [deskriptiven Aspekte] beschrieben werden. Es handelt sich um eine zweistufige Fragebogenuntersuchung.

Die Schülerinnen und Schüler der Experimentalgruppe erhalten einen Fragebogen (vgl. Abschnitt 4.1 in dieser Arbeit) mit zwei Aufgaben: Für die erste Aufgabe werden ihnen 15 zufällig ausgewählte Stationen aus der MINIPHÄNOMENTA gezeigt. Während der Darbietung der Fotos sollen sie angeben, ob sie sich an die Station erinnern können. Um auszuschließen, dass die Schülerinnen und Schüler bei jeder Frage mit Ja (Ja-Sage-Tendenz) antworten, folglich die Aufgabe nicht gewissenhaft erfüllen, werden 2 Kontrollfotos in die Bilderfolge eingestreut. In der zweiten Aufgabe sollen die Schülerinnen und Schüler ein Experiment beschreiben. Die Kinder werden gebeten sich an ihre Lieblingsstation zu erinnern. Diese sollen sie mit bis zu zehn Stichwörtern beschreiben und abschließend den Versuch skizzieren. Während die erste Aufgabe geschlossen ist, handelt es sich hier um eine offene Frage. Um zum einen die Vergleichbarkeit der Ergebnisse und zum anderen ähnliche Bearbeitungszeiten zu gewährleisten, orientiert sich die Bearbeitung der Aufgabe an 10 Stichworten. Die Auswertung der zweiten Aufgabe orientiert sich an einer quantitativen Inhaltsanalyse mit den folgenden Kategorien:

1. Anzahl der genannten Stichpunkte
2. Anzahl der sinnvollen Stichpunkte
3. Aufbau des Versuches wird verständlich
4. Funktion und Durchführung des Versuches werden verständlich
5. Effekt wird verständlich
6. Sinnvolle Erklärungsansätze
7. Affektive Färbung der Stichpunkte
8. Eidetischer Charakter der Erinnerungen

Entsprechend der genannten Kategorien wird die Leistung der Schülerinnen und Schüler anhand einer vierstufigen Skala (Ausnahme: Item 1 und 2) eingeschätzt.

Vor allem aus forschungsmethodischer Sicht wird bei dieser Fragestellung deutlich, dass der Vergleich mit einer Kontrollgruppe größere Fragen aufwirft als er zu lösen im Stande ist. Womit soll die Erinnerung an die Stationen verglichen werden? Möglich wäre der Vergleich mit Unterrichtsgegenständen. Hier ist jedoch einzuwenden, dass dies theoretisch, aufgrund der unterschiedlichen Zielvorstellungen der interaktiven Stationen und des konventionellem Unterrichtes, kaum möglich ist. Hinzu kommen die praktischen Probleme, denn mit welchem Unterrichtsgegenstand sollen die Exponate verglichen werden? Experimente beziehungsweise der Umgang mit ihnen kommen in der Primarstufe nur äußerst selten vor. Insofern entfällt dieser Vergleichsgegenstand. Außerdem werden auch gleiche Inhalte von Lehrerinnen und Lehren didaktisch unterschiedlich aufbereitet, so dass es in der Vergleichsgruppe zu einer enormen Heterogenität kommen würde. Insofern soll der Vergleich mit einer Kontrollgruppe bei dieser Variablen entfallen.

Eine quantitative oder qualitative Itemrevision des ersten Fragebogenteils erscheint nicht sinnvoll, da hier lediglich einfache deskriptive Daten erhoben werden. Die Kategorien des zweiten Teils sollen hingegen mit Ausnahme der Kategorien eins und zwei einer Revision auf mathematisch-statistischer Grundlage unterzogen werden.

Kategorie	Schwierigkeit	Trennschärfe
Anzahl der Stichpunkte	Keine Angaben	Keine Angaben
Anzahl der sinnvollen Stichpunkte	Keine Angaben	Keine Angaben
Aufbau wird verständlich	0,6	0,7
Funktion und Durchführung werden verständlich	0,6	0,5
Effekt wird verständlich	0,5	0,6
Anätze von Erklärungen sind vorhanden	0,4	0,5
Affektive Färbung der Stichpunkte	0,7	0,7
Eidetischer Charakter der Erinnerungen	0,7	0,8

Tabelle 2.4.3.1: Fragebogen zur Erfassung deskriptiver Aspekte;
- Ergebnisse der Itemrevision
Der Verfasser

Mit einer Schwierigkeit von 0,6 handelt es sich um eine durchschnittlich schwere Aufgabe. Die Trennschärfe der Items ist mit 0,6 im sehr hohen Bereich. Das Verfahren kann folglich für die Hauptuntersuchung verwendet werden.

Nun soll die Reliabilität der zweiten Aufgabe des Fragebogens bestimmt werden: Die Genauigkeit der Messung ist mit einer nach der „Spearman-Brown-Formel“ (vgl. Lienert 1969, S. 210) korrigierten split-half-Reliabilität nach den Kriterien von Bortz und Döring (1995, S. 184 f.) mit 0,8 als befriedigend einzuschätzen.

Der Fragebogen erfasst die Erinnerung der Schülerinnen und Schüler an die Stationen in Bezug auf ihren Umfang und die Qualität. Damit ist von einer ausreichend hohen logischen Validität auszugehen.

Damit erfüllt der überarbeitete Fragebogen die gängigen Kriterien der empirischen Sozialforschung. Er kann folglich im weiteren Verlauf der Untersuchung in der dritten und vierten Klassenstufe eingesetzt werden.

2.4.4 Kategorienschema zur Ermittlung der formalen Kompetenz

Zunächst sollen die Untersuchungsmethodik und das Erhebungsinstrument beschrieben werden. Die Datenerhebung zu dieser Variablen geschieht im Zusammenhang mit der Variablen [allgemeines Experimentierverhalten] (vgl. Abschnitt 1.4.5 und 2.4.5 in dieser Arbeit). Im Klassenraum werden insgesamt 11 Versuche aufgebaut. Dabei handelt es sich um 10 Versuche mit einfachen Mitteln, an denen das [allgemeine Experimentierverhalten] ermittelt wird, sowie das „Variopendel“³⁰, das der Bestimmung der [formalen Kompetenz] dient.

Die [formale Kompetenz] wird durch die Kombination einer Verhaltensbeobachtung mit einem Interview an einem konkreten Experiment erhoben. Die folgende Tabelle schildert den Ablauf der Untersuchung:

	Programmpunkt	Inhalt
1	Begrüßung	Die Schülerinnen und Schüler werden am Experiment begrüßt.
2	Aufgabenstellung mit Demonstration	Zunächst wird den Schülerinnen und Schülern die Aufgabe vorgestellt. Es folgt die Nennung und die Demonstration der drei zu manipulierenden Variablen.
3	Aufforderung zur Hypothesenbildung	Die Schülerinnen und Schüler werden aufgefordert Hypothesen zu bilden, wie die Aufgabe gelöst werden kann.
4	Schülerinnen und Schüler experimentieren	Die Schülerinnen und Schüler experimentieren bis zu dem Punkt an dem sie mit der Lösung der Aufgabe zufrieden sind. Dabei denken sie „laut“.
5	Ergebniszusammenfassung	Die Schülerinnen und Schüler werden gebeten zusammenzufassen, was Sie getan haben und warum sie es getan haben. Es folgt ein Fazit wie die Aufgabe gelöst werden konnte.
6	weitere Fragen	Es schließen sich die Fragen zum spezifisch - naturwissenschaftlichem Denken an. Die Schülerinnen und Schüler werden gebeten, sich Gedanken zu machen, warum man Experimente macht und wie Erklärungen für Experimente aussehen.

Tabelle 2.4.4.1: Formale Kompetenz;
- Ablauf der Untersuchung
Der Verfasser

³⁰ Das „Variopendel“ wird dabei lediglich beim ersten Untersuchungstermin eingesetzt. Bei den folgenden Untersuchungen werden ähnliche Versuche verwandt. Kriterien für die Feststellung von Ähnlichkeit waren:

1. Es muss eine konkrete für Kinder fassbare Aufgabe gelöst werden.
2. Drei klar erkennbare Variablen sind manipulierbar.

Vor diesem Hintergrund habe ich mich für folgende Experimente entschieden:

Prätest: „Variopendel“

Posttest I: „Linsensbilder“

Posttest II: „Flaschenrennen“

Die Untersuchung findet in einem separaten Raum statt. So erhalten die Schülerinnen und Schüler die nötige Ruhe zum Experimentieren und es kann verhindert werden, dass die Kinder bereits vor der Untersuchung Erfahrungen mit dem Experiment sammeln.

Das Beobachtungsschema (vgl. Abschnitt 4.5 in dieser Arbeit) basiert auf den im Abschnitt 1.4.4 diskutierten drei Teilaspekten der Variablen (L = ‚Logik‘; D = ‚naturwissenschaftliches Denken‘; E = ‚experimentelle Fähigkeiten‘). Es setzt sich in der Prätestversion aus insgesamt 15 Kategorien zusammen, auf denen die Ausprägung entsprechend der Kodieranweisung (vgl. 4.7 in dieser Arbeit) in der Beobachtungssituation eingeschätzt wird. Zusätzlich zu den drei Skalen (‚logische Kompetenzen‘, ‚experimentelle Kompetenzen‘, ‚spezifisch naturwissenschaftliche Denkweisen‘) und den Basisdaten werden die Dauer des Experimentierens, das Gruppenklima und die Konzentration auf die Aufgabe durch die Methode der Verhaltensbeobachtung erhoben.

Nach diesen methodisch-organisatorischen Aspekten sollen nun die Gütekriterien des Verfahrens bestimmt werden. Ich beginne mit der Überprüfung der von mir vorgeschlagenen Einteilung der Items auf drei Skalen sowie des zugrunde liegenden Drei-Komponenten-Modells. Dies soll durch eine konfirmatorische Faktoranalyse geschehen. Durch die Extraktion von drei Faktoren kann mit 75% ein hoher Anteil der Gesamtvarianz aufgeklärt werden. Die Items laden entsprechend der postulierten Struktur:

Nummer	Codierung	L	E	D
1	L ₁	0,8		
2	L ₃	0,8		
3	L ₂	0,8		
4	L ₄	0,7		
5	L ₅	0,7		
6	L ₆	0,6		
7	E ₄		0,9	
8	E ₅		0,9	
9	E ₆		0,7	
10	E ₂		0,7	
11	E ₇		0,7	
12	E ₁		0,6	
13	E ₃		0,6	
14	D ₁			0,9
15	D ₂			0,7
16	D ₃			0,7
17	D ₄			0,6

Tabelle 2.4.4.2: Formale Kompetenz;
- Ergebnisse der konfirmatorischen Faktoranalyse
Der Verfasser

Auch die Ergebnisse der Itemrevision bestätigen die hohe Güte des Verfahrens. Die durchschnittliche Schwierigkeit liegt mit 0,5 im mittleren Bereich. Die Trennschärfe der Items ist im Durchschnitt mit einem Wert von 0,7 als sehr hoch einzustufen. Keines der Items musste aufgrund der Revision ausgesondert werden. Die einzelnen Ergebnisse werden in der folgenden Tabelle wiedergegeben:

Nummer	Codierung	Schwierigkeit	Trennschärfe
1	L ₁	0,5	0,9
2	L ₂	0,5	0,8
3	L ₃	0,4	0,8
4	L ₄	0,4	0,7
5	L ₅	0,6	0,9
6	L ₆	0,6	0,8
7	E ₁	0,6	0,8
8	E ₂	0,6	0,7
9	E ₃	0,6	0,7
10	E ₄	0,4	0,7
11	E ₅	0,5	0,7
12	E ₆	0,4	0,8
13	E ₇	0,5	0,8
14	D ₁	0,3	0,4
15	D ₂	0,4	0,5
16	D ₃	0,5	0,5
17	D ₄	0,3	0,5

Tabelle 2.4.4.3: Formale Kompetenz;
- Ergebnisse der Itemrevision
Der Verfasser

Die Reliabilität des Verfahrens wurde mit dem cornbachs-Alpha-Index ermittelt und liegt mit einem Wert von 0,9 hoch.

Die logische Validität ist auch bei diesem Verfahren gegeben: Das theoretische Modell konnte durch eine konfirmatorische Faktoranalyse bestätigt werden und es liegen ausreichend geeignete Items für jede Dimension vor.

Das revidierte Verfahren ist dem Abschnitt 4.6 dieser Arbeit zu entnehmen.

Das Verfahren hat sich in der Praxis bewährt. Dies zeigte sich vorrangig durch seine überdurchschnittlich guten Kennwerte. Es kann folglich in der dritten und vierten Klassenstufe der Hauptuntersuchung verwendet werden.

2.4.5 Kategorienschema zur Ermittlung des allgemeinen Experimentierverhaltens

Zur methodisch sicheren Erfassung des [allgemeinen Experimentierverhaltens] schlage ich ein zweistufiges Verfahren vor:

1. Verhaltensbeobachtung
2. Kurzinterview

Beides geschieht im Rahmen des Experimentierfeldes: Insgesamt 11 Versuche stehen den Schülerinnen und Schülern zur Verfügung. An 10 Versuchen können sie völlig selbstständig und frei experimentieren. Der verbleibende Versuch ist der Ermittlung der [formalen Kompetenz] vorbehalten (vgl. Abschnitt 2.4.4 in dieser Arbeit). Die zum freien Experimentieren ausgewählten Versuche wurden nach den Kriterien von Schließmann (2005, S. 42 ff.) ausgewählt. Eine vollständige Liste aller Versuche ist im Anhang unter Punkt 4.11 wiedergegeben.

Ad 1: Verhaltensbeobachtung: Entsprechend der im Abschnitt 1.5.5 erörterten Ergänzungen schlage ich das folgende Kategorienschema vor, wobei eine genaue Kodieranweisung dem Abschnitt 4.10 in dieser Arbeit zu entnehmen ist:

Lernverhalten	Klassifikationskategorie
<ol style="list-style-type: none"> 1. Geht nach kurzem Verweilen weiter 2. Bricht das Experiment in einem frühen Stadium ab 3. Zeigt deutliche Ablehnung gegen das Experiment 	I. ablehnende Reaktion
<ol style="list-style-type: none"> 4. Verbringt Zeit damit, anderen beim Experimentieren zuzusehen 5. Nimmt Kontakt mit Kindern auf, die am Experiment arbeiten 6. Wiederholt den Versuch 	II. diffuses Experimentierverhalten
<ol style="list-style-type: none"> 7. Zeigt positive Emotion als Reaktion auf die Durchführung der Tätigkeit 8. Holt andere Kinder hinzu und zeigt ihnen den Versuch, Teile aus ihm oder berichtet von ihm 9. Betrachtet die „Versuchsorientierung“ genauer 	III. gerichtetes Experimentierverhalten
<ol style="list-style-type: none"> 10. Ist in das Experiment vertieft 11. Testet Variable 12. Diskussion von Kindern über das Experiment 13. Produktive Gruppenarbeit mit Arbeitsteilung 	IV. elaboriertes Experimentierverhalten

Tabelle 2.4.5.1: Allgemeines Experimentierverhalten;
- modifiziertes Kategorienschema zur Ermittlung des allgemeinen Experimentierverhaltens
Der Verfasser

Dabei werden die Verhaltensausprägungen der einzelnen Schülerinnen und Schüler in den Kategorien auf einer vierstufigen Skala eingeschätzt:

- 0: Kategorie nicht erfüllt
- 1: Kategorie teilweise erfüllt
- 2: Kategorie erfüllt
- 3: Kategorie besonders umfassend erfüllt

Ad 2: Kurzinterview: Nachdem die Verhaltensbeobachtung, wie oben beschrieben an einem zufällig ausgewählten Kind und Experiment durchgeführt wurde, folgt ein Kurzinterview, mit dem Ziel nicht beobachtbare Prozesse zu eruieren. In diesem Interview sollen drei Themenbereiche halbstandardisiert ohne Antwortvorgaben erfasst werden:

1. Wahrnehmung des Experimentes: Spannung, Freude beim Experimentieren, Schwierigkeitsgrad des Experimentes
2. kurze Beschreibung des Experimentes
3. Transferleistungen: Einbezug anderer Experimente, Verwendung von Vorwissen, Bezug zur Lebenswelt des Kindes

Die Auswertung erfolgt unter Zuhilfenahme einer quantitativen Inhaltsanalyse. Durch dieses zweistufige Vorgehen kann das experimentelle Verhalten der Schülerinnen und Schüler beschrieben werden. Der Beobachtungsbogen [allgemeines Experimentierverhalten] ist im Anhang unter Punkt 4.8 wiedergegeben.

Wichtig ist es, deutlich hervorzuheben, dass die Variable in ihrem Erfassungsbereich eine inhaltliche Schnittmenge mit der Variablen zur Erfassung der [formalen Kompetenz] hat. Ich möchte jedoch an dieser Stelle betonen, dass die Erhebungssituation für Beide eine völlig unterschiedliche ist (Labor versus Feld). Insofern können die Überschneidungen inhaltlich interpretiert werden

Eine klassische Itemrevision³¹, so wie sie bisher für die übrigen Instrumente durchgeführt wurde, soll für dieses Verfahren entfallen. Dies wird vor allem damit begründet, dass mit den Ergebnissen der einzelnen Items keine komplexen Analysen durchgeführt werden. Vielmehr wird jedes Item stets nur im Zusammenhang mit den übrigen Items einer Kategorie betrachtet. Die Zuweisung einer Versuchsperson erfolgt dabei durch einen Index, der die durchschnittliche Ausprägung der Verhaltenskategorie für die Versuchsperson angibt. Er ist definiert als der Quotient der erreichten Gesamtpunktzahl in der Kategorie, dividiert durch die Anzahl der Items in der Kategorie. Die vier Indices der Verhaltenskategorien werden abschließend miteinander verglichen. Die Versuchsperson wird der Kategorie zugewiesen in der sie den höchsten Index erreicht hat.

³¹ Ich verwende in diesem Abschnitt, abweichend von der sonst üblichen Terminologie für die einzelne Beobachtungskategorie den Begriff Item, um eine Verwechslung mit den Verhaltenskategorien zu vermeiden.

Zur Auswertung des Interviews: Die offenen Fragen werden dabei mit ein bis drei Punkten bewertet. Der Gesamtpunktwert wird durch die Addition der Punktwerte ermittelt. Bei extremen Punktwerten wird die Verhaltenskategorie nach oben beziehungsweise unten korrigiert. Die genauen cut-off-scores sind der folgenden Tabelle zu entnehmen:

Punktintervall	Folgen
7	Erhöhung des Ergebnisses der Verhaltensbeobachtung um eine Kategorie
6 – 4	Beibehalten der Kategorie aus der Verhaltensbeobachtung
3	Abzug einer Kategorie aus der Verhaltensbeobachtung

Tabelle 2.4.5.2: Allgemeines Experimentierverhalten;
- cut-off-scores des Interviews
Der Verfasser

Insofern erscheint die Berechnung mathematisch-statistischer Werte, wie zum Beispiel der Schwierigkeit und der Trennschärfe für *jedes Item* als nicht angemessen.

Können die Charakteristiken alternativ für die Kategorien berechnet werden? Ich beginne mit der Trennschärfe:

Sie kann nicht ermittelt werden, da der für die Korrelation notwendige Referenzwert, zumeist die Gesamtpunktzahl, nicht sinnvoll ermittelt werden kann. Insofern ist die Frage, wie gut jede einzelne Kategorie in ihrer Ausprägung den Gesamttestwert widerspiegelt, forschungspraktisch ohne Bedeutung.

Die Berechnung der Schwierigkeit der Kategorien ist hingegen praktisch relevant und mathematisch möglich. Die Versuchspersonen werden anhand des vorgeschlagenen Indexes auf die vier Kategorien aufgeteilt. Die Schwierigkeit der Kategorien sollte von II bis IV deutlich ansteigen, das heißt in den höheren Verhaltenskategorien sollten sich weniger Untersuchungsteilnehmer befinden, als in den unteren. Es ergibt sich die folgende Verteilung:

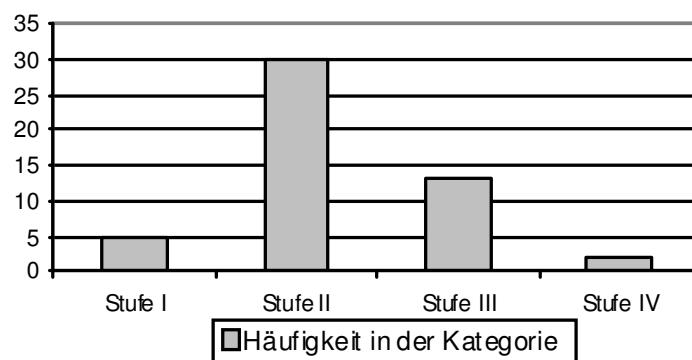


Diagramm 2.4.5.1: Allgemeines Experimentierverhalten;
- Verteilung der Verhaltensstufen im Prätest
Der Verfasser

Überträgt man diese absoluten Zahlen in relative Werte, so ergibt sich die in der Tabelle dargestellte Verteilung:

Nummer	Verhaltenskategorie	relative Häufigkeit im Prätest
1	Stufe I	10%
2	Stufe II	60%
3	Stufe III	26%
4	Stufe IV	4%

Tabelle 2.4.5.3: Allgemeines Experimentierverhalten;
- Verteilung der Verhaltenskategorien im Prätest
Der Verfasser

Mit zunehmender Verhaltensstufe sind die Kategorien seltener besetzt sind, das heißt sie werden schwerer.

Die Itemrevision soll abschließend anhand von qualitativen Kriterien vertieft werden. Leider werden von Seiten der qualitativen Forschung (vgl. Mayring 1995, Flick 2003) keine Kriterien vorgegeben, wie ein gutes Item aussehen sollte.

Ich orientiere mich daher an folgenden Kriterien:

1. Sind die Items sicher in der Untersuchungssituation beobachtbar?
2. Sind die Items hinreichend genau formuliert, das heißt: Ist der durch das Item zu beschreibende Sachverhalt klar gefasst?
3. Gelingt die Differenzierung der Versuchspersonen auf den Items gut?
4. Stellen die Items in der Situation der Beobachtung keine Überforderung des Beobachters in Bezug auf Komplexität oder Umfang dar?
5. Wird der nach der Theorie zu beobachtende Sachverhalt durch die Items genau gefasst?
6. Die Versuchspersonen zeigen darüber hinaus kein weiteres theoretisch fruchtbares Verhalten?
7. Ist das Verfahren insgesamt ökonomisch?

Die meisten der vorgestellten Kategorien werden durch den Fragebogen erfüllt. Es hat sich jedoch gezeigt, dass das halbstandardisierte Interview im Anschluss an den Versuch für die Kinder zu schwierig ist und dass es zu viel Zeit benötigt. Daher wird es so weit wie möglich in ein hochstandardisiertes umgewandelt. Das überarbeitete Verfahren wird im Abschnitt 4.9 dieser Arbeit wiedergegeben.

Der Erhebungsbogen [allgemeines Experimentierverhalten] kann in der modifizierten Form in der Hauptuntersuchung verwendet werden.

2.4.6 Der modifizierte Fragebogen zur Erfassung der aktuellen Motivation

Der FAM kann in seiner Ursprungsform nicht für die Untersuchung verwendet werden. Zum einen ist er mit Blick auf die Stichprobe zu ausführlich, zum anderen müssen die Items sprachlich deutlich einfacher gestaltet werden. Bei der Reduktion der Anzahl der Items erscheint es sinnvoll, sich an den Ergebnissen der Faktoranalyse (siehe unten) zu orientieren: Die folgende Tabelle listet die Faktorladungen der einzelnen Items auf.

Item	Faktor I	Faktor II	Faktor III	Faktor IV
12	.85			
9	.84			
5	.79			
16	.59			
18	.59			
3		-.78		
2		.74		
14		-.72		
13		.53		
17				.75
4				.68
1				.65
7				.62
11				.60
8			.79	
10			.71	
15			.70	
6			.58	

Tabelle 2.4.6.1: Aktuelle Motivation;
- Ladungsmuster der vierfaktoriellen Lösung des FAM
in Anlehnung an Rheinberg, Vollmeyer und Burns 2001, S. 5

In den in dieser Arbeit verwendeten Fragebogen gehen nur diejenigen Items mit der größten Faktorladung (Kriterium: größer gleich .65) ein. So reduziert sich die Anzahl der Items auf 12. Diese Items müssen nun sprachlich an das Niveau von Schülerinnen und Schülern der dritten und vierten Klassenstufe angepasst werden. Dabei ergibt sich das Problem, dass das Antwortverhalten in Fragebögen nicht nur vom Inhalt der Items sondern auch maßgeblich von ihrer Formulierung abhängt (vgl. Mummendey 1995, S. 143 ff.). Aus diesem Grund sollen die Anpassungen möglichst vorsichtig erfolgen. Der überarbeitete Fragebogen ist dem Abschnitt 4.12 in dieser Arbeit zu entnehmen.

Zur Durchführung der Datenerhebung: Die Schülerinnen und Schüler bekommen alle einen Versuchsaufbau „Der cartesische Taucher“³². Im Anschluss an eine kurze Demonstration der Funktion erhalten sie einige Minuten Zeit zum Experimentieren mit dem Aufbau. Danach wird der Versuch eingesammelt und die Aufgabestellung gegeben: Die Schülerinnen und Schüler sollen „den cartesischen Taucher“ mit der Hilfe eines Legekartenprinzips mit 15 Karten erklären. Bevor sie dies jedoch tun, erhalten Sie die modifizierte Form des FAM, mit dessen Hilfe die [Motivation zur Bearbeitung der Aufgabe] erhoben wird.

Die Beurteilung des Instrumentes muss für diese Variable getrennt nach Klassenstufen erfolgen. Ich beginne mit der Ermittlung der Gütekriterien für die dritte Klassenstufe.

Den weiteren Argumentationsverlauf zusammenfassend kann an dieser Stelle gesagt werden, dass der FAM in der modifizierten Form in der dritten Klassenstufe nicht verwendet werden kann.

Die folgende Tabelle gibt die Antworthäufigkeit der Schülerinnen und Schüler für jedes Item und für jede Stufe der Skala an:

	I ₁	E ₁	M ₁	I ₂	M ₂	H ₁	M ₃	H ₂	M ₄	E ₂	H ₃	I ₃
1	0	4	22	4	39	0	22	0	44	8	0	9
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	4	0	9	0	0	0	0	8
4	0	3	3	4	0	8	4	7	0	0	0	4
5	27	23	7	13	5	5	0	10	9	4	9	9
6	80	77	75	86	59	94	72	90	54	95	98	77
<i>gesamt</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>	<i>107</i>

Tabelle 2.4.6.2: Aktuelle Motivation;
- Antwortverhalten in der dritten Klassenstufe
Der Verfasser

³² Bei jedem Testzeitpunkt erhalten die Schülerinnen und Schüler einen anderen Versuch
Prätest: „cartesischer Taucher“
Posttest I: „Teebeutelrakete“
Posttest II: „verlöschende Kerze im Aquarium“

Folgende Schlüsse können gezogen werden:

- Es gibt eine sehr starke Tendenz zum positiven Ende der Skala. Der Skalenwert sechs ist bei allen Items der am häufigsten Gewählte.
- Die mittleren Antwortmöglichkeiten wurden nur selten gewählt.
- Der Skalenwert eins wurde, nach dem Wert sechs, am häufigsten gewählt. Dies ist insbesondere bei Items des Faktors ‚Misserfolgsbefürchtungen‘ der Fall.

Was bedeuten diese Ergebnisse nun für die Motivation der Kinder? Aus der Sicht des Autors legen sie folgenden Schluss nahe: Die Leistungsmotivstruktur von Kindern ist zum Beginn der dritten Klasse ist noch zu undifferenziert, um mit der vorgeschlagenen Methodik gemessen zu werden. Der Fragebogen erfordert differenzierte Kenntnisse der eigenen Fähigkeiten und Möglichkeiten, kurz der eigenen Leistungsfähigkeit. Diese scheinen zum Anfang der dritten Klassenstufe noch nicht vorhanden zu sein. Im Vordergrund steht bei den Kindern vielmehr die Faszination für den konkreten Versuch. Sie spiegelt sich in den extrem hohen Werten auf dem Faktor ‚Interesse‘ wieder. Auffällig ist weiterhin die Häufung der Antwortmöglichkeit eins auf dem Faktor ‚Misserfolgsbefürchtungen‘. Diese Antwortmöglichkeit stellt bei diesen Items die am zweithäufigsten besetzte Kategorie dar. Die Befürchtungen scheinen für die meisten Kinder folglich entweder voll vorhanden zu sein oder sind gar nicht präsent. Differenzierungen sind den Kindern in den meisten Fällen unmöglich. Die Befürchtung, keinen Erfolg zu haben, scheint für Kinder also durchaus möglich zu sein. Sie rückt aber erst dann in den Bewusstseinsbereich der Schülerinnen und Schüler wenn sie explizit danach gefragt werden, denn bei der positiv formulierten Aussage „Ich denke ich kann die Aufgabe lösen“ stimmen die meisten Kinder zu.

Diese Interpretation wird durch Ergebnisse aus der Entwicklungspsychologie unterstützt, denen zu Folge Kinder im Alter von 7 bis 9 Jahren Leistungen nicht auf Fähigkeiten, sondern vielmehr auf Anstrengungen zurückführen (vgl. Oerter, Montada 1995, S. 789 ff.).

Die Ausführungen machen deutlich dass der Fragebogen zur Erfassung der [Motivation] in diesem Zusammenhang ungeeignet ist. Die Erfassung der [Motivation] in dieser Altersstufe scheint mit angemessenem Aufwand leider nicht möglich:

- Alle Befragungsformen weisen den Mangel auf, dass die Leistungsmotivstruktur noch wenig ausgeprägt ist. Die Selbsteinschätzung wird daher immer zu den Extremen der Skala, zumeist zum positiven Ende, tendieren.
- Beobachtungsverfahren erscheinen aus zwei Gründen wenig geeignet: Zum einen erfordern sie einen hohen Zeitaufwand, der aus untersuchungspraktischen Gründen nicht realisierbar ist, zum anderen ist [Motivation] sicher in Teilen, nicht aber vollständig beobachtbar.

Auf die Erfassung der Variablen [Motivation] in der dritten Klassenstufe muss aufgrund eines fehlenden präzisen und ökonomischen Verfahrens verzichtet werden.

In der vierten Klasse zeigt sich ein deutlich anderes Bild. Ich beginne mit der Itemrevision:

Nummer	Codierung	Schwierigkeit	Trennschärfe
1	I ₁	0,9	0,6
2	E ₁	0,7	0,7
3	M ₁	0,7	0,6
4	I ₂	0,9	0,3
5	M ₂	0,4	0,4
6	H ₁	0,8	0,3
7	M ₃	0,6	0,7
8	H ₂	0,8	0,2
9	M ₄	0,5	0,6
10	E ₂	0,8	0,5
11	H ₃	0,8	0,5
12	I ₃	0,6	0,3

Tabelle 2.4.6.3: Aktuelle Motivation;
- Ergebnisse der Itemrevision
Der Verfasser

Entsprechend der bereits genannten Kriterien von Bortz und Döring (1995, S. 199 ff.) werden die Items eins, vier und acht aus dem Fragebogen entfernt. In der Endform hat der Fragebogen damit insgesamt neun Items. Die durchschnittliche Schwierigkeit des Bogens liegt bei 0,6, die durchschnittliche Trennschärfe bei 0,5.

Wie auch bei der Einstellung, soll das zugrundeliegende theoretische Modell des Fragebogens und die Zuordnung der Items zu den Motivationskomponenten unter der Anwendung einer konfirmatorischen Faktoranalyse überprüft werden: Es ergibt sich ein inhomogenes, nur schwer interpretierbares Bild. Setzt man hingegen die Anzahl der Faktoren auf drei herab, so erhält man ein deutlich klareres Bild, das in der folgenden Tabelle wiedergegeben ist:

Nummer	Codierung	1	2	3
1	M ₁	0,8		
2	M ₂	0,7		
3	M ₃	0,7		
4	M ₄	0,7		
5	M ₅	0,7		
6	E ₂		0,8	
7	H ₁		0,8	
8	H ₂		0,7	
9	I ₁			0,9

Tabelle 2.4.6.4: Aktuelle Motivation;
- Ergebnisse der konfirmatorischen Faktoranalyse
Der Verfasser

Durch die Faktoranalyse wird ein zufriedenstellend hoher Anteil der Varianz von 67% aufgeklärt.

Der obigen Tabelle ist zu entnehmen, dass die vorher getrennten Faktoren ‚Erfolgswahrscheinlichkeit‘ und ‚Herausforderung‘ nun gemeinsam auf einem Faktor laden. Wie kann dieser Umstand inhaltlich interpretiert werden?

Zunächst kann festgestellt werden, dass die beiden übrigen Faktoren (‚Interesse‘ (Faktor 3) und ‚Misserfolgsbefürchtungen‘ (Faktor 1)) sich auch in dieser Untersuchung bestätigen lassen. Sie können daher übernommen werden. Das Item „ich glaube ich schaffe diese Aufgabe nicht“ konnte in dieser Untersuchung jedoch nicht dem Faktor ‚Herausforderung‘ (Item lud negativ auf diesem Faktor) zugerechnet werden, sondern vielmehr dem Faktor ‚Misserfolgsbefürchtungen‘, was inhaltlich im Falle der gegebenen positiven Ladung stimmig ist.

Auf dem Faktor zwei (siehe Tabelle 2.3.6.4) laden die folgenden Items:

- Ich glaube ich kann die Aufgabe lösen.
- Ich bin gespannt wie gut ich bei dieser Aufgabe bin.
- Wenn ich die Aufgabe schaffe, bin ich stolz auf mich.

Auf diesem Faktor laden folglich Items, die die Erwartungen an die eigene Kompetenz und den Umgang mit ihr beschreiben. Als Kurzfassung erscheint die Bezeichnung des Faktors mit dem Begriff ‚Kompetenzerwartungen‘ als angemessen.

Problematisch an der vorliegenden Form des Fragebogens ist die zu geringe Anzahl der Items auf dem Faktor ‚Interesse‘. Bedingt durch das hohe Interesse an derartigen Fragestellungen waren die beiden übrigen Items deutlich zu leicht. Benötigt werden demnach Items die im oberen Bereich der Skala besser differenzieren:

Folgende Items seien daher vorgeschlagen:

1. Ich finde es interessant über die Erklärung für den Versuch selber nachzudenken.
2. Ich möchte lieber eine fertige Erklärung bekommen.
3. Bei solchen Aufgaben brauche ich keine Belohnung. Sie machen mir auch so viel Spaß.

Anhand einer kleineren Stichprobe wurden diese Items auf ihre Schwierigkeit und Trennschärfe hin untersucht.

Nummer	Codierung	Schwierigkeit	Trennschärfe	Ladung auf Faktor 3
1	I ₁	0,5	0,6	Ja (0,7)
2	I ₂	0,7	0,5	Ja (-0,5)
3	I ₃	0,9	0,4	Ja (0,8)

Tabelle 2.4.6.5: Aktuelle Motivation;
- Ergebnisse der zweiten Itemrevision
Der Verfasser

Folglich werden die Items eins und zwei mit in den Fragebogen aufgenommen. Die durchschnittliche Schwierigkeit und die durchschnittliche Trennschärfe des Fragebogens verändern sich durch die Hinzunahme der Items nur unbedeutend.

Die Berechnung der Reliabilität erfolgt durch die Bestimmung des arithmetischen Mittels der split-half-Reliabilitäten der Faktoren. Nach der Korrektur durch die „Spearman-Brown-Formel“ (vgl. Lienert 1969, S. 210) ergibt sich eine Reliabilität von 0,8. Sie ist als zufriedenstellend einzuschätzen.

Der Fragebogen weist mit seinen Korrekturen eine ausreichende logische Validität auf. Er repräsentiert, dies belegen die Ergebnisse der Faktoranalyse, die modifizierte Theorie der [aktuellen Motivation].

Das überarbeitete Verfahren ist dem Abschnitt 4.13 zu entnehmen.

Damit erfüllt der überarbeitete Fragebogen die üblichen Kriterien der empirischen Sozialforschung und kann in der Untersuchung für die vierte Klassenstufe verwendet werden.

2.4.7 Wissensdiagnostik

Die Erhebung der Variable [Wissen] erfolgt im zeitlichen Zusammenhang mit der Variablen [Motivation]. Dies sei am Beispiel der Prätestuntersuchung verdeutlicht, in der als Versuch der „cartesische Taucher³³“ zum Einsatz kommt: Alle Kinder der Klasse erhalten zunächst einen fertigen Versuchsaufbau. Nach einer kurzen Demonstration haben sie Zeit, ausgiebig zu experimentieren. Danach wird der Versuch wieder eingesammelt. Die Schülerinnen und Schüler der vierten Klassenstufe bearbeiten nun die modifizierte Form des FAM. Ist diese Aufgabe vollständig bearbeitet, so setzen sie sich in zufälligen Gruppen mit je zwei bis drei Kindern zusammen. Jede Gruppe erhält erneut einen Versuchsaufbau mit der Bitte gemeinsam eine Erklärung für den „cartesischen Taucher“ zu erarbeiten.

Die Bearbeitung der Aufgabe erfolgt in Anlehnung an das Prinzip der Concept-Maps (vgl. Reiska 2002) als ein Kartenlegesystem mit 15 vorgegeben Karten³⁴: acht Knoten und sieben Verknüpfungen. Zur besseren Übersicht sind beide Arten von Karten farblich klar getrennt. Sie sollen von den Kindern in eine für sie stimmige Reihenfolge gebracht werden. Bevor sie mit der Aufgabe beginnen, führe ich das Prinzip anhand eines Beispiels aus einem anderen Bereich ein.

Die Kinder legen nun das Map. Dabei können Sie immer wieder auf das Experiment zur Überprüfung zurückgreifen. Die Begriffe sind so gewählt dass sie lediglich Beobachtbares umfassen. Es handelt sich folglich um eine Kombination von der Erklärung des Phänomens und einer Funktionsbeschreibung, für die alle notwendigen Erkenntnisse aus der konkreten Beobachtung des jeweiligen Experimentes generiert werden können.

³³ An den drei Untersuchungsterminen können nicht die gleichen Versuche verwendet werden, da Erinnerungseffekte nicht auszuschließen wären. Aus diesem Grund werden bei den Posttests andere Versuche genutzt (vgl. Fußnote 15).

³⁴ Aus Gründen der Ökonomie verwende ich in diesem Zusammenhang das „paper and pencil“ Verfahren. Dies erscheint auch vor dem Hintergrund der Erfahrungen von Schließmann (2005, S. 27 ff.) sinnvoll. Während der Erhebungen zu seiner Dissertation machte er die Erfahrungen, dass Kindern im Alter der hier zu untersuchenden Population die Handhabung des Computers noch schwer fällt.

Dieses Verfahren bietet folgende Vorteile:

- Zeitersparnis: Durch die Kombination der Erhebungsrahmen der unterschiedlichen Variablen
- Theoretische Stimmigkeit: Erhoben wird kein Schulwissen im klassischen Sinne sondern das Ergebnis der Denkbemühungen der Kinder in der praktischen Auseinandersetzung mit dem Experiment.
- Schaffung der Möglichkeit eines Vergleiches mit einer Kontrollgruppe: Der erste Zugang, die Wirksamkeit von ASIP mit der von normalem Unterricht zu vergleichen, musste verworfen werden. Dies ist zum einen theoretisch problematisch, da mit der MINIPHÄNOMENTA völlig andere Ziele verbunden sind als mit normalem Unterricht, zum anderen ergeben sich durch dieses Vorgehen massive praktische und forschungsmethodische Probleme: Wie sollte zum Beispiel die Vergleichbarkeit des Unterrichts in den einzelnen Klassen sichergestellt werden? Durch die Standardisierung der Aufgabe für alle Elemente der Stichprobe und durch die bewusste Herausnahme aus dem unterrichtlichen Geschehen können die Ergebnisse der Kontroll- und Experimentalgruppe adäquat verglichen werden.

Die Auswertung der Concept-Maps erfolgt über eine quantitative Inhaltsanalyse mit den folgenden Kategorien:

- Anzahl der richtigen Propositionen
- Anzahl der verwendeten Propositionen
- Anzahl richtiger Insel³⁵
- Qualität der Inseln
- Grad der Verknüpfung der Inseln
- Qualität des Netzes insgesamt

Eine klassische Itemrevision auf mathematisch-statistischer Grundlage ist nur teilweise möglich. Ich beginne mit der Berechnung der Kategorienschwierigkeit, die in der folgenden Tabelle wiedergegeben ist:

³⁵ Der Begriff der Insel wird hier abweichend zum Programm CCMaP verwendet. Sie bezeichnen hier keine allein stehenden Begriffe, sondern vielmehr richtige Elemente in fehlerhaften Abschnitten des Maps.

Nummer	Kategorie	Schwierigkeit
1	Anzahl der verwendeten Teile ³⁶	0,7
2	Anzahl der richtigen Teile	0,3
3	Anzahl der Inseln	0,3
4	Qualität der Inseln ³⁷	0,3
5	Verknüpfung der Inseln	0,3

Tabelle 2.4.7.1: Wissen;
- Ergebnisse der Kategorienrevision
Der Verfasser

Mit einer durchschnittlichen Itemschwierigkeit von 0,4 handelt es sich um ein eher schweres Verfahren. Basierend auf den Kriterien von Bortz und Döring (1995, S. 199 ff.) muss aber keine der Kategorien ausgesondert werden.

Die Berechnung der Trennschärfe kann nicht erfolgen, da die erreichbaren Maximalpunktzahlen zwischen den Kategorien gravierend schwanken. Die Trennschärfe von Items mit einem hohen Maximalpunktwert und damit auch in Relation zu anderen Items mit einem höheren durchschnittlichen erreichten Punktwert würde damit deutlich überschätzt. Die Trennschärfe der Items mit einem niedrigen Maximalpunktwert würde unterschätzt.

Zur zusätzlichen Absicherung der Ergebnisse soll für dieses Instrument flankierend eine qualitative Itemrevision erfolgen. Sie orientiert sich an folgenden Leitfragen:

- Werden durch die Kategorien alle Aspekte der Leistung der Schülerinnen und Schüler erfasst?
- Gelingt durch die Kategorien eine angemessene Differenzierung der Leistungen?
- Sind die Kategorien klar formuliert?

Die vorgeschlagenen Kategorien erfüllen diese Kriterien und können daher in der Hauptuntersuchung verwendet werden.

³⁶ Der Begriff der Insel wird in diesem Zusammenhang, abweichend zur Terminologie von CCMaP, als eine örtlich begrenzte richtige Anordnung der Teile in einer ansonsten inkorrekten Positionierung verstanden. Mathematisch sind bei 15 Teilen, definiert man eine Insel im Mindestfall als die Aneinanderreihung von zwei Teilen, sieben Inseln möglich. Den zur Berechnung der Schwierigkeit erforderlichen Maximalwert habe ich hier allerdings im sozialen Vergleich, als die höchste in der Stichprobe vorkommende Anzahl von Inseln ermittelt.

³⁷ Die Qualität und den Grad der Verknüpfung der Inseln habe ich auf einer sechsstufigen Skala eingeschätzt.

Betrachtet man die Kategorien der Tabelle 2.4.7.1, so fällt deren deutliche Heterogenität auf. Diesen Eindruck bestätigt auch eine explorative Faktoranalyse. Zwei extrahierte Faktoren klären mit insgesamt 63% einen befriedigenden Teil der Gesamtvarianz auf. Die folgende Tabelle zeigt die Faktorladungen der Kategorien:

	Faktor I Qualität	Faktor II Anzahl
Verknüpfung der Inseln (V)	0,8	
Anzahl der richtigen Propositionen (ARP)	0,8	
Qualität der Inseln (Q)	0,6	
Anzahl der Inseln (AI)		0,9
Anzahl der verwendeten Propositionen (AP)		0,7

Tabelle 2.4.7.2: Wissen;
- explorative Faktoranalyse zur Ermittlung der Indexgewichte
Der Verfasser

Die Kategorien des Faktors I beschreiben die ‚Qualität der Antworten‘ der Schülerinnen und Schüler, die des Faktors II den Umfang der Lösung, also die ‚Anzahl der verwendeten Propositionen‘.

Was bedeutet dieses Ergebnis nun für die folgende Auswertung? Zunächst muss festgestellt werden, dass damit die Bildung eines Gesamtindex nicht sinnvoll erscheint, erfassen die Kategorien doch zu unterschiedliches. Ich schlage die Bildung zweier Indizes zur zusammenfassenden Kennzeichnung der Leistungen der Schülerinnen und Schüler vor:

Ein Index ist dabei definiert als „Messwert für ein komplexes Merkmal, der aus den Messwerten mehrerer Indikatorvariablen zusammengesetzt wird“ (Bortz, Döring 1995, S. 133).

Als Gewichte für die Indizes dienen dabei die ermittelten Faktorladungen. Damit berechnen sich die beiden Indizes wie folgt:

$$\text{Index I (Qualität)} = [(V \cdot 0,8) + (ARP \cdot 0,8) + (Q \cdot 0,6)] / 3$$

$$\text{Index II (Umfang)} = [(AI \cdot 0,9) + (AP \cdot 0,7)] / 2$$

Zur Bestimmung der Reliabilität: Sie kann, dies machen die Ausführungen zur Trennschärfe deutlich, nicht als Testhalbierungsreliabilität, wie bei den bisherigen Verfahren geschehen, ermittelt werden. Vielmehr wird sie als Retestreliaibilität bestimmt. Im zeitlichen Abstand von einem Monat wurde das gleiche Testverfahren zweimal bei einer Stichprobe durchgeführt. Die Korrelation der Ergebnisse betrug 0,8. Sie ist damit im befriedigenden Bereich.

Das entwickelte Verfahren ist die altersangemessene konsequente Operationalisierung des Wissensbegriffes dieser Arbeit. Es stellt eine Weiterentwicklung des Ansatzes der Concept-Maps dar, die insbesondere Rücksicht auf die geistigen Fähigkeiten der Schülerinnen und Schüler nimmt. Damit ist von einer ausreichenden logischen Validität auszugehen.

Die Itemrevision wurde erfolgreich durchgeführt. Das Verfahren kann daher in der Hauptuntersuchung verwendet werden.

2.5 Die Stufen der Wirksamkeit

Der Begriff der Wirksamkeit ist theoretisch fundiert. Die einzelnen Verfahren zur Messung der Wirksamkeit der MINIPHÄNOMENTA stehen bereit. Ab wann ist nun die MINIPHÄNOMENTA auf den einzelnen Variablen wirksam und welche Kriterien sind für die Zuweisung der Ergebnisse zu den einzelnen Stufen der Wirksamkeit geeignet?

Zur Beantwortung dieser Fragen rufe sich die Leserin oder der Leser zunächst die Definition des Wirksamkeitsbegriffes aus dem Abschnitt 2.1 in dieser Arbeit in Erinnerung. Basierend auf den Anmerkungen in dem genannten Textteil muss Wirksamkeit immer als Prä-Posttest-Unterschied bestimmt werden. Sofern das treatment, das heißt die MINIPHÄNOMENTA einen Einfluss hat, lassen sich im Posttest Änderungen nachweisen.

Damit nun aber von der Wirksamkeit der MINIPHÄNOMENTA gesprochen werden kann, müssen zwei weitere Kriterien erfüllt sein:

1. Die Veränderung muss positiv interpretierbar sein: Ist zum Beispiel die [formale Kompetenz] der Schülerinnen und Schüler durch die Teilnahme an der MINIPHÄNOMENTA im Durchschnitt um den Wert a gestiegen oder ist die ‚affektive Dimension‘ der [Einstellung] durch das Projekt bei den untersuchten Kindern deutlicher ausgeprägt, so ist dieses Kriterium erfüllt.
2. In einem nächsten Schritt werden die gefundenen Unterschiede auf ihre statistische Bedeutsamkeit überprüft. Aufgrund des explorativen Charakters der Untersuchung kann hier lediglich die Alpha-Fehlerwahrscheinlichkeit bestimmt werden (vgl. Bortz 1995, S. 120 f.). Entsprechend den üblichen Konventionen wird zwischen zwei Signifikanzniveaus unterschieden: hoch signifikant (Wahrscheinlichkeit für einen Alpha-Fehler kleiner als 1 %) und signifikant (Wahrscheinlichkeit für einen Alpha-Fehler kleiner als 5 %). Werte über 5 % sind grundsätzlich nicht signifikant. Dabei ist zu beachten, dass die Höhe des Signifikanzniveaus nur etwas über die Wahrscheinlichkeit einer bestimmten statistischen Fehlerart aussagt, nichts über die Größe des Effektes, das heißt über die Wirksamkeit der MINIPHÄNOMENTA (vgl. Nachtigall, Wirtz 1998, S. 200 f.).

Wenn beide Kriterien erfüllt sind, kann der Frage nach dem Umfang der Wirksamkeit des treatments nachgegangen werden. Die Entscheidung hierüber wird im Rahmen dieser Arbeit anhand von zwei Kriterien vorgenommen:

1. Bestimmung der Effektgröße: Zur Einschätzung der Größe des Effektes wird die zum statistischen Prüfverfahren gehörende Effektgröße ermittelt. Ich unterscheide hier zwischen geringen bis mittleren und hohen Effekten.
2. Umfang der Wirkung: War der Einfluss der MINIPHÄNOMENTA umfassend oder beschränkte sich die Wirkung auf kleine Bereiche (zum Beispiel eine Einstellungskomponente) oder bestimmte Gruppen (zum Beispiel die vierte Klasse).

Die Wirksamkeit der MINIPHÄNOMENTA wird methodisch als ein Prä-Posttest-Unterschied gefasst. Die Fragestellungen, ob das treatment wirksam war und wie wirksam es war, werden in vier Schritten überprüft:

1. Inhaltliche Analyse des Prä-Posttest-Unterschiedes
2. Überprüfung auf statistische Bedeutsamkeit
3. Bestimmung der Effektgröße
4. Umfang der Wirkung des treatments

Der Terminologie und Logik des sozialwissenschaftlichen Experiments folgend unterschiede ich weiterhin zwei methodisch kontrollierte Ebenen der Wirksamkeitsüberprüfung:

1. Ebene I: Wirksamkeitsüberprüfung als Haupteffekt
2. Ebene II: Wirksamkeitsüberprüfung als Interaktionseffekt

Ad 1: Wirksamkeitsüberprüfung als Haupteffekt: Wirksamkeit ist hier definiert als die Veränderung eines Merkmales nach der Zeit. Es wird folglich von der Wirksamkeit ASIPs auf eine Variable ausgegangen, sofern mindestens eine signifikante Prä-Posttest-Differenz mit einer ausreichenden Effektsärke vorliegt. Diese Wirksamkeitsebene wird aufgrund der methodischen Beschaffenheit der Studie (vgl. Abschnitt 2.1 in dieser Arbeit) lediglich für die Differenz Prätest / Posttest I (Untersuchungsstufe I) und die Prä-Posttest-Differenz der Untersuchungsstufe II ermittelt, da in beiden Fällen untersuchungsbedingt keine Ergebnisse in der Kontrollgruppe erhoben werden. Zum Überprüfung auf statistische Bedeutsamkeit werden, je nach Messniveau, t-Test für unabhängige Stichproben beziehungsweise Chi-Quadrat-Verfahren verwendet.

Ad 2: Wirksamkeitsüberprüfung als Interaktionseffekt: Zum Nachweis dieser Wirksamkeitsebene werden über zwei Vergleichszeitpunkte hinaus die Daten einer Kontroll- und einer Experimentalgruppe benötigt. Untersuchungsbedingt kann diese Methode folglich nur in Bezug auf die Differenz zwischen dem Prätest und dem Posttest II eingesetzt werden. Dieses Vorgehen bietet den Vorteil, dass im zeitlichen Verlauf der Wirksamkeit Störvariablen wie Reifungseinflüsse, Testungseffekte und -für diese Untersuchung besonders interessant- systematische Wissenszuwächse, die nicht durch die MINIPHÄNOMENTA entstanden sind, kontrolliert werden können. Zum Nachweis eben dieser Wirksamkeitsebene wird zunächst der Nettoeffekt nach Bortz und Döring (1995, S. 521 f.) berechnet:

	Prätest	Posttest I	Posttest II	
Experimentalgruppe	arithmetisches Mittel I	hier irrelevant	arithmetisches Mittel III	E = arithmetisches Mittel I - arithmetisches Mittel III
Kontrollgruppe	arithmetisches Mittel I'	nicht erhoben	arithmetisches Mittel III'	K = arithmetisches Mittel I' - arithmetisches Mittel III'
Nettoeffekt:				E - K

Tabelle 2.5.1: Stufen der Wirksamkeit;
- Berechnung des Nettoeffekts als Grundlage der Wirksamkeitsebene II
Der Verfasser

Die Überprüfung des Nettoeffektes auf statistische Bedeutsamkeit erfolgt anhand einer zweifaktoriellen Varianzanalyse auf der Basis des allgemeinen linearen Modells (vgl. Bortz 1995, S. 488 f.). Der Nettoeffekt wird dabei als Interaktionseffekt der beiden unabhängigen Variablen Zeit (Prätest, Posttest II) und Gruppenzugehörigkeit (Experimental-, Kontrollgruppe) überprüft. Liegt ein mindestens signifikanter Nettoeffekt vor, wird auch hier die Effektstärke bestimmt. Ist diese ausreichend hoch, so ist damit die Wirksamkeit von ASIP umfassend untersucht, da gezeigt werden konnte, dass das treatment kausalen Einfluss auf die abhängige Variable hat. Störeinflüsse, wie die bereits genannten, können durch den Einbezug einer Kontrollgruppe weitgehend ausgeschlossen werden.

Damit stehen die Elemente der Wirksamkeitsüberprüfung bereit. Es gilt nun die Stufen der Wirksamkeit für beide Ebenen operational zu definieren. Dies geschieht anhand der beiden folgenden Tabellen:

Ebene I: Wirksamkeit als Haupteffekt

Wirksamkeitsstufe	Bezeichnung	Definition
Wirksamkeitsstufe 1	UNWIRKSAM	Die Unterschiede zwischen Prä- und Posttest sind nicht signifikant.
Wirksamkeitsstufe 2	TEILWEISE WIRKSAM	<ol style="list-style-type: none"> 1. Mindestens signifikanter Unterschied 2. Effektstärke: Intervallskalierte Merkmale: Cohans d-Maß dargestellt als Überlappungsbereich nach Bortz und Döring (2006, S. 607 f.): 69% bis 88% Ordinalskalierte Merkmale: ϵ nach Bortz (1995, S. 150 ff.): 0,3 bis 0,8 3. Umfang der Wirkung³⁸: Wenige Bereiche (59% bis 0%)
Wirksamkeitsstufe 3	WIRKSAM	<ol style="list-style-type: none"> 1. Mindestens signifikanter Unterschied 2. Effektstärke: Intervallskalierte Merkmale: Cohans d-Maß dargestellt als Überlappungsbereich nach Bortz und Döring (2006, S. 607 f.): 69% bis 88% Ordinalskalierte Merkmale: ϵ nach Bortz (1995, S. 150 ff.): 0,3 bis 0,8 3. Umfang der Wirkung: Ein Großteil der Bereiche (99% bis 60%)
Wirksamkeitsstufe 4	HOCHWIRKSAM	<ol style="list-style-type: none"> 1. Mindestens signifikanter Unterschied 2. Effektstärke: Intervallskalierte Merkmale: Cohans d-Maß dargestellt als Überlappungsbereich nach Bortz und Döring (2006, S. 607 f.): kleiner als 69% Ordinalskalierte Merkmale: ϵ nach Bortz (1995, S. 150 ff.): größer als 0,8 3. Umfang der Wirkung: Alle Bereiche (100%)

Tabelle 2.5.2: Stufen der Wirksamkeit;
- Wirksamkeitsstufen der Ebene I
Der Verfasser

³⁸ Für die Untersuchungsstufe II muss in Bezug auf die eingeführten Cut-off-scores eine Einschränkung gemacht werden. Da hier keine differentialpsychologischen Unterschiede ermittelt worden sind, gelten für die Zuweisung der Wirksamkeitskategorien lediglich die beiden erstgenannten Kriterien. Um trotzdem eine angemessene Differenzierung erreichen zu können wird hier zusätzlich zwischen kleinen und mittleren Effekten (Überlappungsbereich: 92% bis 89% bzw. 88% bis 69%, ϵ : 0,2 bis 0,29 bzw. 0,3 bis 0,8) unterschieden.

Ebene II: Wirksamkeit als Interaktionseffekt

Wirksamkeitsstufe	Bezeichnung	Definition
Wirksamkeitsstufe 1'	UNWIRKSAM	Kein signifikanter Nettoeffekt
Wirksamkeitsstufe 2'	TEILWEISE WIRKSAM	<ol style="list-style-type: none"> 1. Mindestens signifikanter Unterschied 2. Effektstärke: Intervallskalierte Merkmale: $\eta^2_{\text{part.}}$ nach Bortz und Döring (2006, S. 606): 0,10 bis 0,24 Ordinalskalierte Merkmale: ϵ nach Bortz (1995, S. 150 ff.): 0,3 bis 0,8 3. Umfang der Wirkung: Wenige Bereiche (59% bis 0%)
Wirksamkeitsstufe 3'	WIRKSAM	<ol style="list-style-type: none"> 1. Mindestens signifikanter Unterschied 2. Effektstärke: Intervallskalierte Merkmale: $\eta^2_{\text{part.}}$ nach Bortz und Döring (2006, S. 606): 0,10 bis 0,24 Ordinalskalierte Merkmale: ϵ nach Bortz (1995, S. 150 ff.): 0,3 bis 0,8 3. Umfang der Wirkung: Ein Großteil der Bereiche (99% bis 60%)
Wirksamkeitsstufe 4'	HOCHWIRKSAM	<ol style="list-style-type: none"> 1. Mindestens signifikanter Unterschied 2. Effektstärke: Intervallskalierte Merkmale: $\eta^2_{\text{part.}}$ nach Bortz und Döring (2006, S. 606): größer als 0,24 Ordinalskalierte Merkmale: ϵ nach Bortz (1995, S. 150 ff.): größer als 0,8 3. Umfang der Wirkung: Alle Bereiche (100%)

Tabelle 2.4.3: Stufen der Wirksamkeit;
- Wirksamkeitsstufen der Ebene II
Der Verfasser

Die vorgestellte Verfahrensweise gilt jedoch nicht für alle Variablen. Es gibt zwei Ausnahmen:

1. [Deskriptive Aspekte] auf der Ebene I und II: Die Leserin oder der Leser rufe sich an dieser Stelle erneut die Definition von Wirksamkeit ins Gedächtnis. Unabhängig von ihrer methodischen Erfassung ist sie definiert als die kausale Wirksamkeit des treatments MINIPHÄNOMENTA auf eine abhängige Variable. Erfasst wird dies immer durch den Vergleich eines Prätestergebnisses mit dem eines Posttests. Im Falle der Variablen [deskriptive Aspekte] kann dieser Vergleich nicht erfolgen, da untersuchungsbedingt keine Prätestergebnisse vorliegen. Die Erinnerung der Stationen fasse ich daher in dieser Studie nicht als Ausdruck von Wirksamkeit sondern als die Bedingung für die Wirksamkeit von ASIP.
2. [Allgemeines Experimentierverhalten] auf der Ebene II: Die Variable weist lediglich ein ordinales Skalenniveau auf, daher können varianzanalytische Verfahren nicht angewandt werden. Statt der Berechnung des Nettoeffektes und seiner statistischen Überprüfung als Interaktionseffekt vergleiche ich die Differenzen zwischen den Testzeitpunkten und den Gruppen über eine Konfigurationsfrequenzanalyse (vgl. Bortz 1995, S. 168 f.). Auch hier wird zur Feststellung der Höhe der Wirksamkeit des treatments die Effektstärke bestimmt.

Abschließend sei in diesem Zusammenhang auf den Begriff der praktischen Bedeutsamkeit eingegangen. Bei Evaluationsvorhaben in gut erforschten Bereichen ist es üblich, eine praktisch bedeutsame Effektgröße über das Signifikanzniveau hinaus anzugeben (vgl. Bortz, Döring 1995, S.564). Dies ist hier nicht möglich, da Versuche eine valide Größe zu bestimmen, scheiterten: Eine Bestimmung scheint im Bereich der empirischen Bildungsforschung – Kindheitsforschung nur über den Weg des Experteninterviews möglich. Die befragten Lehrerinnen und Lehrer gaben aber zu bedenken, dass sie kein Maß angeben können, da ihnen hier die Erfahrungen fehlen. Auch eine Bestimmung über Forscherinnen und Forscher ist nicht möglich: Es fehlt schlicht an einer ausreichend breiten Datenbasis im Bereich der physikalischen Elementarbildung.

2.6 Konkretisierung der Fragestellungen: Die Hypothesen der Untersuchung

Die Komplexität des Untersuchungsdesigns bringt eine Vielzahl von zu überprüfenden Hypothesen mit sich. Sie werden an dieser Stelle in Kategorien zusammenfassend dargestellt:

Hypothesencluster der Untersuchungsstufe I:

1. Die Ergebnisse der Kontroll- und der Experimentalgruppe unterscheiden sich im Prätest nicht.
2. In der Experimentalgruppe zeigt sich ein signifikanter Unterschied zwischen Prä- und Posttest I.
3. Dieser Unterschied bleibt bis zum Posttest II erhalten.
4. Die Wirksamkeit der MINIPHÄNOMENTA zeigt sich in allen sechs erfassten Variablen.
5. Es ist anzunehmen, dass die MINIPHÄNOMENTA auf einige Variablen besonders großen Einfluss haben wird.
6. Die Wirksamkeit von ASIP hängt nicht vom Geschlecht der Schülerinnen und Schüler und nicht von der besuchten Klassenstufe ab. Damit ist die MINIPHÄNOMENTA in allen Subgruppen der Stichprobe wirksam.

Hypothesencluster der Untersuchungsstufe II:

1. Die MINIPHÄNOMENTA ist auch nach dieser langen Zeitspanne noch wirksam.
2. Es ist anzunehmen, dass ASIP auf einigen Variablen besonders wirksam ist.

Bei den genannten Hypothesen handelt es sich zumeist um gerichtete Veränderungshypothesen, die in der Form von Unterschiedshypothesen geprüft werden.

3 Darstellung der Ergebnisse

3.1 Einleitung in die Auswertung

Zur besseren Übersicht über den folgenden Auswertungsteil werden in diesem Abschnitt die einzelnen Auswertungsschritte hinsichtlich ihrer zentralen Forschungsfrage und der verwendeten Methodik kurz dargestellt. Zunächst nochmals das Untersuchungsdesign:

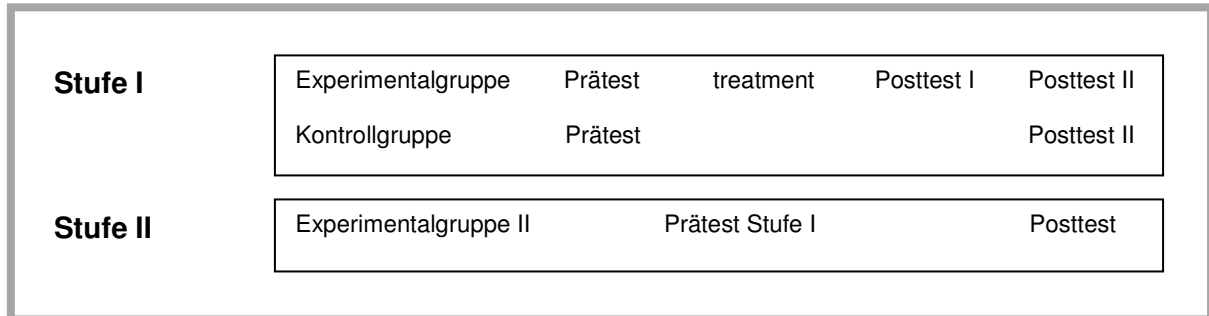


Abbildung 3.1.1: Einleitung in die Auswertung;
- das Untersuchungsdesign der Studie
Der Verfasser

Ich beginne die Auswertung der Untersuchung (vgl. Abschnitt 3.2.1.2 in dieser Arbeit), nach dem Nachweis der Vergleichbarkeit von Experimental- und Kontrollgruppe, mit der Beschreibung der Ergebnisse auf den abhängigen Variablen im Vorfeld des treatments. Hierzu analysiere ich die Prätestdaten der Experimental- und der Kontrollgruppe. Zum Einsatz kommen dabei vorrangig deskriptive Verfahren. Die Ergebnisse können als Indikator für die Physikalische Bildung der Schülerinnen und Schüler verstanden werden.

Die erste Vergleichsmöglichkeit (vgl. Abschnitt 3.2.2.1 in dieser Arbeit) besteht in der Analyse des Unterschiedes zwischen den Prätestergebnissen (Kontroll- und Experimentalgruppe) und den Ergebnissen des Posttest I. Das Ziel dieses Textteils besteht in der Untersuchung der Wirksamkeit der MINIPHÄNOMENTA auf die abhängigen Variablen nach einem Zeitraum von drei Monaten. Mathematisch bediene ich mich hier im Falle der metrischen Variablen eines t-Testes für unabhängige Stichproben, mit dem die Mittelwertsunterschiede auf ihre statistische Bedeutsamkeit hin überprüft werden. Die Variable allgemeines Experimentierverhalten weist lediglich ein ordinales Niveau auf. Zum Vergleich der Kategoriehäufigkeiten wird ein k*I-Chi-Quadrat-Test angewandt. Die Zuweisung zu den Wirksamkeitsstufen geschieht im Anschluss durch den Umfang der Wirksamkeit auf die Variable und die Bestimmung der Effektgröße. Basierend auf den Erörterungen des Abschnittes 2.4 wird die Variable [deskriptive Aspekte] als eine Bedingung der Wirksamkeit von ASIP gefasst. Ihre Auswertung erfolgt daher aufgrund des untersuchungsbedingten Fehlens einer Vergleichsmöglichkeit mit Prätestergebnissen vorrangig deskriptiv.

Im Anschluss an den Nachweis der Wirksamkeit auf der Ebene I folgt der methodisch kontrolliertere Nachweis von Wirksamkeit durch die Verwendung einer zweifaktoriellen Varianzanalyse beziehungsweise bei der Variablen allgemeines Experimentierverhalten einer Konfigurationsfrequenzanalyse (vgl. Abschnitt 3.2.3.2). Durch den Einbezug einer Kontrollgruppe kann die Wirksamkeit des treatments deutlich genauer gemessen werden, da zum Beispiel Reifungseinflüsse und systematische Wissenszuwächse kontrolliert werden können. Zusätzlich wird der zeitliche Verlauf der Wirksamkeit in der Experimentalgruppe durch die vergleichende Gegenüberstellung der Ergebnisse des Posttests I und II untersucht. Die Zuweisung zu den Wirksamkeitsstufen geschieht aufgrund der genannten Indikatoren.

Die Untersuchungsstufe II (vgl. Abschnitt 3.3.1 in dieser Arbeit) liefert Indizien für die Existenz der Langzeitwirksamkeit auf der Ebene I über einen Zeitraum von 27 Monaten. Verglichen werden hier die Ergebnisse des Posttests mit dem Prätest der Untersuchungsstufe I. In Abhängigkeit des Messniveaus der Skalen kommt hier erneut ein t-Test für unabhängige Stichproben beziehungsweise ein k*I-Chi-Quadrat-Test zum Einsatz. Auch hier werden die Ergebnisse der unabhängigen Variablen aufgrund des Umfanges der Wirkung und der bestimmten Effektgröße Wirksamkeitsstufen zugewiesen.

Die Zuweisung zu den Wirksamkeitsstufen in beiden Untersuchungsteilen erfolgt lediglich für die Ergebnisse der Variablen insgesamt und nicht für einzelne Subgruppen der Stichprobe oder des operationalisierten Merkmals. Bezogen auf das Beispiel der [Motivation] bedeutet dies, dass lediglich das gesamte Merkmal einer der Wirksamkeitsstufen zugewiesen wird.

An dieser Stelle sei nochmals auf den explorativen Charakter der Untersuchungsstufe II hingewiesen. Die genannten methodologischen Einschränkungen begrenzen die Validität der Aussagen für diesen Bereich der Untersuchung. Für diese Untersuchung haben die Ergebnisse Ausblickscharakter. Sie untermauern und vertiefen die methodisch umfassend abgesicherten Daten der Untersuchungsstufe I, die das eigentliche Zentrum der Untersuchung darstellen. Entsprechend dieser Aufgabenstellung erfolgt die Auswertung der einzelnen Variablen nicht so umfassend, wie in der Untersuchungsstufe I. Es werden vielmehr nur einige globale Datentendenzen ermittelt.

3.2 Ergebnisse der Untersuchungsstufe I

3.2.1 Der Prätest

3.2.1.1 Die Vergleichbarkeit von Experimental- und Kontrollgruppe im Vorfeld des treatments

Der Vergleich der Variablenwerte von Kontroll- und Experimentalgruppe soll für jede Variable erfolgen. Berücksichtigt werden dabei allerdings nur wenige zentrale Werte. So erfolgt zum Beispiel in Bezug auf die Variable [Einstellung] lediglich der Vergleich des durchschnittlichen Wertes und seiner Streuung. Unberücksichtigt bleiben bei diesem Vergleich der Einfluss des Alters, der Klassenzugehörigkeit, des Geschlechts usw.

Die Variable allgemeines Experimentierverhalten

Das folgende Diagramm gibt die erreichten Verhaltenskategorien der Kontroll- und Experimentalgruppe I aufgrund des unterschiedlichen Stichprobenumfanges in relativen Werten an:

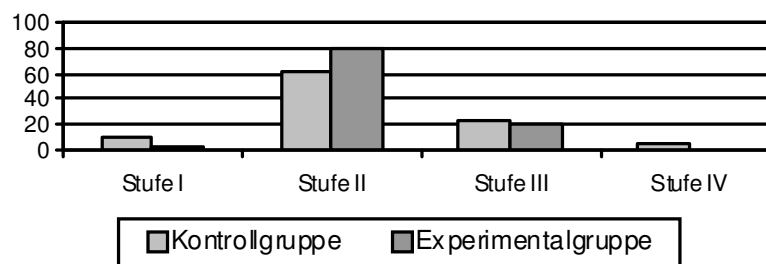


Diagramm 3.2.1.1.1: Allgemeines Experimentierverhalten;
- erreichte Verhaltenskategorien des allg. Experimentierverhaltens in relativen Werten
Der Verfasser

Beide Verteilungen ähneln sich in einer ersten Begutachtung deutlich. Anhand der folgenden Kreuztabelle soll eine statistisch exakte Überprüfung der Ähnlichkeit der beiden Verteilungen unternommen werden:

	Stufe I	Stufe II	Stufe III	Stufe IV	Gesamt
Kontrollgruppe	17	100	37	8	162
Experimentalgruppe	1	49	13	0 ³⁹	63
Gesamt	18	149	50	8	225

Tabelle 3.2.1.1.1: Allgemeines Experimentierverhalten;
- Kreuztabelle zur zum allgemeinen Experimentierverhalten
Der Verfasser

³⁹ Bei der Stufe IV kommen erwartete Zellenhäufigkeiten von kleiner als 5 vor. Nach den Richtlinien von Bortz (1995, S. 170 f.) darf das Testverfahren trotzdem angewandt werden, da der Anteil dieser niedrigen erwarteten Häufigkeiten kleiner als 20% ist.

Die statistische Überprüfung erfolgt durch einen k*I-Chi-Quadrat-Test:

Größe	ermittelter Wert
df	3
χ^2	9,54

Tabelle 3.2.1.1.2: Allgemeines Experimentierverhalten;
- k*I-Chi-Quadrat-Test zum Vergleich von Experimental- und Kontrollgruppe
Der Verfasser

Es muss die Forschungshypothese angenommen werden. Es besteht folglich ein signifikanter Unterschied zwischen den beiden Häufigkeitsverteilungen.

Die Variable Einstellung gegenüber Phänomenen und ihrer Bearbeitung

Das folgende Diagramm zeigt die arithmetischen Mittel und die Standardabweichungen der Experimental- und der Kontrollgruppe.

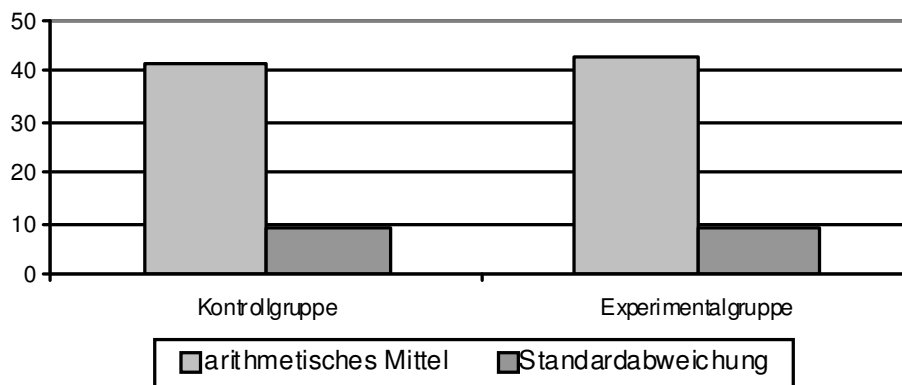


Diagramm 3.2.1.1.2: Einstellung gegenüber Phänomen und ihrer Bearbeitung;
- Basiswerte der Einstellung für Experimental- und Kontrollgruppe
Der Verfasser

Beide Verteilungen sind einander sehr ähnlich. Die Differenz der Werte des arithmetischen Mittels beträgt lediglich 1,1 Punkte, die Differenz der Streuungen ist mit einem Wert von 0,3 ebenfalls sehr gering. Eine statistisch exakte Überprüfung über einen t-Test beziehungsweise F-Test auf die Ähnlichkeit der Verteilungen kann daher an dieser Stelle entfallen.

Die Kontroll- und Experimentalgruppe sind im Hinblick auf die [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] im Prätest vergleichbar.

Die Variable formale Kompetenz

Die durchschnittlich erreichten Rohpunktwerte, sowie deren Standardabweichung werden in dem folgenden Diagramm für Experimental- und Kontrollgruppe ausgewiesen:

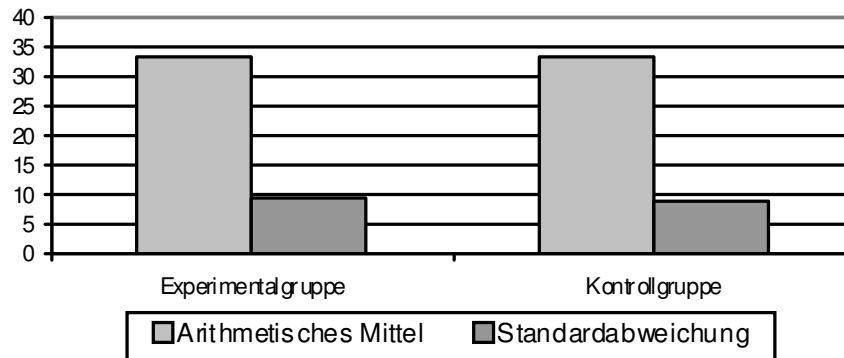


Diagramm 3.2.1.1.3: Formale Kompetenz;
- Basiswerte zur formalen Kompetenz
Der Verfasser

Die arithmetischen Mittel und die Standardabweichungen beider Gruppen sind erneut derart ähnlich, dass eine exakte statistische Überprüfung auf Unterschiede auch in diesem Fall unterlassen werden kann.

Die Kontroll- und Experimentalgruppe sind in Bezug auf die Variable [formale Kompetenz] im Prätest vergleichbar.

Die Variable aktuelle Motivation

In Bezug auf die [aktuelle Motivation] zur tiefergehenden kognitiven Verarbeitung der Inhalte ergeben sich zwischen der Kontroll- und der Experimentalgruppe leichte Unterschiede. Die folgende Tabelle gibt die zentralen Werte wieder:

	Arithmetisches Mittel	Standardabweichung
Experimentalgruppe	46,4	9,2
Kontrollgruppe	44,9	8,7
Differenz	1,5	0,5

Tabelle 3.2.1.1.3: Aktuelle Motivation;
- Basiswerte der aktuellen Motivation bezüglich der Experimental- und Kontrollgruppe
Der Verfasser

Die statistische Überprüfung der Unterschiedshypothesen erfolgt durch einen t-Test und einen F-Test:

Größe	ermittelter Wert
F-Wert	0,16
Sig.	0,68
t-Wert	-1,23
Sig.	0,29

Tabelle 3.2.1.1.4: Aktuelle Motivation;
- ermittelte Werte des t- und F-Testes zum Vergleich von Experimental- und Kontrollgruppe
Der Verfasser

In beiden Fällen kann die Nullhypothese angenommen werden. Es gibt folglich keine signifikanten Unterschiede zwischen der Experimental- und der Kontrollgruppe.

Die Kontroll- und Experimentalgruppe sind in ihrer Ausprägung auf der Variablen [aktuelle Motivation] vergleichbar.

Die Variable Wissen

Bezugnehmend auf die Erörterungen im Abschnitt 2.4.7 werden die Ergebnisse für beide Indices in der Kontroll- und der Experimentalgruppe gegenübergestellt:

	Kontrollgruppe	Experimentalgruppe
Index I		
Arithmetisches Mittel	1,7	1,6
Standardabweichung	0,6	0,6
Index II		
Arithmetisches Mittel	4,8	4,8
Standardabweichung	4,8	3,9

Tabelle 3.2.1.1.5: Wissen;
- deskriptive Werte der beiden Indices der Variablen Wissen
Der Verfasser

Die Differenzen sind in einem Fall groß genug, um sinnvoll über einen Signifikanztest überprüft zu werden. Dies betrifft die beiden Standardabweichungen des Indexes II. Durch den F-Test soll die Frage der Varianzhomogenität statistisch exakt entschieden werden:

Größe	Ermittelter Wert
F-Wert	16,0
Sig.	0,00

Tabelle 3.2.1.1.6: Wissen
- ermittelte Werte des F- Testes zum Index II der Variable Wissen
Der Verfasser

Es liegt folglich ein höchst signifikanter Unterschied vor, so dass die H_1 (Varianzheterogenität) angenommen werden muss. Die Vergleichbarkeit der Gruppen ist folglich in Bezug auf die Varianz nicht gegeben.

Die Kontroll- und Experimentalgruppe sind in ihrer Ausprägung auf der Variablen [Wissen] nur teilweise vergleichbar:

1. keine Unterschiede bei der durchschnittlich erreichten Punktzahl
2. ein signifikanter Unterschied in Bezug auf die Varianz der Messwerte des Index II

In Bezug auf die zu untersuchende Frage der Vergleichbarkeit kann festgestellt werden, dass die Ergebnisse der Kontroll- und der Experimentalgruppe äußerst homogen sind und lediglich in zwei Fällen signifikante Unterschiede aufweisen. Der erste Unterschied besteht in Bezug auf die Variable [allgemeines Experimentierverhalten]. Dieser sollte jedoch nicht überbewertet werden, da der errechnete Chi-Quadrat-Wert nur minimal über dem kritischen Wert liegt. Der zweite Unterschied betrifft die Werte des Index II der Variable [Wissen]. Hier unterschieden sich die Experimental- und die Kontrollgruppe signifikant in Bezug auf die Dispersion eben dieser Werte. Die arithmetischen Mittel weichen jedoch nicht signifikant voneinander ab. Im Durchschnitt zeigen die Schülerinnen und Schüler der Experimental- und Kontrollgruppe folglich die gleiche Leistung. Das Leistungsspektrum in der Kontrollgruppe variiert jedoch stärker. Diese Unterschiede sollten bei der Interpretation der Daten bedacht werden, stellen jedoch die grundsätzliche Vergleichbarkeit nicht in Frage.

Die Werte der Kontroll- und der Experimentalgruppe ähneln sich weitgehend. Sie können im weiteren Verlauf der Untersuchung miteinander verglichen werden.

3.2.1.2 Der Stand vor ASIP: Ergebnisse der Prätestuntersuchung

Im vorherigen, eher methodisch orientierten Kapitel wurde die weitgehende Vergleichbarkeit der Experimental- und Kontrollgruppe hergeleitet. Dies geschah anhand einer groben Darstellung der Ergebnisse der einzelnen Variablen. An dieser Stelle sollen die Ergebnisse umfassend interpretiert werden. Der Schwerpunkt ist dabei nicht mehr forschungsmethodisch. Vielmehr sollen die inhaltlichen Implikationen der Ergebnisse herausgearbeitet werden.

Die Variable allgemeines Experimentierverhalten

Das folgende Histogramm⁴⁰ gibt die Verteilung der von den Schülerinnen und Schülern erreichten Verhaltensstufen⁴¹ an:

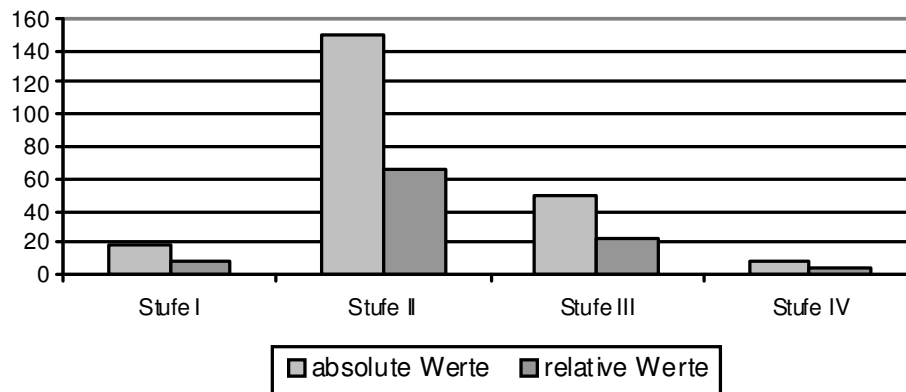


Diagramm 3.2.1.3.1: Allgemeines Experimentierverhalten; Prätest;
- erreichte Verhaltenskategorien in absoluten und relativen Werten
Der Verfasser

Die Grafik zeigt ein bedenkliches Bild des Umganges der Schülerinnen und Schüler mit den Versuchen. Auch wenn es mit 8% relativ wenige negative Reaktionen auf die Versuche gibt, so ist der größte Teil der Schülerinnen und Schüler (66%) lediglich der Verhaltenskategorie II (diffuses Experimentierverhalten) zuzuweisen. Gezieltes, hypothesengeleitetes Arbeiten gelingt ihnen nicht. Vielmehr schauen sie anderen Kindern beim Experimentieren zu, nehmen zu ihnen Kontakt auf und wiederholen den Versuch. Dabei sind sie allerdings emotional nur wenig beteiligt. Vor allem das Vorgehen beim Experimentieren ist defizitär.

Im folgenden Interview gelingt es ihnen zumeist nicht, Parallelen zu anderen Versuchen zu ziehen und auch Alltagsbezüge werden nur von den wenigsten der Schülerinnen und Schüler genannt.

⁴⁰ Die folgenden Auswertungen fußen auf den Prätestergebnissen der Experimentalgruppe I und der Kontrollgruppe. In Bezug auf das [allgemeine Experimentierverhalten] konnten 225 Untersuchungsteilnehmerinnen und Untersuchungsteilnehmer untersucht werden.

⁴¹ Die Begriffe Verhaltensstufe, Stufe, Verhaltenskategorie und Kategorie bezeichnen hier alle den gleichen Sachverhalt.

Insgesamt 22% der Schülerinnen und Schüler konnten der Verhaltenskategorie III (gerichtetes Experimentierverhalten) zugewiesen werden. Ihr Umgang mit den Versuchen ist vor allem durch positive emotionale Reaktionen auf die Experimente gekennzeichnet. Ihre Begeisterung für den Versuch drückt sich auch darin aus, dass sie andere Kinder hinzuholen und den ganzen Versuch oder Teile von ihm demonstrieren, beziehungsweise über ihn berichten. Diese Schülerinnen und Schüler nehmen also deutlich intensiveren Kontakt zu anderen Kindern auf, so dass es, ab dieser Verhaltenskategorie angebracht erscheint von sozialem Lernen der Kinder zu sprechen. Sie waren auch eher in der Lage im Interview Bezüge zu anderen Versuchen aufzubauen und konnten über Alltagsphänomene berichten, die mit dem Versuch im Zusammenhang stehen.

Lediglich 4% der Schülerinnen und Schüler erreichten die Verhaltenskategorie IV (elaboriertes Experimentierverhalten), die durch kompetentes experimentelles Handeln, intensive Prozesse des sozialen Lernens und eine hohe Motivation, d. h. ein deutliches in den Versuch vertieft sein gekennzeichnet ist.

Diese erste Analyse zeigt die deutlichen Schwächen der Kinder im allgemeinen Experimentierverhalten in der dritten und vierten Klassenstufe. Im Folgenden sollen die Daten nun genauer analysiert werden. Ich beginne mit dem Zusammenhang zwischen erreichter Verhaltenskategorie und Geschlecht⁴². Die folgende Kreuztabelle gibt diesen Zusammenhang wieder:

	Stufe I	Stufe II	Stufe III	Stufe IV	Gesamt
Mädchen	6	70	21	4	101
Jungen	12	79	29	4	124
Gesamt	18	149	50	8	225

Tabelle 3.2.1.2.2: Allgemeines Experimentierverhalten; Prätest;
 - Verhaltenskategorien des allg. Experimentierverhaltens in Abhängigkeit vom Geschlecht
 Der Verfasser

Nach der Beurteilung des ersten Eindruckes scheinen beide Verteilungen einander relativ ähnlich zu sein. Um hierüber detaillierte Aussagen treffen zu können, werden die zugehörigen relativen Werte gebildet. Sie sind im folgenden Diagramm dargestellt:

⁴² Bei der Analyse der differentialpsychologischen Unterschiede, bei denen es sich in der Sprache der Varianzanalyse um Einzelvergleiche handelt, muss das Problem der Alpha-Fehler-Kumulierung (vgl. Bortz 1995, S.260 f.) bedacht werden, dass nur durch eine Adjustierung des Alpha-Fehler-Niveaus, zum Beispiel durch eine Bonferoni-Korrektur behoben werden kann. In dieser Untersuchung wird dies nicht unternommen, da es sich bei den Einzelvergleichen ohnehin um a posteriori Hypothesen handelt, denen ohnehin, ob mit oder ohne Korrektur, nur ein explorativer Status zukommt. Sämtliche gefundenen differentialpsychologischen Unterschiede haben damit lediglich hypothetischen Charakter.

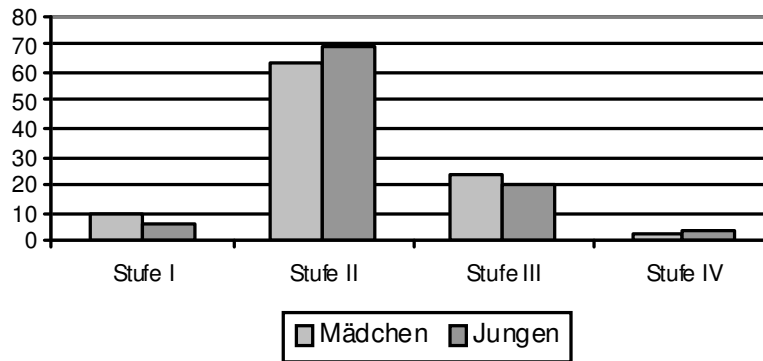


Diagramm 3.2.1.3.2: Allgemeines Experimentierverhalten; Prätest;
- das allgemeine Experimentierverhalten in Abhängigkeit des Geschlechtes
Der Verfasser

Beide Verteilungen sind, das zeigt die Grafik eindeutig, einander derartig ähnlich dass eine statistische Überprüfung nicht erforderlich ist.

Es bleibt die Frage wie dieses Ergebnis zu interpretieren ist, vor allem vor dem Hintergrund der Ergebnisse zur Variablen [formale Kompetenz]. In der Auswertung, dies wird der spätere Textverlauf ergeben, zeigt sich ein sehr ausgeprägter Unterschied zwischen den Kompetenzen der Mädchen und der Jungen. Die Mädchen waren hoch signifikant kompetenter als die Jungen. Hierfür sind aus meiner Sicht vor allem zwei Gründe anzuführen:

1. Beide Variablen erfassen nur in Teilen gleiche Gegenstände. So berücksichtigt die Variable [allgemeines Experimentierverhalten] im Gegensatz zur Variablen [formale Kompetenz] auch emotionale und soziale Aspekte.
2. Weiterhin ist, und dieser Umstand ist aus meiner Sicht noch deutlich entscheidender, die Erhebungssituation für beide Variablen grundsätzlich unterschiedlich. Die Variable [formale Kompetenz] wird in einem separaten Raum in einer Laborsituation bei lediglich zwei bis drei Schülerinnen erhoben. Hingegen geschieht die Erhebung der Variablen [allgemeines Experimentierverhalten] während des Experimentierens der gesamten Klasse.

An dieser Stelle ließe sich folgende Hypothese aufstellen: Der Ansatz der Variablen [formale Kompetenz] folgt dem „testing the limits Paradigma“ (vgl. Lienert 1996, S. 15). Ziel ist es also, die den Versuchspersonen maximal mögliche Leistung zu ermitteln. Hier scheinen Mädchen (vgl. Variable [allgemeines Experimentierverhalten]) deutlich kompetenter als Jungen. Beim gemeinschaftlichen Experimentieren im Klassenverband zeigen sie diese Leistung nicht sondern bleiben hinter ihren Möglichkeiten zurück. Es liegt an dieser Stelle nahe, hier den Einfluss der Jungen zu vermuten, die so das gängige Stereotyp, die besseren Experimentatoren und Naturwissenschaftler seien. Es wäre möglich, dass der Einfluss der Jungen es hier verhindert, dass die Mädchen ihre Möglichkeiten voll entfalten. Das Design der Arbeit bietet keine Möglichkeit diese Vermutung exakt zu überprüfen, so dass sie als Arbeitshypothese späteren Untersuchungen überantwortet werden muss.

Ich fahre fort mit der Ermittlung des Zusammenhanges von Verhaltenskategorie und Klassenstufe: Dieser wird erneut in einer Kreuztabelle dargestellt:

	Stufe I	Stufe II	Stufe III	Stufe IV	Gesamt
Dritte Klassenstufe	15	72	23	0	110
Vierte Klassenstufe	3	77	27	8	115
Gesamt	18	149	50	8	225

Tabelle 3.2.1.2.3: Allgemeines Experimentierverhalten; Prätest;
- die Verhaltenskategorien des allg. Experimentierverhaltens in Abhängigkeit der Klassenstufe
Der Verfasser

Die Tabelle zeigt in den Randbereichen (Verhaltenskategorie I und IV) auffällige Tendenzen, die näher untersucht werden müssen: Die Schülerinnen und Schüler der dritten Klasse erreichen deutlich häufiger lediglich die Stufe I und deutlich seltener die Verhaltenskategorie IV. Die Verteilung in den Stufen II und III ist im Vergleich der beiden Klassenstufen sehr ähnlich, so dass ein Vergleich der gesamten Verteilungen nicht erforderlich ist. Ich beginne mit der Untersuchung zur Stufe I:

Die statistische Überprüfung erfolgt durch einen eindimensionalen Chi-Quadrat-Test:

Größe	ermittelter Wert
df	1
χ^2	8,0

Tabelle 3.2.1.2.4: Allgemeines Experimentierverhalten, Prätest;
- ermittelte Werte des eindimensionalen Chi-Quadrat-Testes für die Verhaltenskategorie I
Der Verfasser

Es ergibt sich ein signifikanter Unterschied. Es wird folglich die H_1 (Unterschiedlichkeit der Verteilungen) angenommen.

Im Falle der Verhaltenskategorie IV erscheint eine statistische Überprüfung aufgrund der Eindeutigkeit des Ergebnisses nicht angebracht. Die Stufe IV wird in der vierten Klasse offenbar leichter erreicht als in der dritten Klasse.

Abschließend sei die Frage untersucht, ob die Gruppengröße Einfluss auf das Erreichen bestimmter Verhaltenskategorien hat. Die Werte seien hier wieder anhand der folgenden Kreuztabelle dargestellt:

	Stufe I	Stufe II	Stufe III	Stufe IV	Gesamt
Alleine	8	74	12	1	95
In der Gruppe	10	75	38	7	130
Gesamt	18	149	50	8	225

Tabelle 3.2.1.2.5: Allgemeines Experimentierverhalten; Prätest;
- Kreuztabelle zur Verhaltenskategorie in Abhängigkeit der Gruppengröße
Der Verfasser

Anhand der ersten Beurteilung der Verteilung komme ich zu dem Schluss, dass höhere Verhaltenskategorien (III, IV) eher innerhalb einer Gruppe erreicht werden. Dies sei im folgenden durch zwei eindimensionale Chi-Quadrat-Teste überprüft:

Verhaltenskategorie III	Verhaltenskategorie IV
df = 1 $\chi^2 = 13,52$	df = 1 $\chi^2 = 4,5$

Tabelle 3.2.1.2.6: Allgemeines Experimentierverhalten; Prätest;
- ermittelte Werte der eindimensionalen Chi-Quadrat-Teste
Der Verfasser

In beiden Fällen kann die H_1 angenommen werden. Die Wahrscheinlichkeit der Schülerinnen und Schüler, in Bezug auf das allgemeine Experimentierverhalten eine hohe Verhaltenskategorie zu erreichen, steigt deutlich, wenn sie in der Gruppe arbeiten.

Zusammenfassend lässt sich zur Variablen [allgemeines Experimentierverhalten] im Vorfeld der MINIPHÄNOMENTA Folgendes feststellen:

- Insgesamt ist das erreichte Niveau des Experimentierverhaltens der Schülerinnen und Schüler sehr niedrig: 70% erreichten lediglich die Verhaltenskategorie II (diffuses Experimentierverhalten) oder I (negative Reaktion).
- Die erreichte Verhaltenskategorie ist vom Geschlecht der Kinder unabhängig.
- In der dritten Klassen zeigen die Kinder häufiger negative Reaktionen (Stufe I) und seltener die Verhaltenskategorie IV (elaboriertes Experimentierverhalten).
- Höhere Kategorien werden eher beim Arbeiten in der Gruppe erreicht als allein.

Die Variable Einstellung gegenüber Phänomenen und ihrer Bearbeitung

Die folgende Tabelle gibt die wichtigsten statistischen Kennwerte des Fragebogens⁴³ wieder:

Bezeichnung	Wert der Verteilung
Arithmetisches Mittel	42,4
Konfidenzintervall (95%)	41,5 – 43,3
Modalwert	37
Standardabweichung	9,3
range	44

Tabelle 3.2.1.2.7: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Prätest;
- Basiswerte der gesamten Verteilung
Der Verfasser

Bei einem Maximalpunktwert von insgesamt 60 Rohpunkten erreichten die Schülerinnen und Schüler im Durchschnitt 42,3 Punkte. Damit weisen sie im Mittel eine positive bis durchschnittliche Einstellung gegenüber dem Untersuchungsgegenstand auf. Vergleicht man den erzielten Wert mit der Einschätzung der Lehrerinnen und Lehrer, die übereinstimmend vermuteten, die Kinder seien begeistert vom Experimentieren, hätten folglich eine extrem positive Einstellung, so zeigt sich ein sehr viel differenzierteres Bild, das von der These der Begeisterung weit entfernt ist.

Dies wird auch durch die weiteren Werte der zentralen Tendenz: Konfidenzintervall und Modalwert unterstützt. So liegt der Modalwert bei lediglich 37 Rohpunkten.

Die Werte zur Dispersion führen ebenfalls zur Verwerfung der These der Begeisterung. Die hohe Range von 44 und die hohe Standardabweichung von 9,3 Testpunkten zeigen, dass die Einstellung in der Stichprobe ein heterogenes Merkmal.

Die Überprüfung auf Normalverteilung erfolgte durch den Kolmogorov-Smirnov-Test, der hier einen Testwert von 1,7 ergibt, so dass von einer Normalverteilung der Werte ausgegangen werden kann. Mit einer Schiefe von - 0,1 kann eine leichte Rechtssteilheit der Verteilung festgestellt werden.

⁴³ Die nachfolgenden Werte für die Variable [Einstellung gegenüber Phänomen und ihrer Bearbeitung] basieren auf der Prätesterhebung in der Untersuchungsstufe I. In der Kontroll- und Experimentalgruppe konnten 436 Fragebögen ausgewertet werden.

Der erste Überblick über die Daten soll durch die Klassifizierung der Summe der individuell erreichten Testpunkte in sechs Kategorien vervollständigt werden. Hier stellt sich die Frage der zu wählenden cut-off-Kriterien. Diese lassen sich wie üblich nur inhaltlich rechtfertigen. Hier ist in Bezug auf die charakterisierte Verteilung zu bedenken, dass der größte Teil der Werte im mittleren bis oberen Bereich liegt. Folglich muss hier das größte Differenzierungsvermögen der cut-off-scores liegen. Darüber hinaus liegt in Bezug auf eine schulische Untersuchung eine Orientierung am Notensystem nahe, zumal das Notensystem darauf ausgelegt ist, im genannten Bereich der Verteilung gut zu differenzieren:

Nummer	Einstellungskategorie	cut-off-score (relativ/absolut)
1	Begeistert	100 – 95 / 60 – 57
2	Sehr positiv	94 – 85 / 56 – 51
3	Positiv	84 – 70 / 50 – 42
4	Ambivalent	69 – 55 / 41 – 33
5	Negativ	54 – 25 / 32 – 15
6	Sehr negativ	24 – 0 / 14 – 0

Tabelle 3.2.1.2.8: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Prätest;
- Einstellungskategorien mit cut-off-scores
Der Verfasser

Das folgende Diagramm zeigt die Verteilung der individuell erreichten Punktwertsummen in den definierten Kategorien:

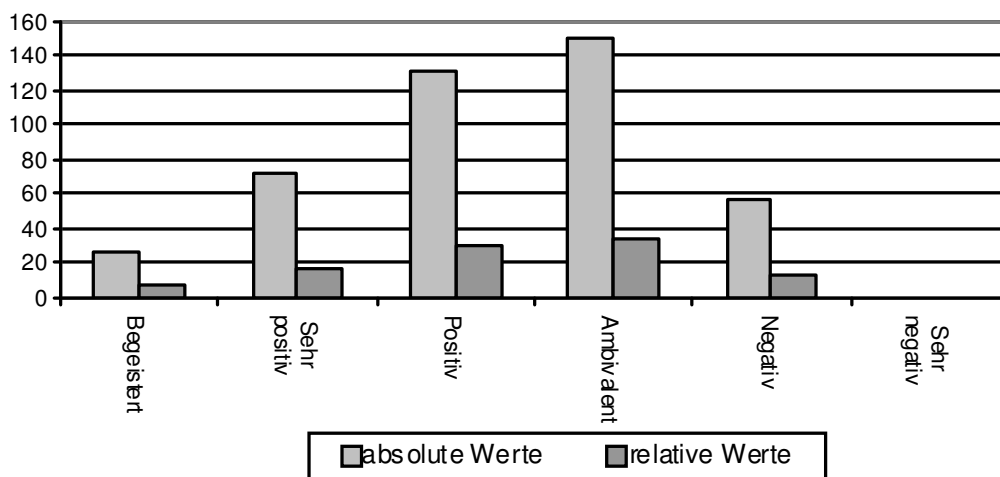


Diagramm 3.2.1.3.3: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Prätest;
- Häufigkeiten der Einstellungskategorien
Der Verfasser

Die Darstellung bestätigt das beschriebene Bild der [Einstellung] der Schülerinnen und Schüler. Die These der begeisterten Einstellung gegenüber dem Experimentieren ist auch durch diese Analyse widerlegt: Lediglich 6% der Schülerinnen und Schüler haben eine ‚begeisterte Einstellung‘.

Der Anteil der Schülerinnen und Schüler mit einer ‚ambivalenten Einstellung‘ ist hingegen bedenklich hoch. Insgesamt 47% der Kinder mussten dieser Kategorie zugewiesen werden. Positiv hingegen fällt auf, dass niemand eine ‚sehr negative, das heißt feindlich-ablehnende Einstellung gegenüber dem Experimentieren‘ aufweist.

Ihnen gegenüber stehen 46% der Schülerinnen und Schüler, die eine ‚sehr positive bis positive Einstellung gegenüber Phänomen und ihrer Bearbeitung‘ aufweisen.

Es können zum jetzigen Analysestand vier Erkenntnisse extrahiert werden:

1. Die Vermutung, die Schülerinnen und Schüler hätten überwiegend eine ‚begeisterte Einstellung‘, ist widerlegt.
2. Eine ‚feindselig ablehnende Einstellung gegenüber dem Experimentieren‘ ist nicht vorhanden.
3. Die Schülerschaft der dritten und vierten Klassenstufe ist in Bezug auf ihre [Einstellung gegenüber dem Experimentieren] zweigeteilt:
 - ‚sehr positive bis positive Einstellung‘
 - ‚ambivalente bis negative Einstellung‘

Beide Kategorienblöcke sind numerisch nahezu identisch besetzt, wobei die mittleren Kategorien, ‚positive und ambivalente Einstellung‘, mit jeweils ungefähr einem Drittel die am häufigsten vergeben Einstellungsklassen sind.

Nach der orientierenden Durchsicht der Daten, soll nun mit der vertiefenden Analyse begonnen werden. Diese befasst sich schwerpunktmäßig mit zwei Themenbereichen:

1. Die Komponenten der Einstellung: Analyse der kognitiven, affektiven und verhaltensbezogenen Dimension der Einstellung gegenüber dem Experimentieren.
2. Differentialpsychologische Analyse: Auswertung der Alters- und Geschlechtsspezifische Unterschiede im Antwortverhalten.

Das folgende Diagramm beschreibt die durchschnittlich erreichten relativen Punktwerte in den drei Dimensionen der Einstellung:

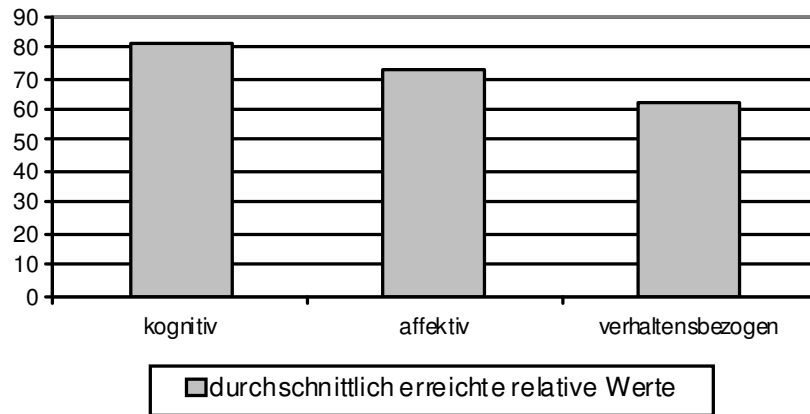


Diagramm 3.2.1.3.4: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Prätest;
- Komponenten der Einstellung
Der Verfasser

Es zeigen sich deutliche Unterschiede: In der [Einstellung] der Schülerinnen und Schüler dominiert die ‚kognitive Komponente‘. Ihm folgt mit etwa 10 Prozentpunkten weniger der ‚affektive Aspekt‘. Erst am Schluss, mit einer erneuten Differenz zum Vorherigen, kommt die ‚verhaltensbezogene Komponente‘ der Einstellung. Vor der inhaltlichen Interpretation dieses Umstandes soll zunächst eine Überprüfung der Unterschiede auf statistische Bedeutsamkeit erfolgen. Die folgende Tabelle listet die Differenzen numerisch auf. Fett gedruckte Werte sind hoch signifikant. Die Berechnung der statistischen Bedeutsamkeit erfolgt durch drei t-Tests für unabhängige Stichproben:

	Affektiv	Verhaltensbezogen	Kognitiv
Kognitiv	9,1 (df = 435 / t = 7,8)		
Affektiv		19,5 (df = 435 / t = 39,2)	
Verhaltensbezogen			10,4 (df = 435 / t = 36,1)

Tabelle 3.2.1.2.9: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Prätest;
- Signifikanzüberprüfung der Dimensionen der Einstellung
Der Verfasser

Alle Unterschiede sind hoch signifikant, können folglich inhaltlich interpretiert werden. Das Experimentieren ist für die Schülerinnen und Schüler somit vorrangig ein ‚kognitiv geprägter Prozess‘. Der ‚affektive Anteil der Einstellung‘ ist signifikant geringer ausgeprägt. Dies ist aus einem Alltagsverständnis heraus evident. Experimente werden vorrangig vorgenommen, um über Sachverhalte kontrolliert nachdenken zu können. Auch deckt sich dieses Ergebnis mit anderen Untersuchungen (vgl. Dengler 1995, S. 26 ff.). Aus physikdidaktischer Sicht erscheint die geringe affektive Betonung jedoch eher bedenklich, sollen die Kinder in der Primarstufe doch einen spielerisch orientierten, emotionalen Zugang zur Physik bekommen. Nur so kann ein langfristiges Interesse und damit eine tiefgehende Beschäftigung mit physikalischen Fragestellungen erwartet werden. Noch bedenklicher ist die geringe Ausprägung der ‚Verhaltensbezogenen Komponente‘. Experimente scheinen für die Schülerinnen und Schüler weniger etwas zu sein, bei dem man aktiv handelt, als vielmehr eine Art der passiven Rezeption. Damit fehlen, dies machen die theoretischen Erörterungen in Abschnitt 1.6 in dieser Arbeit deutlich, wesentliche Aspekte der Tätigkeit des Experimentierens.

Ich fahre mit der **differentialpsychologischen Analyse** fort:

Die folgende Grafik zeigt den Zusammenhang des Geschlechtes mit zentralen Fragebogenparametern:

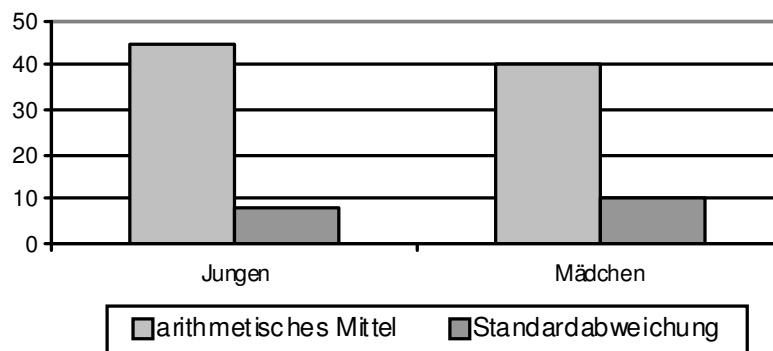


Diagramm 3.2.1.3.5: Einstellung gegenüber Phänomen und ihrer Bedeutung; Prätest;
- der Einfluss des Geschlechtes auf die Einstellung
Der Verfasser

Es ergibt sich ein deutlicher Einfluss des Geschlechts auf die Einstellung der Schülerinnen und Schüler: Der Vergleich der Mittelwerte erfolgt durch einen t-Test für unabhängige Stichproben. Es ergibt sich bei 187 Freiheitsgraden ein t-Wert von 3,25: Dieser ist signifikant. Folglich ist die Einstellung der Jungen statistisch bedeutsam positiver als die der Mädchen. Auch streuen die individuell erreichten kumulierten Testpunkte bei den Jungen weniger als bei den Mädchen. Dieser Unterschied ist allerdings statistisch nicht bedeutsam.

Nun soll die Frage nach dem Einfluss des Alters auf die Einstellung geklärt werden. Das folgende Diagramm gibt eine erste Orientierung: Die Altersklassen sieben und acht Jahre, sowie zehn und elf Jahre wurden zusammengefasst, da systembedingt für eine explizite Auswertung zu wenige Sieben- und zu wenige Elfjährige an der Untersuchung teilgenommen haben:

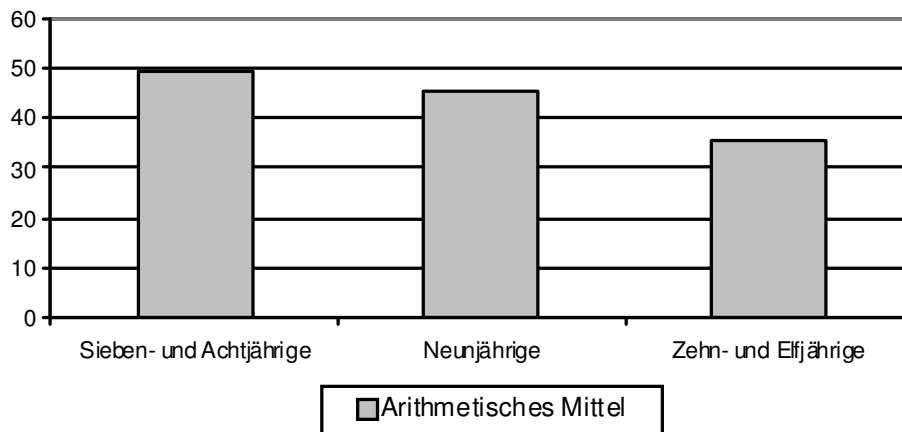


Diagramm 3.2.1.3.6: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Prätest;
- der Einfluss des Alters auf die Einstellung
Der Verfasser

Mit zunehmendem Alter wird die [Einstellung] der Schülerinnen und Schüler negativer, wobei der größte Sprung zwischen den Altersklassen neun und zehn bis elf Jahren liegt. Die Differenzen werden durch drei t-Tests für unabhängige Stichproben untersucht. Die folgende Tabelle zeigt die Ergebnisse:

	Differenz zwischen 7 bis 8 und 9 Jahren	Differenz zwischen 9 und 10 bis 11 Jahren	Differenz zwischen 7 bis 8 und 10 bis 11 Jahren
Differenz	4,3	9,8	14,1
Freiheitsgrade	119	119	132
t-Wert	4,8	12,5	14,4

Tabelle 3.2.1.2.10: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Prätest;
- Signifikanzüberprüfung des Einflusses des Alters
Der Verfasser

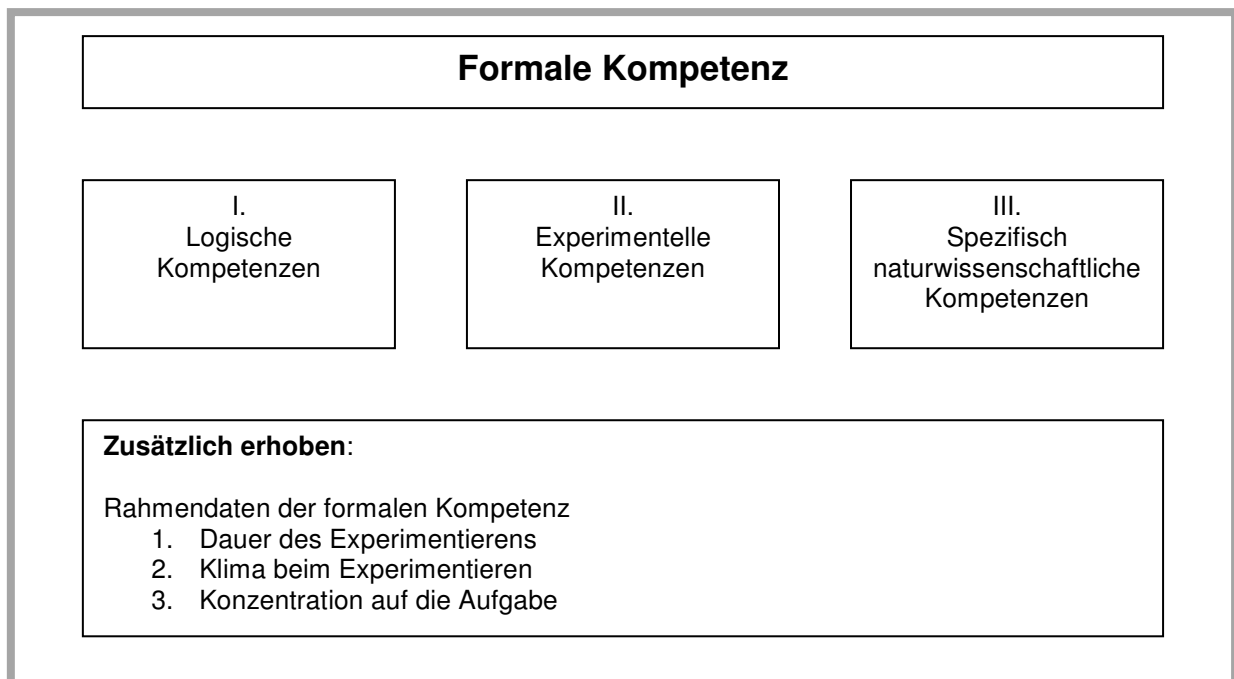
In allen Fällen muss die H_1 angenommen werden. Die Unterschiede sind statistisch hoch bedeutsam. Die [Einstellung] ist stark altersabhängig. Ist sie bei den Sieben- bis Achtjährigen im Durchschnitt noch im positiven Bereich, rutscht sie bis zum Alter von zehn bis elf Jahren im Mittel in den ambivalenten Bereich ab.

Insgesamt führten die vorgenommenen Analysen zu einer Intensivierung des problematischen Eindrucks aus der Exploration der Daten:

1. Im Vergleich zu den ‚kognitiven Aspekten der Einstellung‘ geringe Ausprägung der ‚affektiven und verhaltensbezogenen Dimension‘ beim Experimentieren.
2. Mädchen weisen eine signifikant negativere [Einstellung] auf.
3. Mit steigendem Alter wird die [Einstellung] gravierend negativer.

Die Variable formale Kompetenz

Ich beginne die Auswertung zur besseren Übersicht über den folgenden Textteil mit einem Überblick über die mit dieser Variablen erfassten Gegenstände:



Grafik 3.2.1.2.1: Formale Kompetenz; Prätest;
- Erfassungsbereich der Variablen
Der Verfasser

Zu den Rahmendaten des Experimentierens:⁴⁴. Die folgenden Diagramme geben die zentralen Werte wieder:

⁴⁴ Die folgenden Auswertungsschritte für die Variable [formale Kompetenz] basieren auf der Prätesterhebung in der Experimental- und Kontrollgruppe der Untersuchungsstufe I. Insgesamt konnten 159 Beobachtungen gemacht werden.

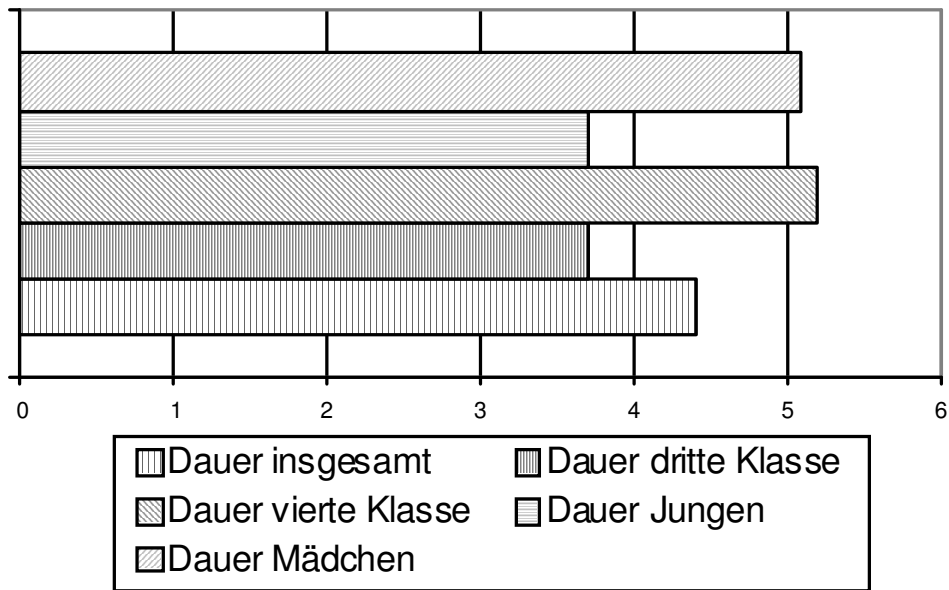


Diagramm 3.2.1.3.7: Formale Kompetenz; Prätest;
- Dauer des Experimentierens
Der Verfasser

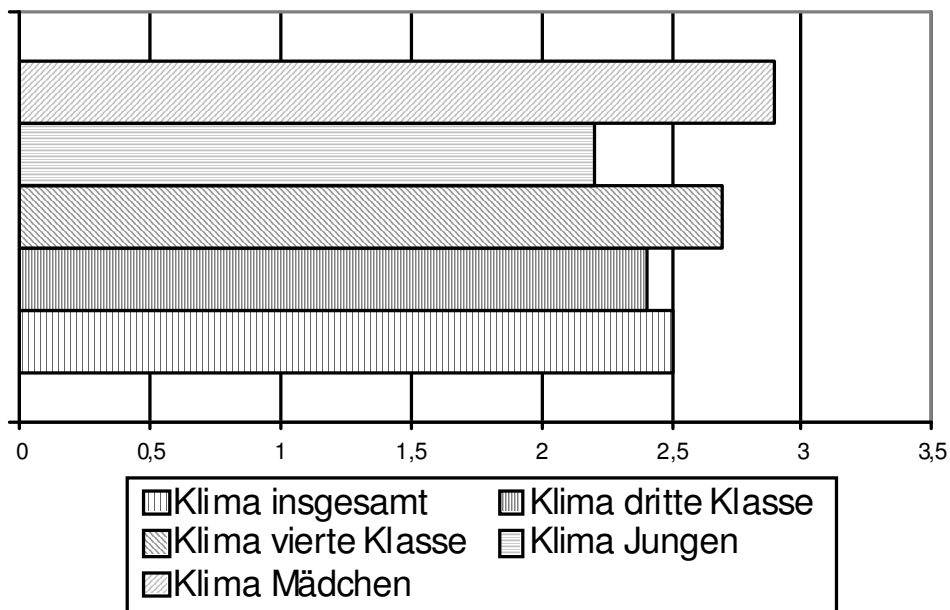


Diagramm 3.2.1.3.8: Formale Kompetenz; Prätest;
- Klima beim Experimentieren
Der Verfasser

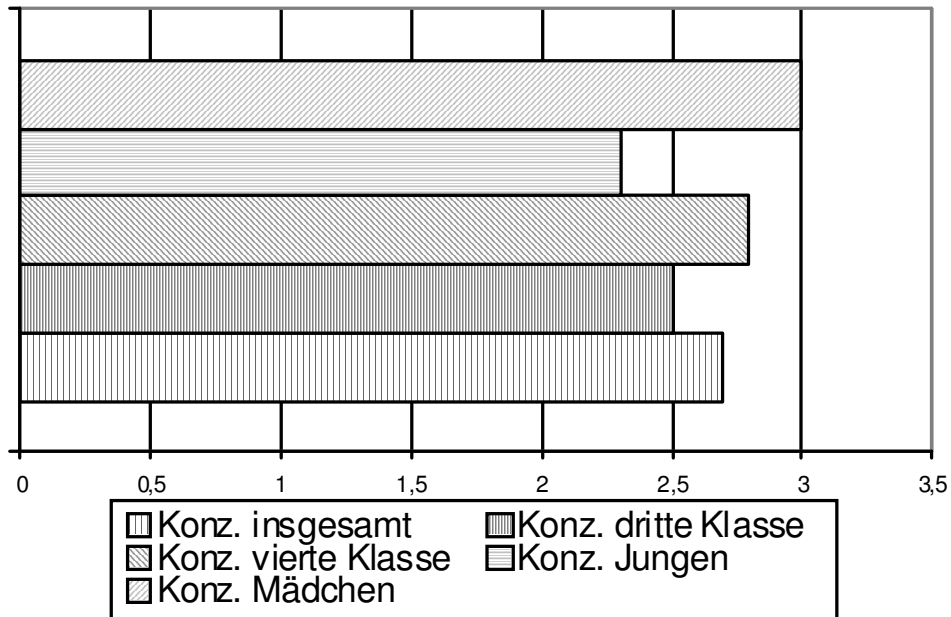


Diagramm 3.2.1.3.9: Formale Kompetenz; Prätest;
- Konzentration auf das Experiment
Der Verfasser

Vergleicht man die Dauer des Experimentierens von durchschnittlich 4,4 Minuten mit den durchschnittlichen Verweilzeiten an Stationen in Science-Centern oder mit denen der MINIPHÄNOMENTA, so könnte der Eindruck eines überdurchschnittlich langen Experimentierens entstehen. Dies ist jedoch nicht der Fall. Zunächst ist das Setting in beiden Fällen ein grundsätzlich Unterschiedliches. Kann beim Besuch eines Science-Centers oder in Bezug auf die MINIPHÄNOMENTA völlig frei experimentiert werden, ist die Situation in der Untersuchung ein klassisches laborpsychologisches Setting: Die Schülerinnen und Schüler erhalten eine feste Aufgabenstellung, experimentieren unter Beobachtung und sollen ihre Gedanken zum Experiment äußern. Sie sollen eine Erklärung finden, das heißt Hypothesen äußern, diese überprüfen und schließlich die Ergebnisse in einer Theorie komprimieren. Bei den 4,4 Minuten handelt es sich damit zum einen nicht um reine Experimentierzeit, zum anderen ist die Aufgabe im Vergleich zu denen in den anderen genannten Settings deutlich anspruchsvoller. Bedenkt man die Aspekte, das Setting und die Komplexität der Aufgabe, so ist die Dauer des Experimentierens mit 4,4 Minuten als überaus niedrig einzustufen.

Das Gruppenklima und die Konzentration auf die Aufgabe liegen mit Werten von 2,5 und 2,7 im mittleren bis leicht erhöhten Bereich. Dieser Umstand ist als positiv zu bewerten.

In Bezug auf das Geschlecht ergibt sich ein deutlicher Trend: Die Werte der Mädchen sind deutlich höher als die der Jungen. Mit Hilfe eines t-Testes für unabhängige Stichproben sollen die Unterschiede auf statistische Bedeutsamkeit überprüft werden. Die Ergebnisse sind in der folgenden Tabelle wiedergegeben:

	Dauer	Klima	Konzentration
Differenz	1,4	0,7	0,7
Freiheitsgrade	77	77	77
t-Wert	-4,0	-3,88	-3,2

Tabelle 3.2.1.2.11: Formale Kompetenz; Prätest;
- Signifikanzüberprüfung der Geschlechtsunterschiede
Der Verfasser

In allen Fällen kann die Forschungshypothese angenommen werden. Die Mädchen arbeiteten länger an der Aufgabe, das Klima bei der Gruppenarbeit war positiver und sie konzentrierten sich stärker auf die Aufgabe. Die Ergebnisse stehen im deutlichen Widerspruch zu denen der [Einstellung]: Die Mädchen haben eine ‚negativere Einstellung gegenüber Phänomenen und ihrer Bearbeitung‘, scheinen aber Jungen in ihren Leistungen, darauf deuten diese Ergebnisse hin, überlegen. Für eine endgültige Entscheidung über die Annahme dieser These sind die bisher vorgebrachten Beweise nicht aussagekräftig genug. Im weiteren Verlauf werde ich auf diesen Aspekt weiter eingehen.

Abschließend soll der Einfluss der Rahmendaten auf das Ergebnis der Untersuchung ermittelt werden: Es wäre evident anzunehmen, dass diejenigen Schülerinnen und Schüler bessere Ergebnisse erzielen, die sich viel Zeit nehmen, gut zusammenarbeiten und sich weitgehend auf die Aufgabe konzentrieren. Überprüft wird diese Hypothese durch eine Produkt-Moment-Korrelation, deren Ergebnisse in der folgenden Tabelle wiedergegeben werden:

	Dauer - erreichter Testwert	Klima - erreichter Testwert	Konzentration - erreichter Testwert
Korrelationsindex	0,8	0,8	0,7

Tabelle 3.2.1.2.12: Formale Kompetenz; Prätest;
- Zusammenhang der Rahmendaten mit der eigentlichen formalen Kompetenz
Der Verfasser

Die H_1 kann in Bezug auf alle Variablen angenommen werden. Es gibt einen stark positiven Zusammenhang, d.h. ein hoher Punktwert geht einher mit einer langen Experimentierdauer, einem guten Klima und einer hohen Konzentration auf die Aufgabe.

Zum jetzigen Stand der Analyse können folgende Ergebnisse festgehalten werden:

- Die durchschnittliche Dauer der Versuchsdurchführung ist mit 4,4 Minuten sehr gering. Gruppenklima und Konzentration bei der Bearbeitung sind im mittleren bis leicht erhöhten Bereich.
- Es gibt deutliche Unterschiede der formalen Kompetenz zwischen Jungen und Mädchen. Die Mädchen experimentierten länger, waren konzentrierter und das Gruppenklima war bei ihnen besser.
- Es gibt einen ausgeprägt positiven Zusammenhang zwischen hohen Rahmendaten und hohen formalen Kompetenz.

Ich fahre in der Auswertung mit der Analyse der eigentlichen [formalen Kompetenz] (vgl. Grafik 3.2.1.2.1 in diesem Abschnitt) fort. Die folgende Tabelle fasst einige deskriptive Werte zusammen:

Bezeichnung	Wert der Verteilung
--------------------	----------------------------

Arithmetisches Mittel	33,4
Konfidenzintervall (95%)	31,9 - 34,8
Modalwert	29
Standardabweichung	9,2
range	39

Tabelle 3.2.1.2.13: Formale Kompetenz; Prätest;
- deskriptive Werte zu der formalen Kompetenz
Der Verfasser

Die Werte zeigen ein bedenkliches Bild der [formalen Kompetenz] der Schülerinnen und Schüler. Sie erreichen im Durchschnitt lediglich 33,4 Rohpunkte, das heißt 49% der Gesamtpunktzahl. Die Streuung der Werte ist relativ groß. Dies beweisen die hohe Standardabweichung und die große range.

Bei der Gesamtverteilung der kumulierten Punkte handelt es sich, dies beweist der Kolmogorov-Smirnov-Test mit einem z-Wert von 0,8, um eine Normalverteilung.

Zur vertiefenden inhaltlichen Interpretation der [formalen Kompetenz] der Schülerinnen und Schüler sollen ihre Leistungen Kompetenzstufen zugewiesen werden, deren cut-off-scores sich erneut am Schulnotensystem orientieren:

Nummer	Kompetenzstufen	cut-off-score (relativ/absolut)
--------	-----------------	---------------------------------

1	sehr hohe formale Kompetenz	100 – 90 / 68 - 61
2	ausgeprägte formale Kompetenz	89 – 70 / 60 - 48
3	durchschnittliche formale Kompetenz	69 – 45 / 47 - 31
4	unbefriedigende formale Kompetenz	44 – 0 / 30 - 0

Tabelle 3.2.1.2.14: Formale Kompetenz; Prätest;
- Kompetenzstufen mit cut-off-scores
Der Verfasser

Das folgende Diagramm zeigt die absoluten und relativen Kategorienhäufigkeiten:

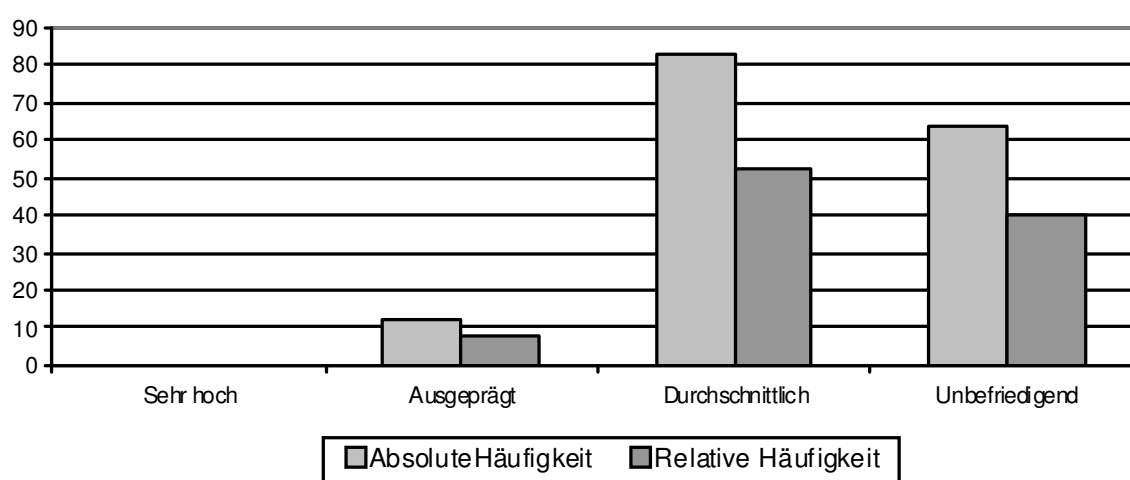


Diagramme 3.2.1.3.10: Formale Kompetenz; Prätest;
- relative und absolute Kategorienhäufigkeiten
Der Verfasser

Das Diagramm verstärkt den problematischen Eindruck der [formalen Kompetenz] der Schülerinnen und Schüler. Kein Kind konnte der Kategorie ‚sehr hohe formale Kompetenz‘ zugewiesen werden. Lediglich 12 Schülerinnen und Schüler weisen eine ‚ausgeprägte formale Kompetenz‘ auf. Etwas mehr als die Hälfte (52%) der Kinder erreichte die Kategorie ‚durchschnittliche formale Kompetenz‘. Der problematischste Wert betrifft die Kategorie ‚unbefriedigende formale Kompetenz‘: 40% der Schülerinnen und Schüler mussten dieser Kategorie zugewiesen werden.

Folgende weitere Ergebnisse können festgehalten werden:

- Die [formale Kompetenz] der Schülerinnen und Schüler ist normalverteilt.
- Die [formale Kompetenz] ist als bedenklich niedrig einzustufen. Das arithmetische Mittel der Verteilung liegt bei 33,4 Testrohpunkten. Damit erreichten die Schülerinnen und Schüler im Durchschnitt lediglich 49% der Gesamtpunktzahl.
- Auch die Einteilung der individuell erreichten kumulierten Testpunkte in Kategorien ergibt ein negatives Bild: 40% der Schülerinnen und Schüler mussten der Kategorie ‚unbefriedigende formale Kompetenz‘ zugerechnet werden.

Bei den bisherigen Analysen wurden die [formale Kompetenz] der Schülerinnen und Schüler insgesamt betrachtet. Es handelt sich hierbei allerdings um ein äußerst komplexes Merkmal, das im Falle der vorliegenden Untersuchung durch drei Dimensionen operationalisiert worden ist. Zur vertiefenden Interpretation sollen die Ergebnisse in den einzelnen Dimensionen untersucht werden. Die folgende Tabelle nennt die zentralen Kennwerte der einzelnen Dimensionen:

	Logisches Denken	Spezifisch naturwissenschaftliches Denken	Experimentelle Kompetenzen
Arithmetisches Mittel	12,0	6,7	15,1
Konfidenzintervall (95%)	11,4 – 12,6	6,4 – 7,1	14,4 – 15,7
Modalwert	14	4	17
Standardabweichung	3,8	2,1	4,3
range	15	10	17

Tabelle 3.2.1.2.14: Formale Kompetenz; Prätest;
- Kennwerte der einzelnen Dimensionen
Der Verfasser

In den Dimensionen spiegelt sich das beschriebene Bild: Im Durchschnitt erreichten die Schülerinnen und Schüler auch hier etwa die Hälfte der erreichbaren Punktzahl. In keiner Dimensionen weichen die Ergebnisse gravierend vom beschriebenen Bild ab. Dennoch gibt es kleinere Unterschiede, die in der folgenden Tabelle wiedergegeben sind. Sie beschreibt die Differenzen der durchschnittlichen erreichten relativen Punktsommen, sowie deren Signifikanzüberprüfung, die durch einen t-Test für unabhängige Stichproben erfolgt:

	Experimentelle Kompetenzen	Spezifisch naturwissenschaftliches Denken	Logisches Denken
Durchschnittlich erreichte relative Punktsomme Logisches Denken	(53,3%) 3,5 (df = 158 / t = -17,7)	(41,0%) _____	(50,0%) _____
Experimentelle Kompetenzen	_____	12,5 (df = 158 / t = 28,9)	_____
Spezifisch naturwissenschaftliches Denken	_____	_____	9,0 (df = 158 / t = 22,8)

Tabelle 3.2.1.2.15: Formale Kompetenz; Prätest;
- Differenzen zwischen den Dimensionen
Der Verfasser

In allen drei Fällen muss die H_1 angenommen werden. Die höchsten Werte erreichten die Schülerinnen bei den ‚experimentellen Kompetenzen‘. Im Durchschnitt wurden hier 53% der verfügbaren Punkte erzielt. Bei den ‚logischen Fähigkeiten‘ wurden im Durchschnitt drei Prozentpunkte weniger erreicht. Deutlich schlechter waren die Leistungen im Bereich des ‚spezifisch naturwissenschaftlichen Denkens‘, der sich inhaltlich mit der Frage des Beobachtens und mit den Grundsätzen des Experimentes befasst. Gerade das genaue Beobachten der auftretenden Phänomene fiel den Kindern sichtlich schwer. Im Durchschnitt erreichten sie auf diesem Item lediglich 1,9 Punkte. Dieses Ergebnis wird bei der Auswertung der Variablen [Wissen] im späteren Textverlauf noch eine Rolle spielen.

Die Unterschiede zwischen den Dimensionen sind signifikant ausgeprägt: Es ergibt sich von schwer zu leicht die folgende Rangreihe: ‚spezifisch naturwissenschaftliches Denken‘, ‚logisches Denken‘ und ‚experimentelle Fähigkeiten‘.

Abschließend zur **differentialpsychologischen Analyse**: Es soll dabei den Fragen des Einflusses von Geschlecht und Klassenzugehörigkeit auf die Leistungen in der Untersuchung nachgegangen werden. Die folgende Tabelle gibt das arithmetische Mittel und die Standardabweichung in Abhängigkeit von Alter und Geschlecht, sowie die Werte der Signifikanzüberprüfung an, die mit Hilfe eines t-Testes für unabhängige Stichproben durchgeführt wurde:

	Arithmetisches Mittel	Standardabweichung	Freiheitsgrade	t-Wert
Dritte Klasse	31,2	9,3	77	-2,8
Vierte Klasse	35,4	8,8	77	
Jungen	30,5	9,6	77	-3,6
Mädchen	36,1	8,1	77	

Tabelle 3.2.1.2.16: Formale Kompetenz; Prätest;
- differentialpsychologische Betrachtung
Der Verfasser

Die Daten zeigen einen eindeutigen Trend. Sowohl die Klassenstufe, als auch das Geschlecht der Kinder haben Einfluss auf das Ergebnis der Untersuchung. Kinder der vierten Klasse lösten die Aufgabe signifikant besser als jene der Klassenstufe drei. Dieser Gedanke ist evident, sind die Schülerinnen und Schüler in der vierten Klasse doch älter, verfügen über einen größeren Erfahrungsschatz und haben mehr schulisches Wissen. Das bessere Abschneiden der Mädchen bestätigt die eingangs in diesem Kapitel geäußerte These der höheren formalen Kompetenz der Mädchen gegenüber den Jungen.

Ergebnisse der differentialpsychologischen Betrachtung:

Schülerinnen und Schüler der vierten Klasse lösen die Aufgabe signifikant besser als solche der Klassenstufe drei.

Mädchen erzielen in Bezug auf die [formale Kompetenz] signifikant bessere Ergebnisse als die Jungen.

Die Variable aktuelle Motivation

Die Erhebung der [aktuellen Motivation] geschieht lediglich in der vierten Klassenstufe. Der Prätest ergab, dass die Motivationsstruktur bei Schülerinnen und Schülern in der dritten Klasse noch nicht differenziert genug ist, um mit der Methode des FAM gemessen zu werden. Insofern besitzen die im Folgenden gemachten Aussagen nur für die vierte Klasse Gültigkeit.

Ich beginne die Auswertung mit einer Tabelle, in der einige wichtige deskriptive Daten⁴⁵ zusammengestellt werden:

Bezeichnung	Wert der Verteilung
Arithmetisches Mittel	45,6
Konfidenzintervall (95%)	44,4 – 46,8
Modalwert	41,0
Standardabweichung	9,0
range	41

Tabelle 3.2.1.2.17: Aktuelle Motivation; Prätest;
- deskriptive Werte zu zur aktuellen Motivation
Der Verfasser

⁴⁵ Die folgende Auswertung für die Variable [aktuelle Motivation] basiert auf den Daten des Prätests der ersten Erhebungsstufe. Insgesamt konnten 225 Bögen ausgewertet werden.

Die Werte der Verteilung ähneln deutlich denen der Einstellung. Im Durchschnitt erreichten die Schülerinnen und Schüler im Fragebogen zur Erfassung der [aktuellen Motivation] 45,6 Rohpunkte und damit 69% der erreichbaren Punkte. Es scheint damit eine akzeptable, wenn auch nicht hohe Motivation, eine Erklärung für den folgenden Versuch zu finden, vorhanden zu sein. Die Streuung der Verteilung ist mit einer Standardabweichung von 9,1, beziehungsweise einer Range von 41 - ähnlich wie die der [Einstellung] - als hoch einzustufen. Es gibt folglich viele Schülerinnen und Schüler, die von diesem Wert unterschiedlich stark abweichen.

Die Verteilung der kumulierten Testwerte ist mit einem z-Wert von 0,8 im Kolmogorov-Smirnov-Test, eine Normalverteilung. Mit einer Schiefe von -0,2 ist sie als rechtsschief zu klassifizieren.

Um die inhaltliche Interpretation der Daten weiter zu vertiefen, werden die individuell erreichten kumulierten Testpunkte Kategorien zugewiesen. Dafür orientiere ich mich an den Erläuterungen zur [Einstellung] im obigen Textverlauf. Analog hierzu definiere ich vier Motivationslagen. Die cut-off-scores sollen jedoch in Anbetracht der deutlicheren Rechtssteilheit der Verteilung etwas strenger gewählt werden, um eine gute Differenzierung der Kategorien zu erreichen:

Nummer	Stufen der aktuellen Motivation	cut-off-score (relativ/absolut)
1	Begeistert und hoch motiviert	100 - 90 / 66 - 59
2	Überdurchschnittlich motiviert	89 - 75 / 58 - 50
3	Motiviert	74 - 50 / 49 - 33
4	Gering motiviert	49 - 0 / 32 - 0

Tabelle 3.2.1.2.18: Aktuelle Motivation; Prätest;
- Motivationsstufen mit cut-off-scores
Der Verfasser

In dem folgenden Diagramm wird die sich auf den cut-off-scores ergebende Verteilung wiedergegeben:

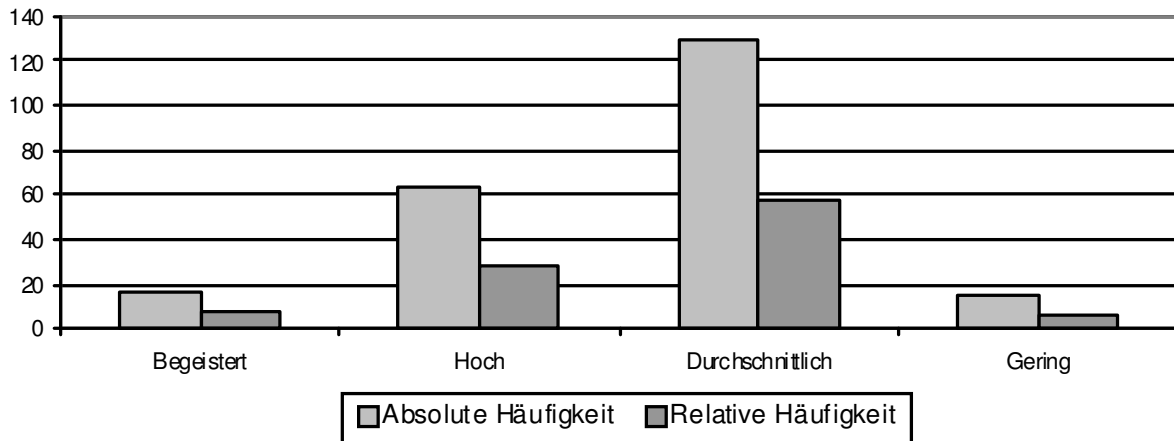


Diagramme 3.2.1.3.11: Aktuelle Motivation; Prätest;
- relative und absolute Kategorienhäufigkeiten
Der Verfasser

Die Verteilung bestätigt den im obigen Textverlauf angedeuteten positiven Eindruck in Bezug auf die [aktuelle Motivation] der Schülerinnen und Schüler, die Aufgabe zu lösen. Lediglich 15 Schülerinnen und Schüler, d.h. 7% sind in Bezug auf die Aufgabe ‚gering motiviert‘. Der größte Anteil zeigt eine ‚durchschnittliche Motivation‘. Insgesamt 36% der Kinder weisen sogar eine ‚hohe, beziehungsweise begeisterte Motivation‘ auf. Vergleicht man diese Werte mit den vorherigen zum [allgemeinen Experimentierverhalten], der [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] und der [formalen Kompetenz], so ergibt sich eine deutliche Unstimmigkeit: Das [allgemeine Experimentierverhalten] war überwiegend wenig elaboriert, ein großer Teil der Schülerinnen und Schüler weisen eine ‚negative Einstellung‘ auf und auch die [formale Kompetenz] ist insgesamt wenig entwickelt, in Teilen sogar defizitär. Wie lässt sich nun vor diesem Hintergrund die eben beschriebene Motivationslage verstehen? Sie zeichnet sich insgesamt weniger dadurch aus, dass sie im Durchschnitt überragend hoch ist, als vielmehr dadurch, dass nur wenige Kinder eine ‚unterdurchschnittliche Motivation‘ aufweisen.

Die Ergebnisse belegen eindeutig, dass bei den Schülerinnen und Schülern durchaus der Wille besteht ein Experiment zu bearbeiten. Sie sind motiviert mit den Versuchen umzugehen und über sie nachzudenken. Bislang wurden sie in diesem Bereich jedoch nur unzureichend gefördert, daher die ‚geringen Kompetenzen‘, die ‚wenig elaborierte Einstellung‘ usw.

Zum jetzigen Stand der Analyse können zwei Ergebnisse festgehalten werden:

1. Die individuell erreichten kumulierten Testwerte der Schülerinnen und Schüler bilden eine rechtssteile Normalverteilung.
2. Im Vergleich zu den bisher dargestellten Merkmalen sind die Ergebnisse der Motivationsanalyse überraschend positiv. Nur ein sehr geringer Teil der Schülerinnen und Schüler weist eine ‚unterdurchschnittliche Motivation‘ auf. Mehr als ein Drittel sind ‚begeistert beziehungsweise hoch motiviert‘.

Nachdem die groben Tendenzen der Daten herausgearbeitet wurden, sollen nun die einzelnen Dimensionen der [aktuelle Motivation] (‚Interesse‘, ‚Kompetenzerwartung‘ und ‚Misserfolgsbefürchtung‘) genauer betrachtet werden. Das folgende Diagramm gibt den durchschnittlich erreichten Punktwert für jede Dimension wieder:

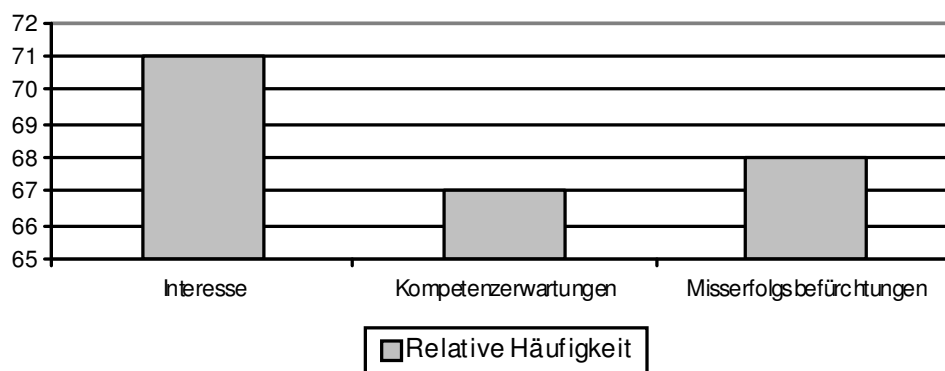


Diagramme 3.2.1.3.12: Aktuelle Motivation; Prätest;
- relative Häufigkeit in den unterschiedlichen Dimensionen⁴⁶
Der Verfasser

Die durchschnittlich erreichte relative Punktzahl in den Kategorien ‚Kompetenzerwartung‘ und ‚Misserfolgsbefürchtung‘ ist mit lediglich einem Prozent Differenz nahezu identisch, so dass eine inhaltliche Interpretation entfallen muss. Das ‚Interesse‘ ist nach einem t-Test für unabhängige Stichproben (Interesse - Kompetenzerwartung: $t = 3,5$, $df = 224$; Interesse - Misserfolgsbefürchtung: $t = -30,6$, $df = 224$) signifikant stärker ausgeprägt als die Werte in den übrigen Dimensionen. Das ‚Interesse‘ wird stärker als die übrigen zwei Dimensionen durch die Attraktivität des Versuches moderiert. Sind die ‚Kompetenzerwartung‘ sowie die ‚Misserfolgsbefürchtung‘ sehr stark auf die Aufgabe, das heißt auf das Finden einer Erklärung bezogen, so wird das ‚Interesse‘ stärker durch den für die Schülerinnen und Schüler spannenden Versuch beeinflusst.

⁴⁶ Bei der inhaltlichen Interpretation der Dimension ‚Misserfolgsbefürchtung‘ muss bedacht werden, dass ein hoher Punktwert geringe ‚Misserfolgsbefürchtungen‘ der Schülerinnen und Schüler abbildet.

Die differenzialpsychologische Analyse bezieht sich im Falle der [aktuelle Motivation] lediglich auf die vierte Klassenstufe, so dass hier nur der Einfluss des Geschlechts auf die individuell erreichten Rohpunkte ermittelt werden muss. Die folgende Tabelle stellt diesen Zusammenhang dar:

	Arithmetisches Mittel	Standardabweichung	t-Wert	Freiheitsgrade
Junge	50,2	7,0	10,28	103
Mädchen	41,6	8,5		

Tabelle 3.2.1.2.19: Aktuelle Motivation; Prätest;
- Einfluss des Geschlechtes auf die Motivation
Der Verfasser

Es besteht ein signifikanter Geschlechtsunterschied zwischen Jungen und Mädchen. Die Jungen haben eine um 8,6 Testpunkte höhere [aktuelle Motivation] die Aufgabe zu bearbeiten als die Mädchen. Dieses Ergebnis setzt den bisherigen Trend der Daten fort. In Bezug auf Fragen der Selbsteinschätzung [Einstellung und aktuelle Motivation] äußern sich die Jungen positiver als die Mädchen. Sie haben eine positivere [Einstellung] und eine höhere [aktuelle Motivation]. In Bezug auf die Fähigkeiten sind die Mädchen hingegen den Jungen überlegen. Sie erreichen bei der ‚formalen Kompetenz‘ signifikant höhere Werte.

In Bezug auf die Auswertung der Variablen [aktuelle Motivation] können abschließend zwei weitere Ergebnisse festgehalten werden.

1. In Bezug auf die Dimensionen der [aktuelle Motivation] sticht das ‚Interesse‘ hervor. Der durchschnittlich erreichte Punktwert ist hier höher als bei den ‚Kompetenzerwartungen‘ und den ‚Misserfolgsbefürchtungen‘. Dies ist durch eine stärkere Moderation dieser Dimension durch den Versuch zu erklären.
2. Jungen haben eine signifikant höhere [aktuelle Motivation] als Mädchen.

Die Variable Wissen

Die folgende Tabelle stellt die wichtigsten deskriptiven Daten⁴⁷ für die einzelnen Variablen dar:

⁴⁷ Die folgenden Aussagen und Interpretationen basieren auf insgesamt 206 ausgewerteten Gruppenarbeiten der Erhebungen der Untersuchungsstufe I (Kontrollgruppe und Experimentalgruppe I) der Variablen [Wissen].

	Kategorien					Indices	
	V ⁴⁸	ARP	Q	AI	AP	Index I (Qualität)	Index II (Quantität)
Arithmetisches Mittel	1,5	3,2	2,3	1,5	12,7	1,7	4,8
Konfidenzintervall (95%)	1,4	3,0	2,1	1,3	12,3	1,6	4,6
	-	-	-	-	-	-	-
	1,7	3,4	2,4	1,6	13,2	1,8	4,9
Modalwert	1	3	2	1	15	1,4	5,3
range	5	8	5	1,1	15	4,4	7,0
Standardabweichung	0,9	1,2	1,0	1,0	3,1	0,6	1,2

Tabelle 3.2.1.2.20: Wissen; Prätest;
- deskriptive Werte
Der Verfasser

Die Werte ergeben in Bezug auf das [Wissen] der Schülerinnen und Schüler ein höchst bedenkliches Bild. Im Durchschnitt konnten Sie nur drei von fünfzehn Kärtchen (vgl. Abschnitte 1.4.7 und 2.4.7) richtig legen, das heißt lediglich 22% der Aufgabe lösen. Diesem korrekten Teil folgten im Schnitt 1,5 Inseln, das heißt Bereiche in denen die Kärtchen in der richtigen Reihenfolge in einem falschen Umfeld lagen. Die Qualität dieser Inseln war im Durchschnitt sehr gering. Sie liegt mit 2,3 Punkten deutlich unter dem Durchschnitt der Skala. Noch schlechter war die Verknüpfung der Inseln untereinander, beziehungsweise die Anknüpfung an die falschen Teile der Aufgabe. Hier erreichten die Schülerinnen und Schüler im Durchschnitt lediglich 1,5 Punkte. Diese negativen Ergebnisse schlagen sich vor allem im Index I (‚Qualität‘) nieder. Im Durchschnitt liegt er bei lediglich 1,7 von 6,8 verfügbaren Punkten. Der Index II (Quantität) hingegen ist relativ hoch: Von 7,8 Indexpunkten erreichten die Untersuchungsteilnehmerinnen und Untersuchungsteilnehmer im Durchschnitt 4,8 Punkte. Dies ist vor allem dadurch bedingt, dass der größte Teil der Schülerinnen und Schüler sich bemühte möglichst viele Teile im Durchschnitt 12,7 zu verwenden.

Es liegt folglich eine Differenz zwischen den beiden Indices ‚Qualität‘ und ‚Quantität‘ vor: Die Schülerinnen und Schüler bemühten sich um eine möglichst vollständige Bearbeitung der Aufgabe, die ‚Qualität der Lösung‘ war jedoch insgesamt sehr gering.

Im Folgenden soll durch den Kolmogorov-Smirnov-Test die Verteilungsform der beiden Indizes überprüft werden. Die Ergebnisse sind in der folgenden Tabelle zusammengefasst:

⁴⁸ Verzeichnis der in der Tabelle verwendeten Abkürzungen:

1. V = Verknüpfung der Propositionen
2. ARP = Anzahl der richtigen Propositionen
3. Q = Qualität der Inseln
4. AI = Anzahl der Inseln
5. AP = Anzahl der Propositionen

	Index I	Index II
Z- Wert	2,14	2,66
Sig.	0,00	0,00

Tabelle 3.2.1.2.21: Wissen; Prätest;
- Test auf Normalverteilung der Indexvariablen
Der Verfasser

In beiden Fällen muss folglich die H_1 angenommen werden, das heißt beide Verteilungen weichen hoch signifikant von der Normalverteilung ab. Es handelt sich um eingipflige, deutlich linksschiefe (Index I) beziehungsweise rechtsschiefe (Index II) Verteilungsformen.

Wie bereits bei den anderen Variablen geschehen, werden die Messwertreihen im Folgenden Kategorien zugewiesen, um eine vertiefende inhaltliche Analyse vornehmen zu können. Ich beginne mit dem Index I (Qualität):

Nummer	Kategorien des Index I	cut-off-score (relativ/absolut)
1	Sehr hohe inhaltliche Qualität	100% - 90% / 6,8 - 6,1
2	Hohe inhaltliche Qualität	89% - 75% / 6,0 - 5,1
3	Ausreichende inhaltliche Qualität	74% - 55% / 5,0 - 3,7
4	Mangelhafte inhaltliche Qualität	54% - 40% / 3,6 - 2,7
5	Ungenügende inhaltliche Qualität	39% - 0% / 2,6 - 0

Tabelle 3.2.1.2.22: Wissen; Prätest;
- cut-off-scores und Kategorien der Index I (Qualität)
Der Verfasser

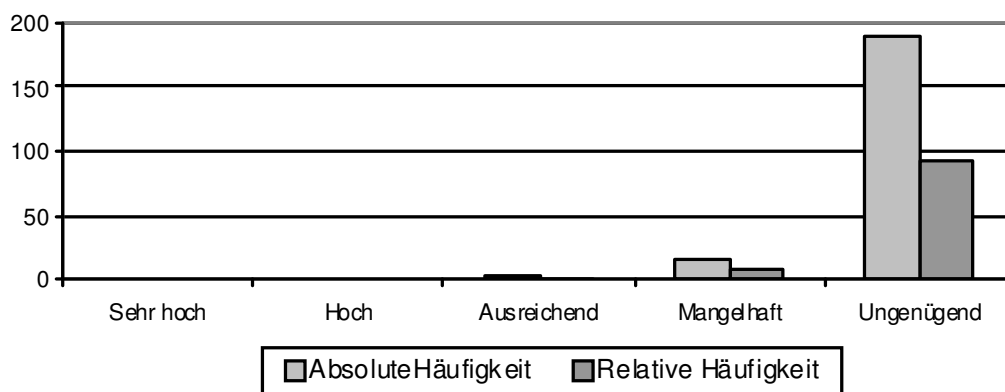


Diagramme 3.2.1.3.13: Wissen; Prätest;
- relative und absolute Häufigkeit in den unterschiedlichen Kategorien des Index I (Qualität)
Der Verfasser

Das Diagramm verdeutlicht die überwiegend geringe inhaltliche Qualität der Arbeiten der Schülerinnen und Schüler. Die Kategorien ‚sehr hoch‘ und ‚hoch‘ sind unbesetzt, der Kategorie ‚ausreichend‘ konnte lediglich eine Gruppenarbeit zugewiesen werden. Die übrigen 189 Arbeiten und damit 92% der Ergebnisse fallen in die Kategorien ‚ungenügend‘ und ‚mangelhaft‘, wobei die Kategorie ‚mangelhaft‘ deutlich stärker besetzt ist.

Analog dem Index I (Qualität) sollen auch die Ergebnisse des Index II (Quantität) in eine Grafik übertragen werden. Die relativen Werte der cut-off-scores aus der Tabelle 3.2.1.2.22 sollen hier übernommen werden, um eine hohe Vergleichbarkeit der Ergebnisse zu gewährleisten. Die Bezeichnungen der Kategorien, sowie die absoluten Werte sind der folgenden Tabelle zu entnehmen.

Nummer	Kategorien des Index II	cut-off-score (relativ/absolut)
1	Sehr hohe Vollständigkeit	100% - 90% / 7,8 - 7,00
2	Hohe Vollständigkeit	89% - 75% / 6,9 - 5,9
3	Ausreichende Vollständigkeit	74% - 55% / 5,8 - 3,9
4	Mangelhafte Vollständigkeit	54% - 40% / 3,8 - 3,1
5	Ungenügende Vollständigkeit	39% - 0% / 3,0 - 0

Tabelle 3.2.1.2.23: Wissen; Prätest;
- cut-off-scores und Kategorien der Index II
Der Verfasser

Die auf dieser Basis ermittelten Ergebnisse sind in der folgenden Grafik wiedergegeben:

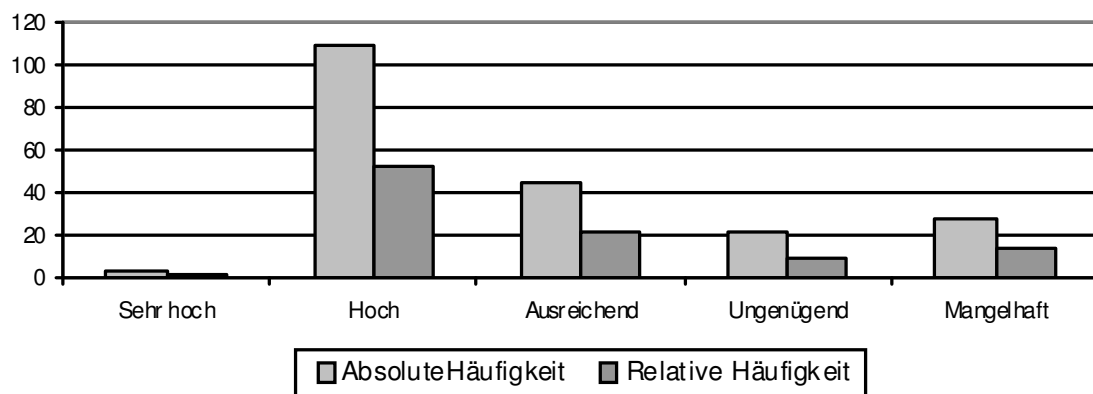


Diagramme 3.2.1.3.14: Wissen, Prätest;
- relative und absolute Häufigkeit in den unterschiedlichen Kategorien des Index II (Quantität)
Der Verfasser

Die These des Unterschiedes in Bezug auf ‚Quantität‘ und ‚Qualität‘ kann angenommen werden. Beide Verteilungen (Diagramm 3.2.3.13 und 3.2.1.14) stehen sich in Bezug auf die Kategorienhäufigkeiten diametral gegenüber.

An dieser Stelle wird die Leserin oder der Leser in Anbetracht der erzielten Ergebnisse fragen, ob die Aufgabe zu schwer gewählt wurde. Hiergegen sprechen vor allem zwei Aspekte: Zunächst die berechneten Schwierigkeitsindizes. Die Aufgabe ist, dies beweist der Abschnitt 2.3.7 in dieser Arbeit, schwer, aber nicht zu schwer gewählt. Weiterhin besteht in Bezug auf das Niveau der Aufgabe ein klarer Lehrplanbezug. Im Zuge der dritten und vierten Klasse sollen die Schülerinnen und Schüler lernen angemessen zu experimentieren (vgl. <http://www.schooloffice-sh.de/download/lehrplan/primar/hsu.pdf> 05.09.2006, 12.00 Uhr).

Das [Wissen], das heißt die Fähigkeit mit dem Versuch umzugehen, genau zu beobachten, diese Beobachtungen zu formulieren und korrekt zu verknüpfen, ist bei den Schülerinnen und Schülern der dritten und der vierten Klassenstufe weniger defizitär, sondern vielmehr überhaupt nicht vorhanden. Sie scheinen mit der Aufgabe, dies zeigt die ‚inhaltliche Qualität‘ der Lösungen, völlig überfordert.

Bislang lassen sich folgende Ergebnisse festhalten:

1. Das [Wissen] der Schülerinnen und Schüler lässt sich, dies konnte faktoranalytisch begründet werden, mit zwei Indizes, ‚Qualität‘ und ‚Quantität‘, beschreiben.
2. Die Schülerinnen und Schüler bemühten sich eine möglichst vollständige Lösung vorzulegen, deren ‚inhaltliche Qualität‘ aber völlig unzureichend ist.
3. Damit bleiben die Schülerinnen und Schüler weit hinter den Anforderungen des Lehrplanes zurück (vgl. Abschnitt 1.2.3 in dieser Arbeit).

Nachdem die groben Tendenzen der Daten vorgestellt wurden, fahre ich mit der differentialpsychologischen Auswertung fort: Ich beginne mit der Analyse des Einflusses der Klassenstufenzugehörigkeit auf die Ergebnisse. Die folgende Tabelle stellt die deskriptiven Werte in Abhängigkeit der Klassenstufe dar:

	Dritte Klasse	Vierte Klasse
Arithmetisches Mittel Index I	1,5	1,9
Standardabweichung Index I	0,4	0,6
Arithmetisches Mittel Index II	4,5	5,0
Standardabweichung Index II	1,4	0,9

Tabelle 3.2.1.2.24: Wissen; Prätest;
- deskriptive Werte in Abhängigkeit der Klassenstufe
Der Verfasser

Beide Indizes sind in der vierten Klassenstufe höher als in der dritten. Die genaue Überprüfung soll an dieser Stelle durch einen t-Test für unabhängige Stichproben erfolgen. Die Normalverteilungsvoraussetzung ist im Falle der vorliegenden Untersuchung nicht erforderlich, da mehr als 30 Messwertpaare untersucht werden (vgl. Bortz 1995, S. 142).

Für beide Indices muss in Bezug auf die Klassenstufe die Forschungshypothese angenommen, das heißt die Schülerinnen und Schüler der Klassenstufe vier schneiden signifikant besser ab, als die der Klassenstufe drei.

Ich fahre nun mit der Analyse der Geschlechtszugehörigkeit auf die Testergebnisse fort:

	Arithmetisches Mittel	Standardabweichung	t-Wert	Freiheitsgrade
Jungen Index I	1,4	0,3	-10,9	101
Mädchen Index I	2,0	0,5		
Jungen Index II	4,2	1,2	-6,4	101
Mädchen Index II	5,3	0,9		

Tabelle 3.2.1.2.25: Wissen; Prätest;
- Ergebnisse der Untersuchung des Einflusses des Geschlechts
Der Verfasser

Für beide Indices kann die Forschungshypothese angenommen werden. Mädchen sind in den durch die Variable [Wissen] gemessenen Leistungen signifikant besser als Jungen. Damit kann der bisherige Trend der Daten in Bezug auf die Kompetenzen der Mädchen bestätigt werden. Mädchen erzielten im Bereich der Selbsteinschätzung ([Einstellung] und [Motivation]) konsistent schlechtere Resultate als die Jungen, das heißt sie wiesen im Durchschnitt eine negativere [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] auf und waren geringer motiviert. Bei der [formalen Kompetenz] und beim [Wissen], das heißt bei den tatsächlichen Fähigkeiten und Fertigkeiten, erwiesen sich die Mädchen als signifikant kompetenter als die Jungen. In dieser Einteilung nimmt die Variable [allgemeines Experimentierverhalten] in Bezug auf die geschlechtsspezifischen Unterschiede eine Sonderstellung ein. Diese ist weiter oben in diesem Abschnitt interpretiert worden, so dass an dieser Stelle nicht erneut darauf eingegangen werden muss.

In Bezug auf die differenzialpsychologische Analyse lassen sich folgende Ergebnisse festhalten:

1. Schülerinnen und Schüler der vierten Klassenstufe lösen die Aufgabe signifikant besser als solche der dritten Klassenstufe.
2. Mädchen sind den Jungen bei der Bewältigung dieser Aufgabe überlegen. Damit bestätigt sich ein allgemeiner Trend der Gesamtdaten. Mädchen sind in Bezug auf die erhobenen Kompetenzvariablen Jungen konsistent überlegen.

3.2.1.3 Zusammenfassung der Ergebnisse

Ich beginne mit der Zusammenfassung der Ergebnisse. Dies geschieht zunächst anhand eines tabellarischen Überblicks. Im Weiteren werden dann die Kennwerte der einzelnen Variablen zusammenfassend inhaltlich interpretiert. Im zweiten Schritt überprüfe ich die Hypothese 1 aus dem Abschnitt 1.6, die ein geringes Niveau der Prätestergebnisse postuliert.

In der folgenden Tabelle habe ich die wichtigsten deskriptiven Werte der einzelnen Variablen aufgelistet:

Allgemeines Experimentierverhalten				
Verhaltenskategorien	I	II	III	IV
Häufigkeit	18	149	50	8
Geschlecht	Kein Einfluss			
Klassenzugehörigkeit	Einfluss in den Randbereichen der Verteilung			
Einstellung gegenüber Phänomen und ihrer Bearbeitung				
Arithmetisches Mittel	42,4 (Maximalwert: 60)			
Standardabweichung	9,3			
Geschlecht	Einstellung der Jungen ist positiver als die der Mädchen.			
Alter	Mit steigendem Alter wird die Einstellung negativer.			
Formale Kompetenz				
Arithmetisches Mittel	33,4 (Maximalwert: 68)			
Standardabweichung	9,2			
Geschlecht	Mädchen erzielen bei der Lösung der Aufgabe bessere Ergebnisse als Jungen.			
Klassenzugehörigkeit	Kinder der vierten Klasse lösen die Aufgabe besser.			
Aktuelle Motivation				
Arithmetisches Mittel	45,6 (Maximalwert: 66)			
Standardabweichung	9,0			
Geschlecht	Mädchen haben eine geringere Motivation als Jungen.			
Klassenzugehörigkeit	Zwischen der dritten und vierten Klassenstufe sinkt die Motivation.			
Wissen				
	Index I (Qualität)		Index II (Quantität)	
Arithmetisches Mittel	1,7 (Maximalwert: 6,8)		4,8 (Maximalwert: 7,8)	
Standardabweichung	0,6		1,2	
Geschlecht	Mädchen lösen die Aufgabe besser als Jungen.			
Klassenzugehörigkeit	Kinder der dritten Klasse schneiden schlechter ab als solche der vierten Klasse.			

Tabelle 3.2.1.3.1: Zusammenfassung; Prättest
- Zusammenfassung der Prätestergebnisse
Der Verfasser

Die Auswertung der Variablen [allgemeines Experimentierverhalten] erfolgt über einen Index, anhand dessen die Leistungen der Schülerinnen und Schüler den Verhaltenskategorien zugeordnet wurden. Die Daten zeigen ein bedenkliches Bild hinsichtlich des [allgemeinen Experimentierverhaltens] der untersuchten Kinder. Der größte Teil der Schülerinnen und Schüler (66%) erreichte lediglich die Verhaltenskategorie II (diffuses Experimentierverhalten). Hier kommt es lediglich zu einer Annäherung an das Experiment. Vertieftes und zielorientiertes Arbeiten ist den Schülerinnen und Schülern in dieser Verhaltenskategorie noch nicht möglich. Es konnte kein Einfluss des Geschlechtes festgestellt werden. Die Zugehörigkeit zur Klassenstufe hat allerdings Einfluss auf die Ausprägungen in den Randbereichen der Verteilung. Die Schülerinnen und Schüler der vierten Klassenstufe erreichten signifikant häufiger die Verhaltenskategorie IV (elaboriertes Experimentierverhalten) und signifikant seltener die Stufe I.

Die Betrachtung des arithmetischen Mittels der [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] erzeugt den Eindruck einer unproblematischen, das heißt im Durchschnitt ‚positiven Einstellung‘. Betrachtet man die Verteilung jedoch genauer anhand der kategorisierten, erreichten Punktsummen so fällt auf, dass eine deutliche Zweiteilung vorherrscht. Etwa die Hälfte der Schülerinnen und Schüler weist eine ‚begeisterte, sehr positive oder positive Einstellung‘ auf, die andere eine ‚ambivalente bis negative‘. Es besteht ein deutlicher Einfluss des Alters und des Geschlechts: So haben Mädchen eine negativere Einstellung als Jungen und mit steigendem Alter wird die Einstellung sukzessive negativer.

Die [formale Kompetenz] ist bei den Schülerinnen und Schülern defizitär ausgeprägt. Im Durchschnitt erreichten die Kinder lediglich 49% der verfügbaren Rohpunkte. Kategorisiert man die individuell erreichten Punktsummen, so haben die meisten Schülerinnen und Schüler lediglich eine ‚durchschnittliche bis unbefriedigende formale Kompetenz‘. Erneut spielen das Geschlecht und die Klassenzugehörigkeit eine wichtige Rolle. So verbessert sich mit steigender Klassenstufe die [formale Kompetenz] und die Mädchen sind den Jungen deutlich überlegen.

Die [aktuelle Motivation] der Schülerinnen und Schüler liegt im Durchschnitt im positiven Bereich der Skala. So sind die Kategorien ‚hoch motiviert‘ und ‚durchschnittlich motiviert‘ die am stärksten besetzten. Die Kategorie ‚gering motiviert‘ ist mit 7% die am schwächsten besetzte. Problematisch hingegen ist, dass die [Motivation] genauso wie die [Einstellung] mit steigendem Alter abnimmt, und dass Mädchen eine geringere [Motivation] haben als die Jungen.

In Bezug auf das [Wissen] offenbaren sich die bedenklichsten Ergebnisse aus den hier erhobenen Daten. Auch wenn sich die Schülerinnen und Schüler um eine vollständige Lösung der Aufgabe bemühten (vgl. Index II), so ist die ‚inhaltliche Qualität‘ zu 92% im mangelhaften Bereich. Die Schülerinnen und Schüler waren nicht in der Lage, mit dem Versuch adäquat umzugehen, das heißt in diesem Zusammenhang vor allem genau zu beobachten, die Beobachtungen zu verknüpfen und angemessen in die Syntax des Maps zu übertragen. Genau wie bei der [formalen Kompetenz] waren die Mädchen bei der Lösung dieser Aufgabe signifikant besser als die Jungen. Außerdem war der Einfluss der Klassenstufenzugehörigkeit insofern spürbar, als dass Kinder in der vierten Klassenstufe die Aufgabe besser lösten, als solche der dritten Klassenstufe.

Die Forderungen des Lehrplans nach einem angemessenen Umgang mit Experimenten und der Erklärung einfacher naturwissenschaftlicher Gesetzmäßigkeiten, wirken im Hinblick auf die erhobenen Daten illusionär. Sie werden in der breiten Masse der Schülerschaft nicht erfüllt.

3.2.2 Der Posttest I

3.2.2.1 Die Wirksamkeit von ASIP drei Monate nach der Teilnahme

Ähnlich der Vorgehensweise im Abschnitt 3.2.1.2 werden hier nun die Ergebnisse des ersten Posttests drei Monate nach der Teilnahme an ASIP ausführlich dargestellt, interpretiert und diskutiert. Eine Zusammenfassung der Ergebnisse, sowie die Auswertung der Daten im Hinblick auf die Hypothesen wird im folgenden Kapitel unternommen. Zur besseren Übersicht orientiere ich mich in der Reihenfolge der Darstellung der Variablen an dem bereits genannten Abschnitt 3.2.1.2, wobei im Prätest natürlich die Erinnerungen an die Stationen mit der ich hier beginne noch nicht erfasst werden konnten.

Die Variable deskriptive Aspekte

Die Erinnerung der Schülerinnen und Schüler an die Stationen wird, dies wurde bereits in Abschnitt 2.5 dieser Arbeit erwähnt, nicht als ein Ausdruck von Wirksamkeit gefasst, sondern ist vielmehr eine Bedingung ihrer Möglichkeit. Insofern wird dieser Variablen keine Wirksamkeitsstufe zugewiesen, sondern es ist vielmehr im Verlauf dieses Abschnitts zu klären, ob und in welchem Umfang sich die Kinder an die Versuche des Programms erinnern. Nur wenn dies der Fall ist, können die übrigen Variablen bezüglich der Wirksamkeit der MINIPHÄNOMENTA sinnvoll untersucht werden.

Ich beginne mit der Analyse des ersten Teils des Fragebogens, der Erinnerung der Schülerinnen und Schüler an die Stationen. Forschungsmethodisch wurde dies durch die Darbietung von Bildern von zufällig ausgewählten Stationen erhoben. Die Schülerinnen und Schüler sollten anhand eines Bildes von einer Station angeben, ob sie sich an die jeweilige Station erinnern oder nicht. Insgesamt gingen 188 Fragebögen in die folgende Auswertung ein. Daraus mussten 29 Fragebögen, das heißt 13% wegen einer zu ausgeprägten Ja-sage-Tendenz ausgesondert werden. Dies geschah, wenn eine Erinnerung an zwei an der Schule nicht angebotenen Stationen angegeben wurde.

Die folgende Tabelle fasst einige zentrale deskriptive Werte des ersten Teils des Fragebogens zusammen:

Bezeichnung	Wert
Arithmetisches Mittel	10,8
Konfidenzintervall (95%)	10,4 – 11,1
Modalwert	13
Standardabweichung	2,4
range	10

Tabelle 3.2.2.1.1: deskriptive Aspekte; Posttest I;
- deskriptive Werte der Erinnerungsbilder
Der Verfasser

Die Werte zeigen deutlich, wie umfassend sich die Schülerinnen und Schüler an die Stationen erinnern. Im Durchschnitt konnten Sie sich bei einem arithmetischen Mittel von 10,8 an 83% der Stationen erinnern. 36% und damit mehr als ein Drittel der Schülerinnen und Schüler, konnten sich an alle Stationen erinnern. Positiv zu bewerten ist auch die geringe Standardabweichung. Damit weichen die Werte im Durchschnitt nur im geringen Umfang vom hoch liegenden arithmetischen Mittel ab.

Zur vertiefenden Interpretation der Daten ist die Verteilung der Variablen in dem folgenden Diagramm abgebildet:

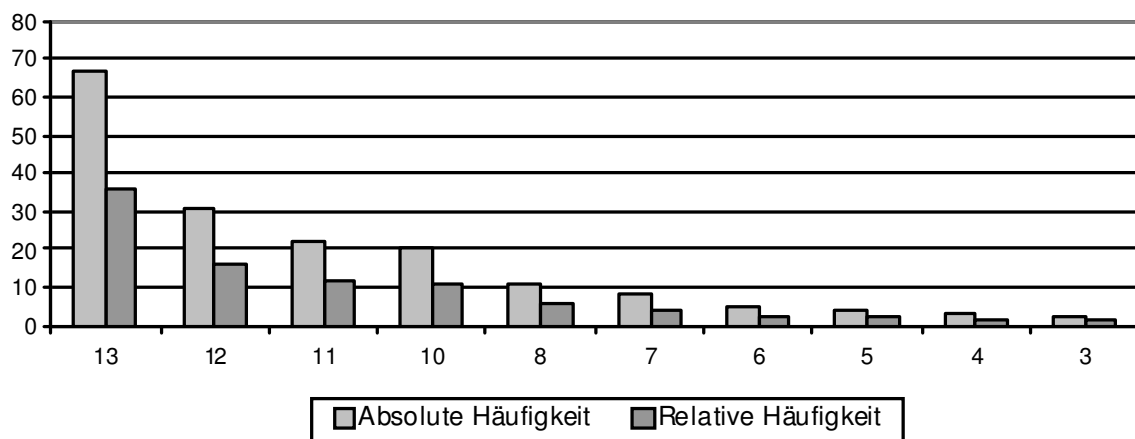


Diagramm 3.2.2.1.1: Deskriptive Aspekte; Posttest I;
- Erinnerungshäufigkeit
Der Verfasser

Das Diagramm bestätigt den geschilderten positiven Eindruck: Ein Großteil der Schülerinnen und Schüler kann sich an eine Vielzahl von Stationen erinnern. Interessant ist dabei weiterhin der Umstand des stetigen Abnehmens der absoluten und relativen Häufigkeit zusammen mit dem Absinken der Anzahl der erinnerten Bilder.

Die Schülerinnen und Schüler erinnern sich im Durchschnitt sehr umfassend an die Stationen.

Zum ersten Teil des Fragebogens soll an dieser Stelle abschließend untersucht werden, ob die Klassenzugehörigkeit der Schülerinnen und Schüler oder deren Geschlecht Einfluss auf die Anzahl der erinnerten Stationen haben. Dies geschieht mit einem t-Test für unabhängige Stichproben. Die Normalverteilungsvoraussetzung muss an dieser Stelle nicht erfüllt sein, da die Anzahl der analysierten Paare größer gleich 30 ist. Die Ergebnisse der beiden Tests werden in der folgenden Tabelle wiedergegeben. Die Alterswerte der Randbereiche wurden, wie bereits in Abschnitt 3.2.1.2 vorgenommen, zusammengefasst, da die Kategorien sieben Jahre und elf Jahre zu selten besetzt sind, um zu validen Ergebnissen zu gelangen.

	Geschlecht		Alter		
	Jungen	Mädchen	7 -8	9	10 – 11
Arithmetisches Mittel	10,7	10,8	11,1	10,6	10,8

Tabelle 3.2.2.1.2: deskriptive Aspekte; Posttest I;
- Erinnerung und differentialpsychologische Merkmale
Der Verfasser

Die Differenzen zwischen der Anzahl der erinnerten Stationen sind derartig gering, dass eine Signifikanzüberprüfung hier entfallen kann.

Das Alter und das Geschlecht der Schülerinnen und Schüler üben keinen Einfluss auf den Umfang der Erinnerung an die Stationen aus.

Ich fahre nun mit der Auswertung des zweiten Teils des Fragebogens fort: Die Stichworte, sowie die Zeichnungen der Schülerinnen und Schüler wurden durch die Anwendung einer quantitativen Inhaltsanalyse mit insgesamt acht Kategorien ausgewertet. Die folgende Tabelle gibt einen Überblick über die Ausprägung in den unterschiedlichen Kategorien:

Bezeichnung der Kategorie	Arithmetisches Mittel	Standardabweichung
Anzahl der Stichpunkte	7,1	2,6
Anzahl der richtigen Stichpunkte	6,3	2,6
Aufbau des Versuches	3,4	0,8
Funktion des Versuches	3,3	0,9
Effekt des Versuches	3,3	0,9
Erklärungsansätze für den Versuch	1,5	0,9
Affektive Äußerungen zum Versuch	3,3	0,7
Eidetischer Charakter der Äußerungen	3,3	0,7

Tabelle 3.2.2.1.3: deskriptive Aspekte; Posttest I;
- deskriptive Werte des zweiten Teils des Fragebogens
Der Verfasser

Die ‚Anzahl der durchschnittlich genannten Stichpunkte‘ liegt mit 7,1 Nennungen im oberen Bereich. Besonders erfreulich ist der Umstand zu bewerten, dass die ‚Anzahl der Stichpunkte‘ und die ‚Anzahl der richtigen Stichpunkte‘ sich nur wenig unterscheiden: Die Schülerinnen und Schüler nannten nur wenige falsche oder irrelevante Aspekte.

Ebenfalls positiv zu bewerten ist der Umstand, dass die Beschreibung der Versuche durch die Kinder größtenteils deutlich ‚affektiv‘ gefärbt sind. Die Erinnerung an die Stationen hat ‚eidetischen Charakter‘, das heißt die Schülerinnen und Schüler speichern diese in Form von Vorstellungsbildern, die zumeist mit einer konkreten Handlung verbunden sind. Dies wird vor allem dadurch deutlich, dass ein Großteil der Schülerinnen und Schüler die Stationen nicht alleine im Sinne einer technischen Zeichnung malte, sondern vielmehr mit einem oder mehreren Kindern, die an der Station tätig sind. Die Beschreibung der Station gelingt den Kindern, dies zeigen die Werte in den Kategorien ‚Aufbau‘, ‚Funktion‘ und ‚Effekt‘, gut und umfassend. Interessant ist der Umstand zu bewerten, dass die Schülerinnen und Schüler in der Beschreibung der Versuche kaum Erklärungen lieferten. Es ist zu fragen, wie diese Tatsache inhaltlich zu interpretieren ist. Im ersten Zugang könnte man geneigt sein, den geringen Wert auf der Variablen ‚Erklärung‘ negativ zu bewerten. Zum Beispiel in der Form, dass die Stationen zu „Spielobjekten“ umfunktioniert werden und deren naturwissenschaftlich-physikalischer Ansatz in den Hintergrund gedrängt wird. Diese Form der Erklärung scheint mir nicht angemessen. Vielmehr bewerte ich diesen Umstand äußerst positiv. Er ist ein Indiz dafür, dass die Schülerinnen und Schüler einen positiven handlungs- und emotionsbezogenen Zugang zu Naturwissenschaften bekommen. Diese Interpretation wird durch den ‚eidetischen Charakter der Erinnerungen‘ und durch die ‚affektiven Äußerungen‘ der Schülerinnen und Schüler gestützt.

Die folgende Tabelle gibt einige zentrale Werte hinsichtlich der Punktsumme der Inhaltsanalyse wieder. In dieser blieben die ‚Anzahl der Stichpunkte‘ sowie die ‚Anzahl der richtigen Stichpunkte‘ unberücksichtigt, da sie keine unmittelbar relevanten Aussagen zur Beschreibung der Stationen liefern:

Bezeichnung	Wert
Arithmetisches Mittel	18,4 (Maximalwert: 24)
Konfidenzintervall (95%)	17,8 – 18,9
Modalwert	21
Standardabweichung	3,6
range	18,0

Tabelle 3.2.2.1.4: deskriptive Aspekte; Posttest I;
- zentrale Werte der Inhaltsanalyse
Der Verfasser

Das arithmetische Mittel liegt in Bezug auf die erreichbare Höchstpunktzahl extrem hoch. Die Schülerinnen und Schüler erreichten im Durchschnitt 77% der verfügbaren Rohpunkte. Auch der Modalwert liegt mit 21 Punkten sehr hoch. Ihn erreichten allein 20% der an der Untersuchung Teilnehmenden. Die Standardabweichung ist mit 3,6 eher gering.

Bei der Verteilung der individuell erreichten, kumulierten Werte der Inhaltsanalyse handelt es sich aufgrund einer deutlichen Rechtssteilheit der Verteilung von -0,8, dies ergab der Kolmogorov-Smirnov-Test, nicht um eine Normalverteilung.

Für eine vertiefende inhaltliche Interpretation der Daten kategorisiere ich die in der Inhaltsanalyse individuell erreichten Punktwerte. Aufgrund der deutlichen Rechtsschiefe in der Verteilung, sowie der Anknüpfung an das Setting Schule, orientiere ich mich für die Definition der Erinnerungskategorien lose am oberen Bereich (sehr gut bis befriedigend) des Schulnotensystems. Die folgende Tabelle definiert Kategorien und cut-off-scores:

Nummer	Kategorien der Qualität der Erinnerung	cut-off-score	
		relativ	absolut
1	sehr gute Qualität der Erinnerung	100% - 90%	24 - 21
2	gute Qualität der Erinnerung	89% - 75%	20 - 18
3	befriedigende Qualität der Erinnerung	74% - 60%	17 - 14
4	unbefriedigende Qualität der Erinnerung	59% - 0%	13 - 0

Tabelle 3.2.2.1.5: deskriptive Aspekte; Posttest I;
- Kategorien und cut-off-scores
Der Verfasser

Die folgende Grafik stellt die relativen und absoluten Häufigkeiten der Kategorien dar:

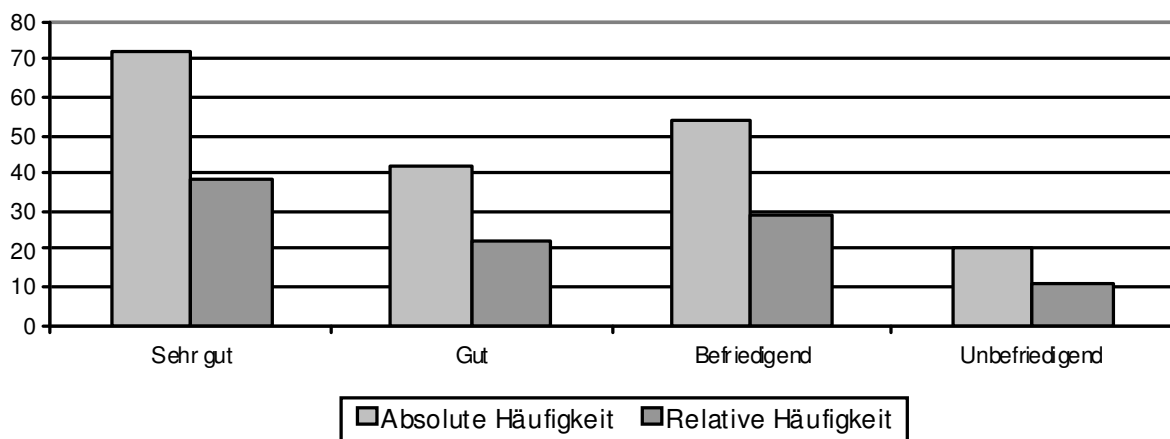


Diagramm 3.2.2.1.2: Deskriptive Aspekte; Posttest I;
- Häufigkeit der Kategorien der Qualität der Erinnerung
Der Verfasser

Die Kategorie ‚sehr gute Erinnerung an die MINIPHÄNOMENTA‘ trifft auf die Leistungen von mehr als einem Drittel der Schülerinnen und Schüler zu. Sie ist mit 38% die am stärksten besetzte Kategorie der Verteilung. 89% der Schülerinnen und Schüler gelingt eine ‚sehr gute bis befriedigende Erinnerung‘ an die Stationen. Lediglich 11% der an der Untersuchung Teilnehmenden erinnern sich in einem ‚ungenügenden Umfang‘ an die von ihnen ausgewählten Stationen der MINIPHÄNOMENTA.

Den Schülerinnen und Schülern gelingt es im Durchschnitt gut, eine von ihnen ausgewählte Station der MINIPHÄNOMENTA umfassend und korrekt zu beschreiben. Sie können den ‚Aufbau‘, die ‚Funktion‘ und den ‚Effekt‘ korrekt formulieren. Die Erinnerung hat ‚eidetischen Charakter‘ und ist deutlich emotional gefärbt.

Abschließend sollen zwei differentialpsychologische Fragestellungen bearbeitet werden. Sie befassen sich mit dem Einfluss dem Zusammenhang des Alters und Geschlechtes auf die Beschreibung der ausgewählten Stationen.

Die folgende Tabelle stellt deskriptive Werte und Kenndaten deren interferenzstatistischer Überprüfung zusammen:

	Geschlecht		Alter		
	Jungen	Mädchen	7 -8	9	10 - 11
Arithmetisches Mittel	17,8	19,0	16,7	19,7	18,2
			7 - 8 / 9	7 - 8 / 10 - 11	9 / 10 - 11
t-Wert	-2,5		-4,1	-0,9	2,3
Freiheitsgrade	91		43	43	59

Tabelle 3.2.2.1.6: deskriptive Aspekte; Posttest I;
- deskriptive Werte zu den differentialpsychologischen Fragestellungen
Der Verfasser

Es existiert ein signifikanter geschlechtsspezifischer Unterschied in Bezug auf die Qualität der Erinnerungen. Dieser ist vor allem darauf zurück zu führen, dass die Mädchen zwar ähnlich viele Stichpunkte wie die Jungen nennen, jedoch weniger Fehler machen, beziehungsweise irrelevante Aspekte nennen. Dieser Aspekt der höheren Kompetenz der Mädchen spielt in der Untersuchung eine zentrale Rolle, da er in Bezug auf eine Vielzahl von Variablen festgestellt werden konnte.

Auch der Zusammenhang des Alters mit dem Ergebnis der Inhaltsanalyse ist statistisch bedeutsam ausgeprägt. So steigt der durchschnittlich erreichte Punktwert mit steigendem Alter an, das heißt die Qualität der geschilderten Erinnerungen nimmt zu. Vergleicht man dieses Ergebnis mit dem Einfluss des Alters mit den Ergebnissen im ersten Fragebogenteil, so wird deutlich, dass die Anzahl der Erinnerungen mit steigendem Alter hier abnimmt.

Betrachtet man nun die einzelnen Variablen der Inhaltsanalyse, so wird deutlich, dass signifikante Unterschiede auch hier vor allem zwischen den sieben- bis acht- und neunjährigen Schülerinnen und Schülern bestehen. Die älteren Kinder schildern insbesondere ‚Aufbau‘ und ‚Funktion‘ signifikant besser, ihre Aussagen sind ‚affektiv‘ stärker geladen und besitzen einen stärkern ‚eidetischen Charakter‘.

In Bezug auf die Anzahl der verwendeten Stichpunkte sowie der richtigen Stichpunkte ergibt sich ein ähnliches Bild. Insgesamt nimmt die Anzahl der verwendeten und der richtigen Stichpunkte mit steigendem Alter zu. Die deutlichste Zunahme ist zwischen den Altersgruppen sieben bis acht Jahre und neun Jahre zu verzeichnen.

Basierend auf den angeführten Ergebnissen kann somit von einer umfassenden Erinnerung der Schülerinnen und Schüler an die MINIPHÄNOMENTA ausgegangen werden.

Die Variable allgemeines Experimentierverhalten

Die folgenden Ausführungen bezüglich des [allgemeinen Experimentierverhaltens] der Schülerinnen und Schüler beruhen auf der Auswertung von 63 Beobachtungen. Dieser im Vergleich zum Prätest verhältnismäßig geringe Stichprobenumfang (vgl. Abschnitt 3.2.1.2 in dieser Arbeit) ist auf zwei Aspekte zurück zu führen:

1. Im Prätest wurden die Beobachtungsbögen der Experimentalgruppe I und der Kontrollgruppe ausgewertet. Hier hingegen werden untersuchungsbedingt lediglich die Ergebnisse der Experimentalgruppe verarbeitet.
2. Auch wenn man den unter Punkt eins genannten Aspekt berücksichtigt, ist die Zahl der auszuwertenden Beobachtungen immer noch zu gering: Der sich ergebende Differenzbetrag musste aufgrund von Beobachtungs- und Kodierfehlern ausgesondert werden. Dies betraf insgesamt 37 Beobachtungsbögen, die zum einen fehlerhaft oder unvollständig ausgefüllt waren, zum anderen offensichtliche Fehlertendenzen (Tendenz zur Mitte, Milde-Härte-Fehler usw.) aufwiesen.

Insgesamt erscheint die Datenbasis aber dennoch breit genug, um zu validen Ergebnissen gelangen zu können. Die Auswertung der Erhebungsbögen erfolgt erneut über einen Index, dessen Ermittlung und Interpretation bereits in Abschnitt 3.2.1.2 in dieser Arbeit dargelegt wurde und daher an dieser Stelle nicht wiederholt werden muss.

Das folgende Diagramm zeigt die Verteilung der durch die Schülerinnen und Schüler erreichten Verhaltenskategorien⁴⁹ für den Prätest und den Posttest I. Aufgrund des deutlich unterschiedlichen Stichprobenumfanges erscheint eine Darstellung in relativen Werten angebracht:

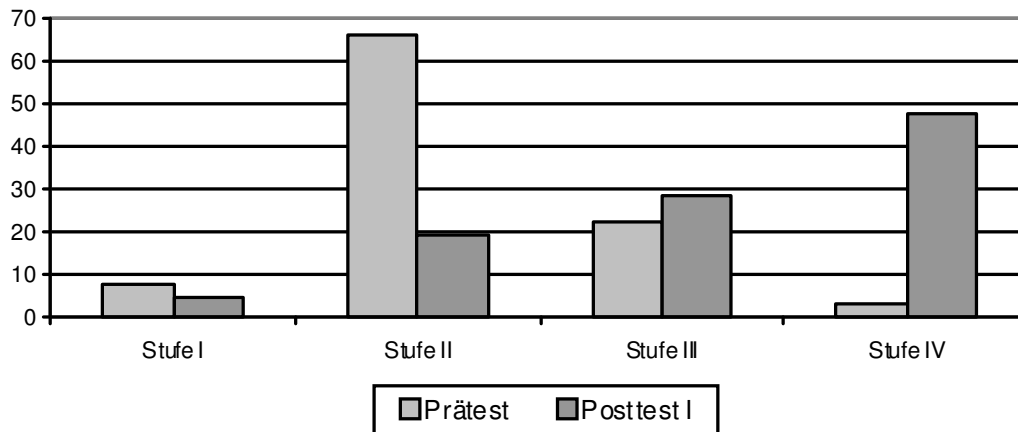


Diagramm 3.2.2.1.3: Allgemeines Experimentierverhalten; Posttest I;
- erreichte Verhaltenskategorien in Prä- und Posttest I
Der Verfasser

Die Grafik zeigt eindrucksvoll den überaus positiven Einfluss der MINIPHÄNOMENTA. Bevor mit der inhaltlichen Interpretation dieser Ergebnisse begonnen werden kann, soll nachgewiesen werden, dass beide Verteilungen sich statistisch bedeutsam unterscheiden, dass die Unterschiede nicht durch Zufallsschwankungen bedingt sind. Dies geschieht durch einen k*1-Chi-Quadrat-Test, dessen Ergebnisse in der folgenden Tabelle festgehalten sind:

Größe	ermittelter Wert
df	3
χ^2	93,6

Tabelle 3.2.2.1.7: Allgemeines Experimentierverhalten; Posttest I
- Überprüfung der Prätest- Posttest-Unterschiede
Der Verfasser

Der Unterschied zwischen beiden Verteilungen ist hoch bedeutsam. Auf dieser Basis wurde die Effektgröße ermittelt. Sie ist mit einem Wert von $\epsilon = 0,9$ (vgl. Bortz 1995, S. 150 ff.) als hoch zu klassifizieren.

⁴⁹ Die Definition der Verhaltenskategorien beziehungsweise Stufen wurde im Abschnitt 2.3.5 vorgenommen und soll an dieser Stelle nicht wiederholt werden.

Der größte Teil der Schülerinnen und Schüler (66%) erreichten vor der Teilnahme an der MINIPHÄNOMENTA lediglich die Verhaltenskategorie II („diffuses Experimentierverhalten“). Damit waren sie emotional wenig beteiligt, nicht wirklich in den Versuch vertieft, testeten keine Variablen, und in Bezug auf das soziale Lernen waren lediglich Ansätze festzustellen. Nach der Teilnahme an der MINIPHÄNOMENTA hat sich der Umgang der Schülerinnen und Schüler mit Experimenten grundsätzlich gewandelt. Die Verhaltenskategorie IV („elaboriertes Experimentierverhalten“) ist nun die am stärksten besetzte Kategorie. Ihr sind insgesamt 48% der Schülerinnen und Schüler zuzurechnen. Damit zeigt fast die Hälfte der untersuchten Kinder kompetentes experimentelles Verhalten, ist in die Versuche vertieft und emotional am Experimentiergeschehen beteiligt. Ferner arbeiten die Kinder intensiv und zielgerichtet in Kleingruppen an den Versuchen, sie tauschen sich aus, diskutieren den Versuchsverlauf, die Ergebnisse, sowie Ansätze zur Erklärung.

Weiterhin erfreulich ist der Rückgang der ‚ablehnenden Reaktionen auf die Versuch‘, sowie das Ansteigen der relativen Häufigkeit in der Verhaltenskategorie III („gerichtetes Experimentierverhalten“).

Insgesamt kann festgehalten werden, dass die Schülerinnen und Schüler nach der Teilnahme an ASIP deutlich kompetenter waren. In fast der Hälfte der Fälle erreichten sie nun die Verhaltenskategorie IV („elaboriertes Experimentierverhalten“). Die Wahrscheinlichkeit für das Auftreten der Stufen I („negative Reaktionen“) und II („diffuses Experimentierverhalten“), die im Prätest am häufigsten besetzten Kategorien, sank deutlich, während die Verhaltenskategorie III („gerichtetes Experimentierverhalten“) häufiger vergeben werden konnte. Dieser positive Effekt wird durch die hohe Effektgröße von $\epsilon = 0,9$ belegt.

Zur Vertiefung des Analysestandes soll der Einfluss des Geschlechtes, der Klassenzugehörigkeit und des Arbeitens in der Gruppe auf das Erreichen einer bestimmten Verhaltensstufe ermittelt werden. Die folgende Kreuztabelle gibt diesen Zusammenhang wieder:

	Stufe I	Stufe II	Stufe III	Stufe IV	Gesamt
Männlich	0	5	11	15	31
Weiblich	3	7	7	15	32
Dritte Klasse	0	6	12	15	33
Vierte Klasse	3	6	7	14	30
Alleine	2	1	7	3	13
In der Gruppe	1	6	14	29	50

Tabelle 3.2.2.1.8: Allgemeines Experimentierverhalten; Posttest I;
- Kreuztabelle der zu untersuchenden Merkmale
Der Verfasser

Ich beginne mit dem Einfluss des Geschlechtes auf das Erreichen einer Verhaltenskategorie: Wie bereits im Prätest festgestellt, kann auch im Posttest I kein Einfluss festgestellt werden.

Im Prätest konnte der Einfluss der Klassenzugehörigkeit auf den Randbereich der Verteilung (Verhaltenskategorie I und IV) nachgewiesen werden: Die Schülerinnen und Schüler der vierten Klassenstufe erreichten statistisch bedeutsam häufiger die Verhaltenskategorie IV als die der Klassenstufe drei. Weiterhin mussten die Kinder der dritten Klassenstufe statistisch bedeutsam häufiger der Stufe III zugewiesen werden. Dieser Effekt ist im Posttest I nicht mehr nachweisbar. Der Unterschied in Bezug auf die Verhaltenskategorie IV ist derartig gering, dass eine Signifikanzüberprüfung unterbleiben kann. In Bezug auf den vorhandenen Unterschied in der Stufe I erscheint eine Interpretation unangebracht, da insgesamt nur sehr wenige Untersuchungsteilnehmerinnen und Untersuchungsteilnehmer dieser Kategorie zugewiesen wurden: Zum einen ist dadurch eine Signifikanzüberprüfung über Standardverfahren nicht mehr möglich (vgl. Bortz 1995, S.159 f.), zum anderen würde der inhaltlichen Interpretation keine größere Validität zukommen. Insofern kann an dieser Stelle über diese Frage nicht endgültig entschieden werden. Sie bleibt weiteren Forschungsvorhaben überantwortet. Als Ergebnis kann an dieser Stelle daher festgehalten werden, dass der Einfluss der Klassenzugehörigkeit auf den oberen Bereich der Verteilung (Erreichen der Verhaltenskategorie IV) nivelliert wurde. Die Klassenstufe kann folglich nicht mehr als Indikator gesehen werden, der die Wahrscheinlichkeit des Auftretens einer hohen Lerntiefestufe moderiert.

In Bezug auf die Sozialform des Arbeitens traten Gruppen- und Einzelarbeit im Prätest ähnlich häufig auf. Dies ist im Posttest keineswegs mehr der Fall: Zu 79% arbeiten die Schülerinnen und Schüler in Gruppen. Nach der Teilnahme der Schülerinnen und Schüler an der MINIPHÄNOMENTA gibt es folglich eine deutlich ausgeprägtere Tendenz, in Kleingruppen zu arbeiten. Diese zeigt sich durchgängig in allen Verhaltenskategorien, so dass auch hier die Datentendenz des Prätests nicht mehr besteht.

Nach der genannten Definition ist die Variable [allgemeines Experimentierverhalten] im Posttest I

HOCHWIRKSAM (Ebene I)

- Die Schülerinnen und Schüler waren nach der Teilnahme an ASIP deutlich kompetenter. In fast der Hälfte der Fälle erreichten sie nun die Verhaltenskategorie IV (,elaboriertes Experimentierverhalten'). Die Wahrscheinlichkeit für das Auftreten der Stufen I (,negative Reaktionen') und II (,diffuses Experimentierverhalten'), die im Prätest am häufigsten besetzte Kategorie, sank deutlich, während die Verhaltenskategorie III häufiger vergeben werden konnte. Die beschriebene Wirksamkeit muss mit einer Effektgröße von $\epsilon = 0,9$ als sehr hoch klassifiziert werden.
- Die Klassenstufe ist kein Indikator mehr für das Auftreten der Verhaltenskategorie IV (,elaboriertes Experimentierverhalten'). Da weder im Prätest noch im Posttest I ein Einfluss des Geschlechts auf das [allgemeine Experimentierverhalten] nachweisbar war, beträgt der Umfang der Wirkung insgesamt 100%.

Die Variable Einstellung gegenüber Phänomenen und ihrer Bearbeitung

Im Posttest I konnten in Bezug auf die [Einstellung] 217 Fragebögen ausgewertet werden. Für einen ersten Vergleich der Ergebnisse aus Prä- und Posttest werden einige wichtige deskriptive Verteilungsmerkmale der individuell erreichten Rohpunktsummen in der folgenden Tabelle einander gegenübergestellt:

	Prätest	Posttest I
Arithmetisches Mittel	42,4	49,2
Konfidenzintervall (95%)	41,5 – 43,3	48,1 – 50,2
Modalwert	37	56
Standardabweichung	9,3	8,1
range	44	39

Tabelle 3.2.2.1.9: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Posttest I;
- deskriptive Kennwerte
Der Verfasser

Bevor eine inhaltliche Interpretation der Daten unternommen wird, soll zunächst die Signifikanz des Mittelwertsunterschiedes durch einen t-Test für unabhängige Stichproben überprüft werden. Die folgende Tabelle fasst die Ergebnisse des Prüfverfahrens zusammen:

Bezeichnung	Wert
df	216
t – Wert	2,1

Tabelle 3.2.2.1.10: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Posttest I;
-Ergebnisse des t-Test
Der Verfasser

Das arithmetische Mittel der Posttestverteilung ist, im Vergleich zu dem der Prätestverteilung, um 6,2 Rohpunkte gestiegen. Dieser Unterschied ist signifikant. Auf dieser Basis wurde eine hohe Effektgröße, das heißt ein geringer Überlappungsbereich der Prä- und der Posttestverteilung (vgl. Bortz, Döring 2006, S. 607 f.) von 68% ermittelt.

War die [Einstellung der Schülerinnen und Schüler gegenüber Phänomenen und ihrer Bearbeitung] im Vorfeld der MINIPHÄNOMENTA ‚durchschnittlich bis leicht positiv‘, so ist sie nun ‚ausgeprägt positiv‘. Dieses Bild des Einflusses von ASIP wird auch durch die Streuungsmaße unterstrichen. Insgesamt ist die Dispersion der Werte deutlich zurückgegangen. Die Daten haben sich folglich auf dem beschriebenen hohen Niveau stabilisiert.

Die Posttestverteilung der individuell erreichten kumulierten Punktschichten weicht, dies zeigt der Kolmogorov-Smirnov Test mit einem Z-Wert von 2,0, signifikant von der Normalverteilung ab. Es handelt sich mit einer Schiefe von -0,97 um eine ausgeprägt rechtsschiefe Verteilung.

Zur vertiefenden inhaltlichen Interpretation sollen nun die individuell erreichten kumulierten Punktschichten Kategorien zugewiesen werden. Um einen direkten Vergleich von Prä- und Posttestverteilung zu ermöglichen, verwende ich die bereits eingeführten cut-off-scores, die dem Abschnitt 3.2.1.2 in dieser Arbeit zu entnehmen sind. Aufgrund des deutlich unterschiedlichen Stichprobenumfanges in Prä- und Posttest verwende ich zur folgenden Darstellung relative Werte:

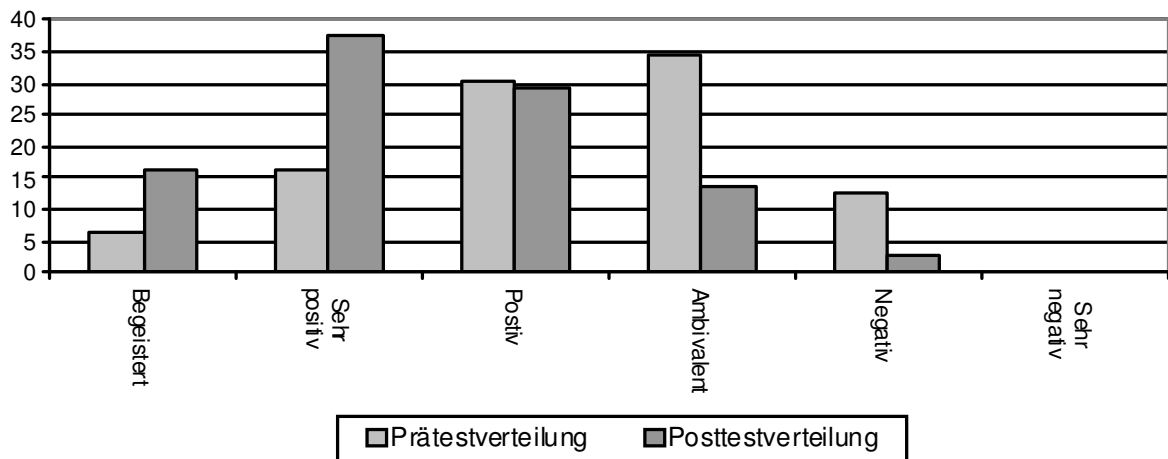


Diagramm 3.2.2.1.4: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Posttest I;
 - Häufigkeit der Einstellungskategorien in Prä- und Posttest
 Der Verfasser

Beide Verteilungen zeigen deutlich unterschiedliche Kategorienhäufigkeiten. Zur genaueren Auswertung sollen die Unterschiede in einer Kreuztabelle aufgelistet, sowie durch einen k*I-Chi-Quadrat-Test auf Signifikanz überprüft werden:

	Begeistert	Sehr positiv	Positiv	Ambivalent	Negativ	Sehr negativ	Gesamt
Prätest	26	71	132	151	56	0	436
Posttest	36	81	64	30	6	0	217
Gesamt	62	152	196	181	62	0	653

Tabelle 3.2.2.1.11: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Posttest I;
 - Kreuztabelle der Kategorienhäufigkeiten
 Der Verfasser

Der k*I-Chi-Quadrat-Test ergab mit einem χ^2 - Wert von 82,9 (df = 5) einen hoch signifikanten Unterschied zwischen beiden Verteilungen. War im Prätest die Kategorie ‚ambivalente Einstellung‘ die am häufigsten besetzte, so hat sich dies im Zuge der Teilnahme an der MINIPHÄNOMENTA geändert. Die deutliche Mehrheit der Schülerinnen und Schüler (83%) konnte nun den obersten drei Kategorien zugeordnet werden, wobei die Kategorie ‚sehr positive Einstellung‘ der Modalwert der kategorisierten Werte ist.

Die Teilnahme an der MINIPHÄNOMENTA hatte einen deutlichen Einfluss auf die [Einstellung der Schülerinnen und Schüler gegenüber Phänomenen und ihrer Bearbeitung]:

- Das arithmetische Mittel der individuell erreichten kumulierten Testwerte stieg um 6,2 Punkte. Dies entspricht einem hohen Effekt.
- Die Streuung der Werte ging zurück, stabilisierte sich auf hohem Niveau.
- Die deutliche Mehrheit der Schülerinnen und Schüler (83 %) weist nach der Teilnahme an der MINIPHÄNOMENTA eine ‚begeisterte, sehr positive oder positive Einstellung gegenüber Phänomenen und ihrer Bearbeitung‘ auf.
- Die von den Lehrerinnen und Lehrern geäußerte These der Begeisterung, im Prätest widerlegt, kann für den Posttest angenommen werden.

Ich fahre nun mit den Komponenten der [Einstellung] fort. Das folgende Diagramm stellt die durchschnittlich erreichten relativen Punktwerte der einzelnen Dimensionen von Prä- und Posttest gegenüber:

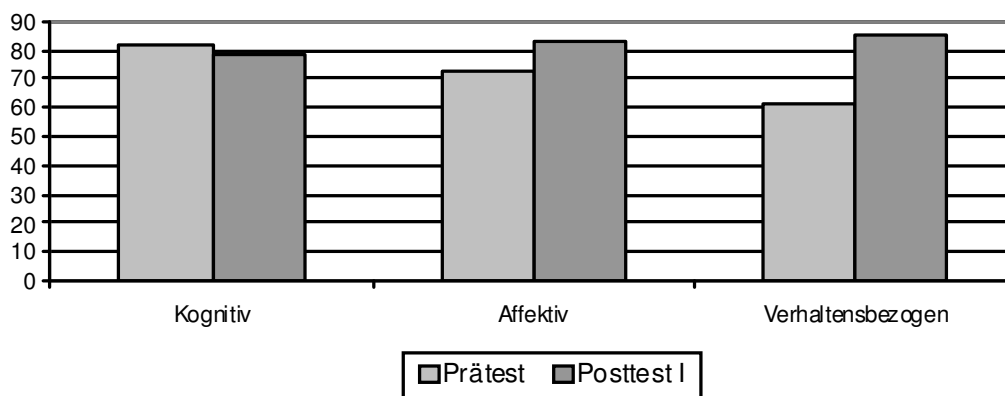


Diagramm 3.2.2.1.5: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Posttest I; - durchschnittlich erreichte relative Rohpunktsumme in den einzelnen Dimensionen der Einstellung
Der Verfasser

Die Differenzen der durchschnittlich erreichten relativen Punktzahlen in den einzelnen Dimensionen haben sich nach der Teilnahme an der MINIPHÄNOMENTA deutlich angeglichen. War die ‚kognitive Dimension‘ im Prätest diejenige mit der durchschnittlich höchsten relativen Punktzahl, gefolgt von der ‚affektiven Dimension und schließlich der ‚verhaltensbezogenen Dimension‘, so hat sich dieses Bild grundlegend geändert. Die [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] ist nicht mehr durch ‚kognitive Faktoren‘ dominiert, ganz im Gegenteil ging der ‚kognitive Anteil‘ mit einem t-Wert von -6,0 und $df = 216$ sogar signifikant zurück.

Wird das Experimentieren der Schülerinnen und Schüler damit zur Spielerei? Die Frage muss entschieden verneint werden. Zum einen lässt sich ein derartig linearer Zusammenhang von Verhalten und [Einstellung] nicht sinnvoll begründen, da die [Einstellung] nur eine Verhaltensdisposition (vgl. Abschnitt 1.4.2 in dieser Arbeit) unter vielen ist, zum anderen ist der ‚kognitive Anteil‘ immer noch hoch. Vielmehr konnten die Schülerinnen und Schüler im Zuge der MINPHÄNOMENTA einen deutlich ‚emotionalen Zugang‘ zum Experimentieren aufbauen. Weiterhin verlor das Experimentieren im Zuge der MINPHÄNOMENTA seine Passivität und wurde zu einer aktiven Tätigkeit.

Die Prätestergebnisse zeigten, vor allem auch in Bezug auf die differentialpsychologische Auswertung, sehr problematische Aspekte, so war die [Einstellung] der Mädchen negativer als die der Jungen und sank deutlich mit steigendem Alter der Schülerinnen und Schüler. Die folgende Tabelle stellt die Ergebnisse der differentialpsychologischen Analyse aus Prä- und Posttest gegenüber:

Prätest	Arithmetisches Mittel	Standardabweichung
Mädchen	40,2	9,8
Jungen	45,0	7,8
Posttest	Arithmetisches Mittel	Standardabweichung
Mädchen	46,6	9,0
Jungen	51,7	6,2
Prätest	Arithmetisches Mittel	Standardabweichung
7 - 8	49,5	7,6
9	45,2	6,2
10 - 11	35,4	6,8
Posttest	Arithmetisches Mittel	Standardabweichung
7 - 8	52,4	5,4
9	52,0	6,8
10 - 11	44,9	8,7

Tabelle 3.2.2.1.12: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Posttest I;
- Werte der differentialpsychologischen Analyse
Der Verfasser

Insgesamt kann festgestellt werden, dass die signifikanten alters- und geschlechtsspezifischen Unterschiede bestehen bleiben. Mädchen haben damit auch nach der Teilnahme an der MINIPHÄNOMENTA eine negativere [Einstellung] als Jungen. Weiterhin sinkt auch nach dem treatment die [Einstellung] mit steigendem Alter. Insofern war die MINIPHÄNOMENTA nicht in der Lage diese Unterschiede zu nivellieren. Positiv kann aber festgestellt werden, dass die MINIPHÄNOMENTA in allen Subgruppen der Stichprobe wirksam war.

Die MINIPHÄNOMENTA war in Bezug auf die Variable [Einstellung gegenüber Phänomen und ihrer Bearbeitung]

HOCHWIRKSAM (Ebene I)

1. hoher Effekt des treatments mit einem geringen Überlappungsbereich der Prä- und Posttestverteilung von 68%
2. deutlicher Einfluss auf die Dimensionen der [Einstellung]: Die Schülerinnen und Schüler erhielten einen handlungsorientierten und affektiv gefärbten Zugang zu Experimenten.
3. Geschlechts- und altersspezifische Unterschiede blieben bestehen, dies jedoch auf insgesamt deutlich erhöhtem Niveau. Damit ist die MINIPHÄNOMENTA auf die Subgruppen zu 100% wirksam.

Die Variable formale Kompetenz

Die folgenden Ergebnisse für diese Variable basieren auf insgesamt 47 ausgewerteten Beobachtungsbögen. Vergleicht man die Anzahl der Beobachtungen mit der des Prätests, so fällt auf, dass selbst wenn man mit einbezieht, dass im Prätest Kontroll- und Experimentalgruppe zusammengefasst wurden, hier deutlich weniger Beobachtungen gemacht werden konnten. Diese Tatsache ist durch den Umstand zu erklären, dass die Schülerinnen und Schüler im Durchschnitt deutlich länger experimentierten. Man könnte geneigt sein, diese Tatsache negativ zu interpretieren. Es wäre möglich, dass die Schülerinnen und Schüler länger gebraucht haben, um das Experiment zu verstehen, mit ihm umzugehen und Ansätze zur Erklärung zu äußern. Dies war jedoch keineswegs der Fall. Vielmehr gingen sie mit dem Experiment deutlich kompetenter um. Waren die Schülerinnen und Schüler im Prätest eher „steif“, hatten Berührungsängste und zeigten sich experimentell unerfahren, so hat sich dieses Bild deutlich verändert. In der folgenden Analyse soll dieser zunächst subjektive Eindruck durch empirisches Material untermauert werden.

Ich beginne mit der Gegenüberstellung der Rahmendaten:

	Prätest			Posttest I		
	Dauer des Experimentierens	Klima beim Experimentieren	Konzentration beim Experimentieren	Dauer des Experimentierens	Klima beim Experimentieren	Konzentration beim Experimentieren
Arithmetisches Mittel	4,4	2,5	2,7	7,1	3,4	3,4
Standardabweichung	2,0	0,9	1,0	2,1	0,7	0,6

Tabelle 3.2.2.1.13: Formale Kompetenz; Posttest I;
- deskriptive Daten
Der Verfasser

Die folgende Tabelle zeigt die Ergebnisse des t-Tests für unabhängige Stichproben, der auf der Basis der in Tabelle 3.2.2.1.14 dargestellten Werte durchgeführt wurde:

	Dauer des Experimentierens	Klima beim Experimentieren	Konzentration beim Experimentieren
df	46	46	46
t-Wert	-13,6	-7,5	6,7

Tabelle 3.2.2.1.14: Formale Kompetenz; Posttest I;
- Ergebnisse des t-Tests
Der Verfasser

Damit haben sich nach der MINIPHÄNOMENTA die ‚Konzentration‘ der Schülerinnen und Schüler, sowie das ‚Klima beim Experimentieren‘ deutlich verbessert. Weiterhin nahm die Zeit, die die Schülerinnen und Schüler experimentierten und sich gedanklich mit dem Experiment beschäftigten, statistisch bedeutsam zu, sie verdoppelte sich nahezu.

Die folgende Tabelle stellt einen Zusammenhang zwischen den bereits beschriebenen Basisvariablen (Dauer des Experimentierens, Klima während des Experimentierens und Konzentration auf die Aufgabe) sowie der Klassenstufenzugehörigkeit und dem Geschlecht der Schülerinnen und Schüler her. Im Prätest konnte gezeigt werden, dass die Werte der Variablen mit steigender Klassenstufe und vor allem bei Mädchen stark zunehmen:

Prätest	Arithmetisches Mittel			Standardabweichung		
	Dauer	Klima	Konzentration	Dauer	Klima	Konzentration
Jungen	3,7	2,2	2,3	1,9	0,8	0,9
Mädchen	5,1	2,9	3,0	1,7	0,9	0,9
Posttest I	Arithmetisches Mittel			Standardabweichung		
	Dauer	Klima	Konzentration	Dauer	Klima	Konzentration
Jungen	6,2	3,3	3,3	2,0	0,9	0,7
Mädchen	8,0	3,4	3,4	1,8	0,5	0,5
Prätest	Arithmetisches Mittel			Standardabweichung		
	Dauer	Klima	Konzentration	Dauer	Klima	Konzentration
Dritte Klasse	3,7	2,4	2,5	1,8	0,8	1,0
Vierte Klasse	5,2	2,7	2,8	1,9	1,0	1,0
Posttest I	Arithmetisches Mittel			Standardabweichung		
	Dauer	Klima	Konzentration	Dauer	Klima	Konzentration
Dritte Klasse	5,8	3,3	3,2	1,7	0,8	0,7
Vierte Klasse	8,7	3,5	3,6	1,4	0,7	0,5

Tabelle 3.2.2.1.15: Formale Kompetenz; Posttest I;
- der Einfluss der differential-psychologischen Aspekte auf die Basisvariablen
Der Verfasser

Signifikante Unterschiede ergeben sich zwischen der Klassenstufenzugehörigkeit bzw. der Geschlechtszugehörigkeit für den Posttest I lediglich für die Variable ‚Dauer des Experimentierens‘. Hier arbeiten Mädchen und Kinder der vierten Klassenstufe deutlich länger, als Jungen und Kinder der dritten Klassenstufe. Trotz der weiteren Existenz dieser Unterschiede darf nicht vergessen werden, dass die erzielten Ergebnisse konsistent auf einem höheren Niveau liegen. Es gibt folglich für alle vier Gruppen eine Steigerung der ‚Experimentierdauer‘.

Die MINIPHÄNOMENTA verbesserte somit die definierten Rahmenbedingungen des ‚Experimentierens‘ in allen Kategorien deutlich. Die MINIPHÄNOMENTA war jedoch nicht in der Lage, sämtliche geschlechts- und altersspezifischen Unterschiede zu nivellieren.

Nachdem der Einfluss der MINIPHÄNOMENTA auf die Rahmenaspekte des Experimentierens geklärt ist, folgt die Analyse der Wirksamkeit von ASIP auf die [formale Kompetenz] selbst. Die folgende Tabelle stellt einige wichtige deskriptive Daten der individuell erreichten kumulierten Rohpunktwerte zusammen:

	Prätest	Posttest I
Arithmetisches Mittel	33,4	50,5
Konfidenzintervall (95%)	31,9 – 34,8	47,3 – 53,7
Modalwert	29	50
Standardabweichung	9,2	10,7
range	39	46

Tabelle 3.2.2.1.16: Formale Kompetenz; Posttest I;
- deskriptive Daten
Der Verfasser

Der Unterschied in Bezug auf die Mittelwerte zwischen den beiden Verteilungen ist gravierend. Die MINIPHÄNOMENTA hat einen sehr großen Einfluss auf die formale Kompetenz der Schülerinnen und Schüler. Im Durchschnitt erreichten sie im Posttest I 17,1 Testrohpunkte mehr als vor der Teilnahme an ASIP. Dieser Unterschied ist mit einem t-Wert von 10,8 und 45 Freiheitsgraden hoch signifikant. Mit einer geringen Überlappung der Prä- und der Posttestverteilung von 36% konnte ein hoher Effekt ermittelt werden. Auffällig ist weiterhin, dass die Dispersionsmaße, ohnehin im Prätest schon relativ hoch, weiter zugenommen haben. Die formale Kompetenz scheint ein höchst heterogenes Merkmal zu sein, wobei ASIP diese Heterogenität weiter zu fördern scheint. Die formale Kompetenz wird somit durch ASIP nicht bei allen Schülerinnen und Schülern gleichmäßig entwickelt.

Auch bei der Verteilung im Posttest I handelt es sich mit einem Kolmogorov-Smirnov Z-Wert von 0,8 um eine Normalverteilung. Diese ist jedoch mit einer Schiefe von -0,7 deutlich rechtssteil.

Um die Eigenschaften beider Verteilungen besser vergleichen zu können, wird auch hier eine Zuweisung der individuell erreichten kumulierten Testwerte zu Kategorien vorgenommen. Ich verwende an dieser Stelle die im Abschnitt 3.2.1.2 eingeführten Kategorien und cut-off-scores. Aufgrund des deutlich unterschiedlichen Stichprobenumfanges verwende ich für die folgende Grafik relative Werte:

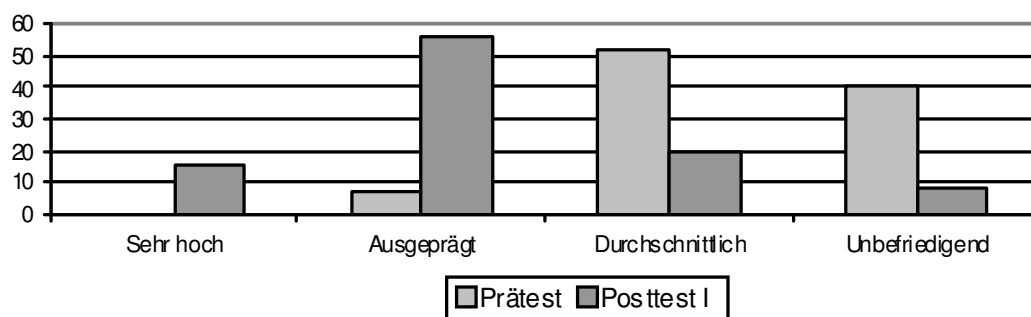


Diagramm 3.2.2.1.6: Formale Kompetenz; Posttest I
- Kategorien der formalen Kompetenz
Der Verfasser

Die Kategorienhäufigkeiten der formalen Kompetenz hat sich, vergleicht man den Prä- und Posttest I, grundlegend geändert: Waren die meisten Schülerinnen und Schüler des Prätests den Kategorien ‚durchschnittlich‘ bis ‚unbefriedigend‘ zuzuordnen, so ist der Großteil der Leistungen im Posttest im ‚ausgeprägten bis sehr guten Bereich‘. Nur noch ein sehr geringer Anteil der Schülerinnen und Schüler (9%) zeigte eine ‚unbefriedigende formale Kompetenz‘.

Die [formale Kompetenz] der Schülerinnen und Schüler haben sich durch die Teilnahme an der MINIPHÄNOMENTA entscheidend verbessert. Sie liegen überwiegend im ‚ausgeprägten bis sehr guten Bereich‘.

Nach der Klärung des grundsätzlichen Einflusses der MINIPHÄNOMENTA sollen nun zwei Aspekte vertiefend untersucht werden. Der erste betrifft die Dimensionen der [formalen Kompetenz], der zweite den Einfluss der MINIPHÄNOMENTA auf differentialpsychologische Fragestellungen.

Zunächst zur ersten Fragestellung: Die Tabelle zeigt die durchschnittlich erreichten relativen Punktzahlen in den einzelnen Dimensionen des Posttests I:

	Experimentelle Kompetenzen	Spezifisch naturwissenschaftliches Denken	Logisches Denken
Durchschnittlich erreichte relative Punktzahl in der Dimension	76,4	68,6	73,7

Tabelle 3.2.2.1.17: Formale Kompetenz; Posttest I
- Ergebnisse der Dimensionen
Der Verfasser

Die MINIPHÄNOMENTA war in allen drei Bereichen wirksam. Die grundsätzlichen Tendenzen sind aber dennoch erhalten geblieben: Die ‚experimentellen Kompetenzen‘ sind am ausgeprägtesten. Ihnen folgen das ‚logische Denken‘ und schließlich das ‚spezifisch-naturwissenschaftliche Denken‘. Die Kompetenzen konnten in allen drei Bereichen entscheidend gestärkt werden.

Abschließend sei an dieser Stelle auf die differentialpsychologischen Aspekte eingegangen: Im Prätest erreichten die Schülerinnen und Schüler der vierten Klasse, sowie die Mädchen einen höheren Punktwert als Kinder der Klassenstufe drei und die männlichen Untersuchungsteilnehmer. Der Einfluss der beiden Faktoren wird anhand eines t-Testes für unabhängige Stichproben durchgeführt, dessen Ergebnisse in der folgenden Tabelle wiedergegeben sind:

	Arithmetisches Mittel	Standardabweichung	Freiheitsgrade	t-Wert
Dritte Klasse	47,5	11,5	19	-1,9
Vierte Klasse	54,5	8,3		
Jungen	48,0	12,5	21	-1,7
Mädchen	52,9	8,5		

Tabelle 3.2.2.1.19: Formale Kompetenz; Posttest I;
- Ergebnisse des t-Testes
Der Verfasser

Die MINIPHÄNOMENTA war auf alle Subgruppen der Stichprobe wirksam. Signifikante Unterschiede in Bezug auf Klassenzugehörigkeit oder Geschlecht existieren nach der Teilnahme an der MINIPHÄNOMENTA nicht mehr.

Die MINIPHÄNOMENTA war in Bezug auf die Variable [formale Kompetenz]

HOCHWIRKSAM (Ebene I)

- Die ‚Rahmenbedingungen des Experimentierens‘ verbesserten sich entscheidend.
- Mit einem Überlappungsbereich der Prä- und der Posttestverteilung von 36% konnte ein hoher Effekt nachgewiesen werden.
- Die MINIPHÄNOMENTA war in Bezug auf alle Dimensionen der [formalen Kompetenz] wirksam.
- Die MINIPHÄNOMENTA förderte die ‚experimentellen Kompetenzen‘ der Schülerinnen und Schüler so umfassend, dass Geschlechts- und Klassenzugehörigkeitsunterschiede statistisch unbedeutsam wurden. Sie ist damit auf 100% der Subgruppen wirksam.

Die Variable aktuelle Motivation

Die folgende Analyse basiert auf insgesamt 110 ausgewerteten Fragebögen zur [aktuellen Motivation], die von Schülerinnen und Schülern der vierten Klasse bearbeitet wurden. Wie bereits in den Abschnitten 2.4.6 und 3.2.1.2 in dieser Arbeit dargestellt, kann das Instrument des modifizierten FAM leider erst ab dieser Klassenstufe eingesetzt werden, da vorher die Motivationsstruktur der Schülerinnen und Schüler nicht ausreichend entwickelt ist.

Ich beginne die Auswertung mit der folgenden Tabelle, in der einige wichtige deskriptive Daten der Prä- und Posttestverteilung der individuell erreichten kumulierten Rohtestpunkte einander gegenübergestellt werden:

	Prätest	Posttest
Arithmetisches Mittel	45,6	53,5
Konfidenzintervall (95%)	44,4 – 46,8	52,3 – 54,6
Modalwert	41,0	52
Standardabweichung	9,0	6,0
range	41	28

Tabelle 3.2.2.1.19: Motivation; Posttest I;
- deskriptive Werte
Der Verfasser

Beide Verteilungen zeigen deutlich unterschiedliche Eigenschaften: Die Mittelwertsdifferenz zwischen dem Prä- und dem Posttest I beträgt 7,9 Punkte. Damit ist die [Motivation] der Schülerinnen und Schüler sich tiefgehend mit einem Phänomen zu beschäftigen und sich um eine eigene Erklärung zu bemühen, nach der Teilnahme an der MINIPHÄNOMENTA deutlich gestiegen. Die Mittelwertsdifferenz ist hoch signifikant (t-Wert = 4,1, df = 109). Von der Voraussetzung des statistisch bedeutsamen Unterschiedes ausgehend wurde die Effektstärke ermittelt. Der Überlappungsbereich der Prä- und der Posttestverteilung (vgl. Bortz, Döring 2006, S. 607 f.) ist mit 68% gering. Es gibt folglich einen hochwirksamen Einfluss der MINIPHÄNOMENTA auf die Motivation der Kinder der vierten Klassenstufe. Ebenfalls erfreulich ist die deutliche Abnahme der Dispersionsmaße. Damit ist die Motivation der Schülerinnen und Schüler homogen auf einem hohen Niveau.

Die Posttestverteilung ist, dies ergab der Kolmogorov-Smirnov-Test, hinreichend normalverteilt (z-Wert = 0,6). Sie ist jedoch, dies beweist die Schiefe von -0,3, rechtssteil. Damit haben mehr Schülerinnen und Schüler eine positive, das heißt eine ‚hohe Motivation‘, als eine ‚unbefriedigende Motivation‘.

Entsprechend dem Vorgehen im Prätest, sollen nun hier auch die individuell erreichten kumulierten Rohtestpunkte Kategorien zugewiesen werden. Um die Vergleichbarkeit der Daten sicher zu stellen, verwende ich hier die gleichen Kategorien und cut-off-scores, die auch zur Interpretation der Prätestergebnisse verwendet wurden. Da beide Datenreihen deutlich unterschiedliche Versuchspersonenzahlen aufweisen, werden die folgenden Werte relativ dargestellt:

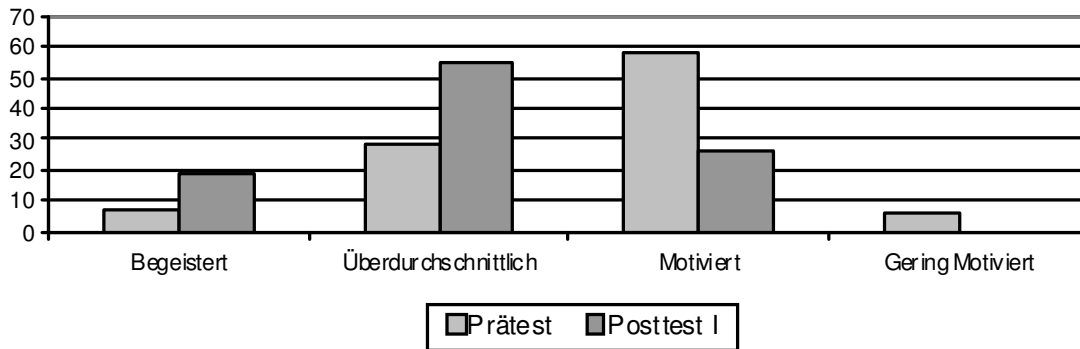


Diagramm 3.2.2.1.7: Motivation; Posttest I;
- Motivationskategorien
Der Verfasser

Die Grafik zeigt den immensen Einfluss der MINIPHÄNOMENTA auf die [Motivation] der Schülerinnen und Schüler: War im Prätest die Kategorie ‚motiviert‘ die am häufigsten besetzte und gab es nur wenige Schülerinnen und Schüler mit einer ‚begeistert hohen Motivation‘, so hat sich dies Bild nach der Teilnahme an ASIP deutlich gewandelt. Die Kategorie ‚überdurchschnittlich hoch motiviert‘ wird nun von 55% der Schülerinnen und Schüler erreicht. 18% der Untersuchungsteilnehmerinnen und Untersuchungsteilnehmer erreichten sogar die oberste Kategorie, ‚begeistert hohe Motivation‘. Weiterhin ist der Umstand erfreulich, dass keiner der Schülerinnen und Schüler im Posttest eine ‚unterdurchschnittliche Motivation‘ aufwies. War die Motivation der Kinder im Vorfeld der MINIPHÄNOMENTA bereits überraschend hoch, sich mit einem Versuch genauer auseinander zu setzen, eine eigene, subjektiv tragfähige Erklärung zu finden, so muss sie nun als sehr hoch bezeichnet werden.

Diese Daten widerlegen ein oft vorgebrachtes Argument gegen das Prinzip des genetischen Lernens. Viele pädagogische Professionelle (Lehrerinnen und Lehrer und Erzieherinnen und Erzieher) befürchten durch die Nichtgabe der Erklärungen zu den Versuchen die Schülerinnen und Schüler zu frustrieren und damit langfristig eher zu verschrecken als für naturwissenschaftliche Fragestellungen zu begeistern. Dies sind typische Berührungsprobleme und Ängste der in der Praxis Tätigen. Bei der MINIPHÄNOMENTA werden jedoch keinerlei Erklärungen gegeben (vgl. Abschnitt 1.1 und 1.3.2 in dieser Arbeit). Die [Motivation] der Schülerinnen und Schüler sinkt aber durch dieses Vorgehen nicht, vielmehr steigt sie, dies zeigen die obigen Daten, drastisch an.

Infolge der Teilnahme an der MINIPHÄNOMENTA konnte drei Monate später eine deutlich **höhere Motivation** der Schülerinnen und Schüler gemessen werden, sich tiefgehend mit naturwissenschaftlichen Versuchen auseinanderzusetzen.

Damit sind die grundsätzlichen Tendenzen der Daten ausgewertet. Abschließend sollen in Bezug auf diese Variable zwei Aspekte herausgearbeitet werden:

1. Der Einfluss der MINIPHÄNOMENTA auf die einzelnen Dimensionen der [Motivation]
2. Differentialpsychologische Analyse des Einflusses der MINIPHÄNOMENTA

Ad 1: Das folgende Diagramm gibt die durchschnittlich erreichten Punktwerte in jeder Dimension für Prä- und Posttest I wieder:

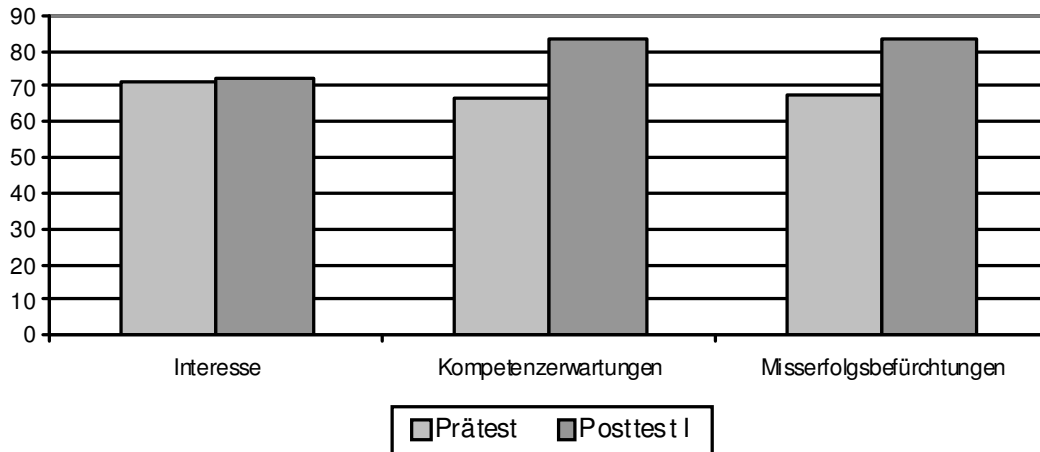


Diagramm 3.2.2.1.8: Motivation; Posttest I;
- Dimensionen der Motivation
Der Verfasser

Es ergibt sich ein erstaunliches Bild: Im Prätest äußerten die Schülerinnen und Schüler ein starkes ‚Interesse‘ an der Fragestellung. Darüber hinaus zeigten sie durchschnittliche ‚Kompetenzerwartungen‘ und ‚Misserfolgsbefürchtungen‘. Auf das ohnehin schon hohe ‚Interesse‘ an derartigen Fragestellungen hatte die MINIPHÄNOMENTA nur geringen Einfluss. Anders bei den ‚Kompetenzerwartungen‘ und den ‚Misserfolgsbefürchtungen‘. In beiden Bereichen hatte das Projekt großen Einfluss: Nach der Teilnahme schätzten die Schülerinnen und Schüler ihre eigenen Kompetenzen positiver ein und äußerten weniger ‚Misserfolgsbefürchtungen‘⁵⁰.

Ad 2: Im Prätest bestanden in Bezug auf die Ergebnisse der Jungen und Mädchen deutliche Unterschiede: Die Jungen hatten eine hoch signifikant höhere Motivation, die Aufgabe zu bearbeiten, als die Mädchen. In der folgenden Tabelle sind die Ergebnisse der Überprüfung dieses Unterschiedes im Posttest wiedergegeben. Sie geschieht durch einen t-Test für unabhängige Stichproben:

	Arithmetisches Mittel	Standardabweichung	Freiheitsgrade	t-Wert
Jungen	54,0	4,1	48	2,2
Mädchen	53,0	7,2		

Tabelle 3.2.2.1.20: Motivation; Posttest I
- Signifikanzüberprüfung durch einen t-Test für unabhängige Stichproben
Der Verfasser

⁵⁰ Ein hoher Punktwert zeigt geringe Misserfolgsbefürchtungen an.

Die Punktedifferenz zwischen den Jungen und Mädchen ist damit deutlich geringer geworden. Trotzdem ist sie signifikant, so dass weiterhin ein bedeutsamer geschlechtsspezifischer Unterschied im Antwortverhalten zwischen Mädchen und Jungen besteht.

Die MINIPHÄNOMENTA war in Bezug auf die [aktuelle Motivation zur tiefergehenden kognitiven Beschäftigung mit einem Phänomen]:

HOCHWIRKSAM (Ebene I)

- Es wurde eine hohe Effektstärke ermittelt, dies ergibt der geringe Überlappungsbereich der Prä- und Posttestverteilung von 68%.
- Der größte Teil der Schülerinnen und Schüler ist ‚begeistert‘, ‚hoch motiviert‘ oder ‚überdurchschnittlich hoch motiviert‘. Keiner der Untersuchungsteilnehmerinnen oder Untersuchungsteilnehmer ist ‚unbefriedigend motiviert‘.
- Die MINIPHÄNOMENTA ist insbesondere auf die ‚Kompetenzerwartungen‘ und die ‚Misserfolgsbefürchtungen‘ wirksam.
- Die geschlechtsspezifischen Unterschiede in der Motivationshöhe haben deutlich abgenommen, sind aber immer noch statistisch bedeutsam. Die MINIPHÄNOMENTA ist somit zu 100% auf die Subgruppen der Stichprobe wirksam.

Die Variable Wissen

Im Posttest I konnten in Bezug auf die Variable [Wissen] insgesamt 84 Gruppenarbeiten von Schülerinnen und Schülern der dritten und vierten Klassenstufe ausgewertet werden. Die Ergebnisse der Inhaltsanalyse können, dies wurde im Abschnitt 3.2.1.2, unter der Verwendung von zwei Indizes (Qualität und Quantität) zusammengefasst werden. Es handelt sich um additive gewichtete Indices, deren Gewichte faktoranalytisch bestimmt wurden.

Die folgende Tabelle stellt zentrale deskriptive Werte der Indizes des Posttests I in Relation zu den Ergebnissen des Prätests dar:

	Index I (Qualität)	Index II (Umfang)
Prätest		
Arithmetisches Mittel	1,7	4,8
Standardabweichung	0,6	1,2
Posttest I		
Arithmetisches Mittel	4,1	5,1
Standardabweichung	1,4	0,3

Tabelle 3.2.2.1.21: Wissen; Posttest I;
- deskriptive Werte des Prä- und Posttests I
Der Verfasser

Die Werte zeigen deutlich den Einfluss der MINIPHÄNOMENTA. Sowohl der Index, der die Qualität, als auch derjenige, der die Quantität der Lösung beschreibt, sind gestiegen. Dabei ist allerdings zu beachten, dass der Index I deutlich stärker nach der Teilnahme an der MINIPHÄNOMENTA gestiegen ist, als der Index II. Beide Mittelwertunterschiede werden durch einen t-Test für unabhängige Stichproben auf statistische Bedeutsamkeit überprüft. Die Ergebnisse des Tests sind in der folgenden Tabelle zusammengefasst:

	Mittelwertsdifferenz	Freiheitsgrade	t-Wert
Index I Prätest – Index I Posttest I	2,4	83	16,5
Index II Prätest – Index II Posttest I	0,3	83	4,6

Tabelle 3.2.2.1.22: Wissen; Posttest I;
- Ergebnisse des t-Testes
Der Verfasser

Die Differenzen der beiden Indizes sind statistisch hoch bedeutsam. Die Effektstärken klaffen hingegen deutlich auseinander. Im Falle des Index I (Qualität) überlagern sich die Prä- und die Posttestverteilung (vgl. Bortz, Döring 2006, S. 607 f.) mit 19% sehr gering. Der Index II (Quantität) weist hingegen eine deutliche höhere Überlagerung der beiden Verteilungen von 88% auf, die lediglich im mittleren Effektstärkebereich liegt. Insgesamt lässt sich aber trotzdem die Aussage treffen, dass die MINIPHÄNOMENTA die Schülerinnen und Schüler umfassend gefördert hat: Die Kinder lernten infolge der Teilnahme an der MINIPHÄNOMENTA einen Versuch durchzuführen, genau zu beobachten und auf Basis dieser Beobachtungen eigene Ideen und Gedanken zur Begründung der Phänomene anzuführen.

Wie bereits bei den anderen Variablen geschehen, sollen auch diese Messwertreihen Kategorien zugewiesen werden. Um eine hohe Vergleichbarkeit sicher zu stellen, verwende ich die gleichen Kategorien und cut-off-scores, wie im Prätest. Die beiden folgenden Diagramme stellen die Ergebnisse für die beiden Indizes dar:

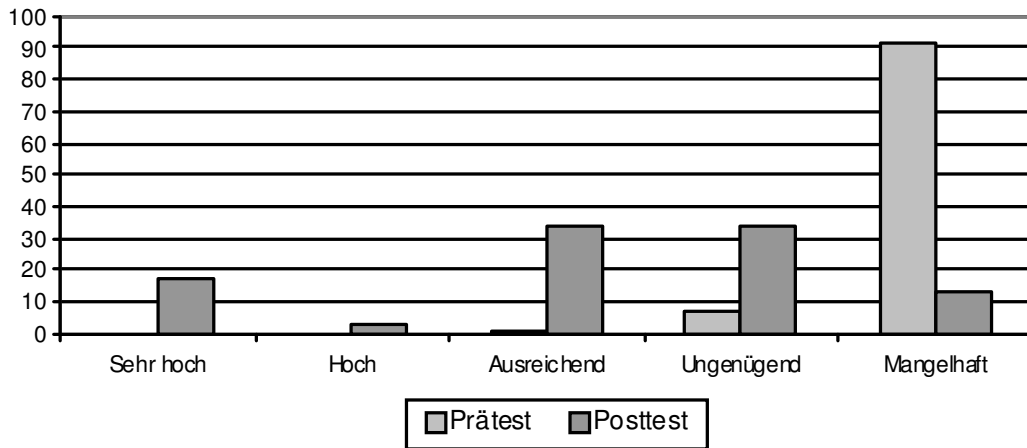


Diagramm 3.2.2.1.9: Wissen; Posttest I;
- Kategorienhäufigkeiten zum Index I (Qualität)
Der Verfasser

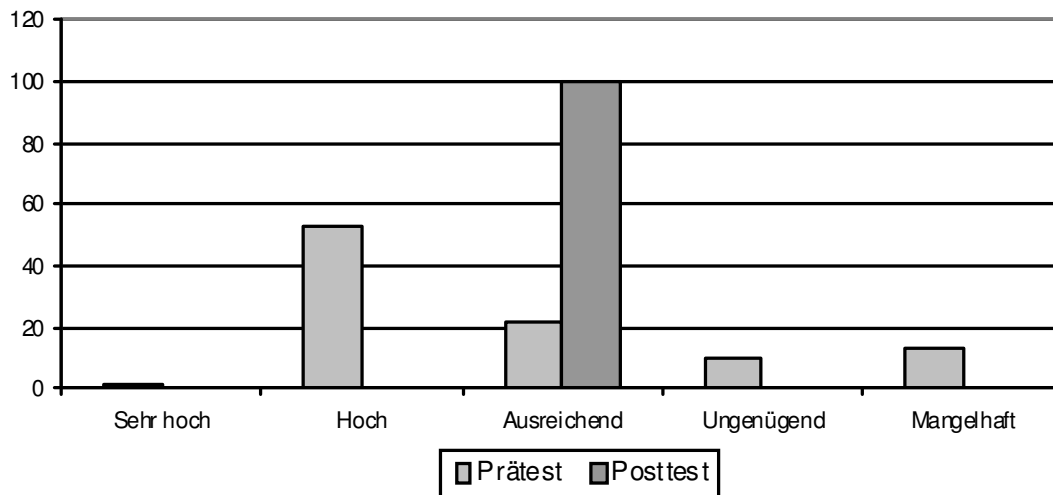


Diagramme 3.2.2.1.10: Wissen; Posttest I;
- Kategorienhäufigkeiten zum Index II (Quantität)
Der Verfasser

In Bezug auf den ersten Index I (Qualität) ergibt sich ein deutliches und überaus erfreuliches Bild. War die ‚inhaltliche Qualität‘ der Arbeiten im Prätest fast durchgängig ungenügend, so hat sich dies durch die MINIPHÄNOMENTA grundlegend verändert: Nun ist mehr als die Hälfte der Schülerinnen und Schüler in der Lage die Aufgabe mit einer ‚sehr guten bis ausreichenden Qualität‘ zu bearbeiten. Bedenklich ist, dass immer noch 46% der Lösungen der Kinder eine ‚mangelhafte bis ungenügende Qualität‘ aufweisen. Bedenkt man aber das äußerst niedrige Prätestniveau der Aufgaben, so stellen die Veränderungen insgesamt eine erhebliche Verbesserung der Sachlage dar.

War der Umfang der Lösungen der Schülerinnen und Schüler insgesamt hoch signifikant gestiegen, so offenbart die Zuweisung des Index II (Quantität) zu den Kategorien ein interessantes Bild: Im Vorfeld der Teilnahme an der MINIPHÄNOMENTA streuten die Werte deutlich. Die Schülerinnen und Schüler schienen der möglichst vollständigen Bearbeitung der Aufgabe ein großes Gewicht einzuräumen. Dies scheint sich nach der Teilnahme an der MINIPHÄNOMENTA deutlich geändert zu haben. Alle Schülerinnen und Schüler sind aufgrund ihrer Werte der Kategorie ‚ausreichender Umfang‘ zugewiesen worden. Die Kinder der Experimentalgruppe scheinen damit infolge des Projektes die Wichtigkeit eines hohen Umfanges der Bearbeitung der Aufgabe nicht mehr so hoch einzustufen. Der Vergleich mit dem Diagramm 3.2.2.1.10 ergibt, dass sie ihr Augenmerk nun eher auf die ‚Qualität‘ der Beantwortung der Aufgabe richten. Trotzdem wird dabei der Umfang nicht vernachlässigt. Er ist in allen Fällen, dies sei an dieser Stelle nochmals betont, im ausreichenden Bereich.

Der Einfluss der MINIPHÄNOMENTA ist in Bezug auf beide Indizes (Qualität und Quantität) deutlich zu erkennen:

- Die ‚Qualität‘ der Lösungen stieg mit einer Effektgröße von 2,6 beträchtlich.
- Der Umfang der Lösungen stabilisierte sich mit einer Effektgröße von 0,3 auf einem durchschnittlichen Niveau.
- Beides zusammen kann als eine Verschiebung des Fokus interpretiert werden: Die Schülerinnen und Schüler richten ihr Augenmerk nach der Teilnahme an der MINIPHÄNOMENTA deutlich stärker auf die ‚Qualität‘ ihrer Aussagen anstatt auf die ‚Quantität‘, ohne diese jedoch dabei zu vernachlässigen.

Nachdem die groben Daten des Einflusses der MINIPHÄNOMENTA offen gelegt wurden, fahre ich an dieser Stelle mit der differentialpsychologischen Analyse fort. Im Prätest konnte der Einfluss des Geschlechtes, sowie der der Klassezugehörigkeit eindeutig nachgewiesen werden. Mädchen lösten die Aufgabe hoch signifikant besser als Jungen und Schülerinnen und Schüler der vierten Klassenstufe schnitten ebenfalls besser ab als jene der Klassenstufe drei. Dies gilt sowohl für die ‚Qualität‘ der Lösung, als auch für deren ‚Quantität‘. Die folgende Tabelle stellt die Mittelwertsdifferenzen und die Ergebnisse der durchgeführten t-Tests für unabhängige Stichproben dar:

Nummer	Index	Mittelwertsdifferenz	t-Wert	Freiheitsgrade
Klassenzugehörigkeit				
1	Index I (Qualität)	Schülerinnen und Schüler der vierten Klassenstufe 0,7 Indexpunkte mehr	-3,5	36
2	Index II (Quantität)	Schülerinnen und Schüler der vierten Klassenstufe 0,1 Indexpunkte mehr	-0,8	36
Geschlecht				
1	Index I (Qualität)	Mädchen 1,7 Indexpunkte höher	-6,9	36
2	Index II (Quantität)	Mädchen 0,1 Indexpunkte höher	-1,5	36

Tabelle 3.2.2.1.23: Wissen; Posttest I;
- Ergebnisse der t-Tests
Der Verfasser

Die Tabelle zeigt deutlich die Abschwächung der differentialpsychologischen Merkmale: Bestanden vorher in allen Fällen hoch signifikante Unterschiede, so ist dies im Nachhinein der Teilnahme an der MINIPHÄNOMENTA nur noch in einem Fall gegeben. Der Index II (,Quantität') zeigt keinerlei Abhängigkeit mehr von differentialpsychologischen Merkmalen. Der die ,Qualität' der Lösung beschreibende Index I ist hingegen von der Klassenstufenzugehörigkeit und vor allem vom Geschlecht der Schülerinnen und Schüler abhängig.

Insgesamt kann festgestellt werden, dass die MINIPHÄNOMENTA in Bezug auf die Variable [Wissen]

HOCHWIRKSAM (Ebene I)

ist:

- Die Effektgrößen der beiden Indizes sind deutlich unterschiedlich. Sie liegen mit einer Überlappung der Prä- und der Posttestverteilungen von 19% (Index I) und 88% (Index II) im hohen bzw. im mittleren Bereich. Mit einer mittleren Überlappung beider Verteilungen von 53,5% liegt die Effektgröße im sehr hohen Bereich.
- Die Schülerinnen und Schüler richteten im Posttest I ihren Fokus stärker auf die ,Qualität ihrer Aussagen' als auf deren ,Quantität', ohne dabei diese zu vernachlässigen.
- Der Einfluss differentialpsychologischer Faktoren wurde durch die Teilnahme an der MINIPHÄNOMENTA deutlich abgeschwächt. Das treatment wirkt also zu 100% auf die Subgruppen der Stichprobe.

3.2.2.2 Zusammenfassung der Ergebnisse

Die folgende Tabelle fasst die Wirksamkeitsstufen der einzelnen Variablen, sowie einige zentrale Forschungsbefunde zusammen:

Variable	Wirksamkeitsstufe Ebene I	Zusammenfassung der Ergebnisse
Deskriptive Aspekte	Erinnerung vorhanden: Weitere Analyse sinnvoll	<p>Die Schülerinnen und Schüler erinnern sich im Durchschnitt an 10,8 von 13 gezeigten Stationen.</p> <p>Die Erinnerung an die Stationen muss als umfassend bewertet werden. Im Durchschnitt können die Schülerinnen und Schüler Aufbau und Funktion gut beschreiben. Sie betonen in ihrer Erinnerung insbesondere affektive und verhaltensbezogene Anteile.</p> <p>Die Geschlechtszugehörigkeit hat keinen Einfluss auf Qualität und Umfang der Erinnerung. Das Alter der Schülerinnen und Schüler hingegen übt einen maßgeblichen Einfluss aus.</p>
Allgemeines Experimentierverhalten	HOCHWIRKSAM	<p>Die Schülerinnen und Schüler wiesen nach der Teilnahme eine deutlich höhere Verhaltenskategorie auf. In fast der Hälfte der Fälle erreichten sie Kategorie IV.</p> <p>Die Wahrscheinlichkeit für das Auftreten einer niedrigen Verhaltenskategorie sank erheblich.</p> <p>Die Klassenstufe ist kein Indikator mehr für das Auftreten der Verhaltenskategorie IV. Die Schüler experimentierten zum überwiegenden Teil in Kleingruppen.</p>
Einstellung	HOCHWIRKSAM	<p>Mit einer Überlappung der Prä- und der Posttestverteilung von 68% liegt die Effektstärke im hohen Bereich.</p> <p>Die Einstellung ist damit im Durchschnitt im sehr positiven Bereich.</p> <p>Die Schülerinnen und Schüler erlangten eine affektiv gefärbte und handlungsorientierte Einstellung gegenüber dem Experimentieren.</p> <p>Die geschlechts- und altersspezifischen Unterschiede blieben bestehen, jedoch auf höherem Niveau.</p>
Formale Kompetenz	HOCHWIRKSAM	<p>Die Rahmenbedingungen des Experimentierens (Experimentierzeit, Konzentration auf die Aufgabe usw.) haben sich entscheidend verbessert. Die durchschnittliche kumulierte Prä-Posttest-Differenz ist hoch signifikant.</p> <p>Die MINIPHÄNOMENTA ist in Bezug auf die formale Kompetenz wirksam. Dies drückt sich insbesondere in der geringen Überlappung der Prä- und der Posttestverteilung von 36% aus</p> <p>Geschlechts- und altersspezifische Unterschiede sind im Posttest nicht mehr signifikant.</p>
Aktuelle Motivation	HOCHWIRKSAM	<p>Mit einer Überlappung der Prä- und der Posttestverteilung von 68% ergibt sich ein hoher Effekt.</p> <p>Die Motivation der Schülerinnen und Schüler ist im Mittel überdurchschnittlich hoch bis begeistert.</p> <p>Die MINIPHÄNOMENTA war insbesondere auf die Kompetenz- und Misserfolgserwartung wirksam.</p> <p>Die geschlechtsspezifischen Unterschiede sind immer noch statistisch bedeutsam, haben jedoch abgenommen.</p>

Wissen

HOCHWIRKSAM

Beide Indizes waren im Posttest hoch signifikant stärker ausgeprägt als im Prätest. Im Durchschnitt liegt mit einer Überlappung der Prä- und der Posttestverteilung von 53,5% ein hoher Effekt vor.

Die Schülerinnen und Schüler richteten im Posttest I ihren Focus stärker auf die Qualität ihrer Aussagen.

Der Einfluss differentialpsychologischer Faktoren auf die Variable Wissen wurde durch die MINIPHÄNOMENTA deutlich abgeschwächt.

Insgesamt war damit das treatment **HOCHWIRKSAM**.

Tabelle 3.2.2.2.1: Ergebnisse; Posttest I;
- Zusammenfassung der Ergebnisse des Posttests I
Der Verfasser

Die MINIPHÄNOMENTA war damit in Bezug auf alle Variablen HOCHWIRKSAM. Sie förderte damit die Schülerinnen und Schüler in Bezug auf physikalische Fragestellungen umfassend.

3.2.3 Der Posttest II

3.2.3.1 Die Wirksamkeit von ASIP sechs Monate nach der Teilnahme

Die Variable deskriptive Aspekte

In diesem Abschnitt wird der Frage nachgegangen, ob beziehungsweise wie weit sich die Schülerinnen und Schüler an die Stationen der MINIPHÄNOMENTA erinnern. Für diese Variable werden erneut keine Wirksamkeitsstufen definiert. Wie bereits im Posttest I geschehen, soll auch hier die Erinnerung an die Stationen als Bedingung der Möglichkeit von Wirksamkeit gefasst werden.

Im Posttest I mussten lediglich 29 Fragebögen, dies entspricht 13 Prozent der erhobenen Daten, aufgrund einer zu ausgeprägten Ja-Sage-Tendenz ausgesondert werden. Ein Fragebogen wurde nicht in die Bewertung einbezogen, sofern die Schülerin oder der Schüler an beide in der Schule nicht vorhandenen Stationen eine Erinnerung angab. Die Zahl der nach diesem Kriterium auszusondernden Bögen stieg im Posttest II deutlich an: 46 Fragebögen und damit 22 Prozent der erhobenen Daten wurden wegen einer offensichtlich nicht gewissenhaften Bearbeitung der Fragestellung von der Auswertung ausgeschlossen. Dieser deutliche Anstieg ist aus der Sicht des Verfassers nicht auf die fehlerhafte Erinnerung der Schülerinnen und Schüler zurückzuführen sondern ist vielmehr durch die Wiederholung der ohnehin niedrig-komplexen Aufgabe bedingt⁵¹. Es handelt sich folglich um einen Lern- oder Testungseffekt.

Das positive Bild der Erinnerungsfähigkeit der Schülerinnen und Schüler an die Stationen des Posttests I kann im Posttest II bestätigt werden. Die folgende Tabelle stellt zentrale deskriptive Werte der Anzahl der erinnerten Stationen des Posttest II dar:

Bezeichnung	Posttest II
Arithmetisches Mittel	10,9
Konfidenzintervall (95%)	10,6 – 11,2
Modalwert	13
Standardabweichung	2,0
range	9

Tabelle 3.2.3.1.1: Deskriptive Aspekte; Posttest II;
- deskriptive Werte der Erinnerungsbilder
Der Verfasser

⁵¹ Für diese Erklärung spricht, dass der Effekt bei den übrigen Variablen aufgrund der deutlich komplexeren Aufgabenstellung nicht auftrat.

Die Schülerinnen und Schüler können sich damit auch nach einem Zeitraum von sechs Monaten noch an eine Vielzahl der Stationen der MINIPHÄNOMENTA erinnern. Dabei ist bemerkenswert, dass die Erinnerung in den drei Monaten zwischen Posttest I und Posttest II nicht schlechter geworden ist. Lediglich die Anzahl der Schülerinnen und Schüler, die sich an alle Stationen erinnern konnten ist von 36 Prozent auf 29 Prozent gefallen.

Die Schülerinnen und Schüler konnten sich auch nach einem zeitlichen Abstand von sechs Monaten zur Teilnahme an der MINIPHÄNOMENTA noch umfassend an die Stationen erinnern.

Zum ersten Teil des Fragebogens zur Erfassung der deskriptiven Aspekte soll an dieser Stelle abschließend gefragt werden, ob differentialpsychologische Merkmale, hier das Alter und das Geschlecht, Einfluss auf die Anzahl der erinnerten Stationen haben. Es ist festzustellen, dass es keinen signifikanten Zusammenhang zwischen der Anzahl der erinnerten Station und der Geschlechtszugehörigkeit beziehungsweise dem Alter der Kinder gibt. Dies ist evident, bedenkt man das auch im Posttest I keine Zusammenhänge beziehungsweise Unterschiede nachgewiesen werden konnten.

Es besteht kein Zusammenhang zwischen der Anzahl der erinnerten Stationen und den erfassten differentialpsychologischen Merkmalen.

Ich fahre mit der Auswertung des zweiten Teils des Fragebogens fort. Die folgende Tabelle fasst zentrale deskriptive Werte der Analysekatgorien zusammen:

Bezeichnung der Kategorie	Arithmetisches Mittel	Standardabweichung
Anzahl der Stichpunkte	7,1	2,0
Anzahl der richtigen Stichpunkte	5,3	1,8
Aufbau des Versuches	3,0	0,7
Funktion des Versuches	2,9	0,8
Effekt des Versuches	2,8	0,9
Erklärungsansätze für den Versuch	1,4	0,7
Affektive Äußerungen zum Versuch	3,1	0,7
Eidetischer Charakter der Äußerungen	3,2	0,7

Tabelle 3.2.3.1.2: Deskriptive Aspekte; Posttest II;
- deskriptive Werte der Analysekatgorien
Der Verfasser

Obwohl die Werte insgesamt - im Vergleich zum Posttest I - leicht abgesunken sind, bestätigen sie dennoch die umfassende Erinnerung der Schülerinnen und Schüler an einen Versuch der MINIPHÄNOMENTA. Mit Werten weit oberhalb des Durchschnitts der Skala gelingt es ihnen ‚Aufbau‘, ‚Funktion‘ und ‚Effekt‘ einer Station korrekt zu beschreiben. Besonders erfreulich ist weiterhin, dass die Kinder auch nach sechs Monaten einen ‚affektiv‘ geladenen und handlungsorientierten Zugang zu der Station aufweisen.

Dieses erfreuliche Bild in den Analysekatégorien schlägt sich auch im Gesamtwert der Inhaltsanalyse wieder, dessen zentrale deskriptive Werte in der folgenden Tabelle angeführt werden:

Bezeichnung	Posttest II
Arithmetisches Mittel	16,6
Konfidenzintervall (95%)	16,1 – 17,1
Modalwert	15
Standardabweichung	3,2
range	15

Tabelle 3.2.3.1.4: Deskriptive Aspekte; Posttest II;
- deskriptive Werte der Punktsomme der Inhaltsanalyse
Der Verfasser

Der Gesamtpunktwert ist nur wenig gefallen. Insgesamt erreichten die Versuchspersonen immer noch 69 Prozent der erreichbaren Punkte. Auch nach sechs Monaten ist der Gesamtwert der Inhaltsanalyse damit noch im deutlich erhöhten Bereich der Skala.

Klassifiziert man die individuell erreichten Punktsommen der Inhaltsanalyse des Posttests II entsprechend der in Abschnitt 3.2.2.1 eingeführten Kategorien und cut-off-scores, so ergibt sich das in der Grafik dargestellte Bild:

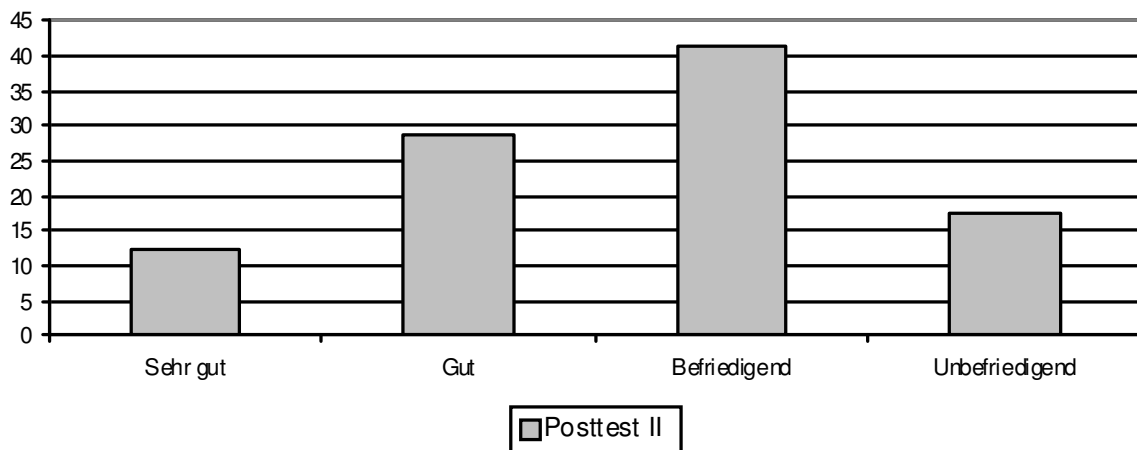


Diagramm 3.2.3.1.1: Deskriptive Aspekte; Posttest II;
- Häufigkeit in den Kategorien
Der Verfasser

Die Qualität der Erinnerung ist auch aufgrund dieser Ergebnisse insgesamt als positiv zu bewerten. 83 Prozent der Schülerinnen und Schüler erreichten eine ‚sehr gute bis befriedigende Erinnerungsleistung‘.

Die Erinnerung der Schülerinnen und Schüler ist in Qualität und Umfang als erfreulich zu bewerten.

Abschließend sei nach dem Einfluss differentialpsychologischer Aspekte, hier Alter und Geschlecht, auf die Qualität der Erinnerung gefragt. Die folgende Tabelle stellt die arithmetischen Mittel der Inhaltsanalyse der Subpopulationen aus dem Posttest II dar:

Arithmetische Mittel

Geschlecht	
Jungen	16,3
Mädchen	17,1
Alter	
7 bis 8 Jahre	16,1
9 Jahre	17,2
10 bis 11 Jahre	16,2

Tabelle 3.2.3.1.5: Deskriptive Aspekte; Posttest II;
- differentialpsychologische Einflüsse
Der Verfasser

Die im Posttest I bestehenden differentialpsychologischen Unterschiede in der Qualität der Erinnerung sind im Posttest II nicht mehr existent. Die Mitglieder der einzelnen Subpopulationen erzielen damit ähnliche individuelle Punktschichten. Beim vertiefenden Vergleich der Datenreihen fällt weiterhin eine relativ homogene Abnahme der durchschnittlichen Punktschichten über die einzelnen Subpopulationen der Stichprobe auf. Die einzige Ausnahme bilden die Sieben- bis Achtjährigen, bei denen der Rückgang der Punktschicht deutlich weniger stark ausgeprägt ist als bei den übrigen Gruppen. Bei ihnen hat sich damit die Fähigkeit eine Station zu beschreiben besonders langzeitwirksam entwickelt.

Die Erinnerung an die Stationen ist auch nach dem Zeitraum von sechs Monaten noch umfassend vorhanden. Insofern kann die Analyse der Wirksamkeit der MINIPHÄNOMENTA auf die übrigen Variablen durchgeführt werden.

Die Variable allgemeines Experimentierverhalten

Ich beginne mit dem zeitlichen Verlauf der Wirksamkeit auf der Ebene I. Zu diesem Zweck werden die Häufigkeitsverteilungen des [allgemeinen Experimentierverhaltens] der beiden Posttests in dem folgenden Diagramm vergleichend gegenübergestellt:

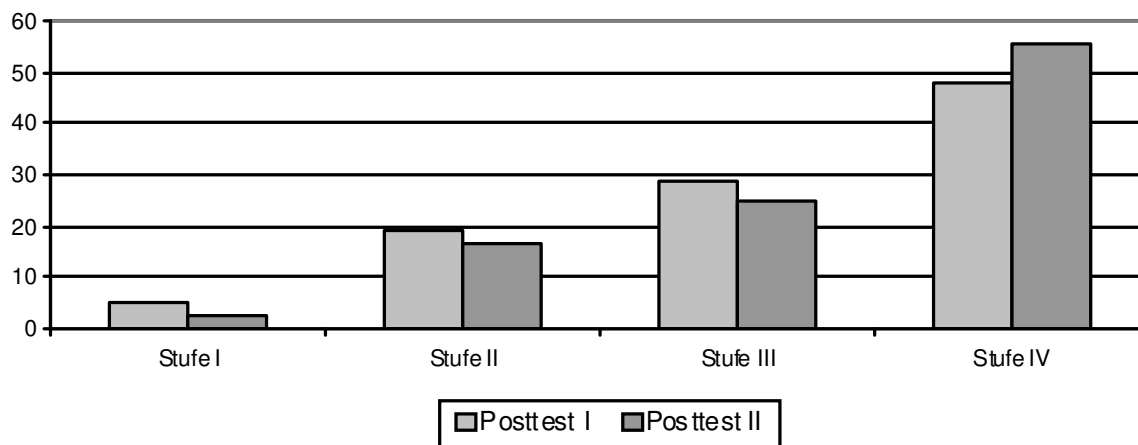


Diagramm 3.2.3.1.2: Allgemeines Experimentierverhalten; Posttest II;
- Häufigkeiten in den Kategorien in den Posttests I und II
Der Verfasser

Beide Verteilungen scheinen einander sehr ähnlich zu sein. Im Posttest II war das allgemeine Experimentierverhalten sogar noch etwas positiver als im Posttest I: Dies zeigt sich vor allem im Anstieg der relativen Häufigkeit der Verhaltenskategorie ‚elaboriertes Experimentierverhalten‘ (Stufe IV). Beide Verteilungen werden durch einen k^*1 -Chi-Quadrat-Test auf ihre Unterschiedlichkeit überprüft:

Größe	Ermittelter Wert
df	3
χ^2	19,7

Tabelle 3.2.3.1.6: Allgemeines Experimentierverhalten; Posttest II;
- Signifikanzüberprüfung der beiden Häufigkeitsverteilungen
Der Verfasser

Es ergibt sich ein hoch signifikanter Unterschied. Wie ist dieses weitere Ansteigen der abhängigen Variablen zu erklären? Hier sind zwei Interpretationen denkbar: Zunächst könnte es sich beim Anstieg im Posttest II um einen Lerneffekt handeln. Die Schülerinnen und Schüler sind durch das wiederholte Erheben dieser Variablen und das damit einhergehende Experimentieren kompetenter geworden. Eine andere Deutungsmöglichkeit ist, unter der Berücksichtigung der Größe des Unterschiedes, wahrscheinlicher: Durch die Teilnahme an der MINIPHÄNOMENTA verändert sich, dies konnte Sauer (2005, vgl. Abschnitt 1.4.4 in dieser Arbeit) nachweisen, der Unterricht maßgeblich. Oft werden verstärkt erfahrungsorientierte und handlungsbezogene Elemente in den schulischen Alltag eingebunden. Diese Veränderung des Unterrichtes, die mit einer stärkeren Konzentration auf das Experimentieren einhergeht, wäre in der Lage den gefundenen Unterschied sinnvoll zu erklären. Es muss an dieser Stelle betont werden, dass zur endgültigen Annahme dieser Interpretation weitere, in dieser Arbeit nicht erhobene Daten nötig wären. Bei dem Gesagten handelt es sich folglich nicht um eine abgesicherte Interpretation sondern vielmehr um einen explorativen Entwurf.

Das [allgemeine Experimentierverhalten] wurde in dem Zeitraum zwischen dem Posttest I und dem Posttest II positiver. Dies ist vermutlich durch eine Veränderung des Unterrichtes durch die MINIPHÄNOMENTA zu erklären.

Der Nachweis der Wirksamkeit auf der Ebene II⁵² kann für die Variable [allgemeines Experimentierverhalten] aufgrund ihres ordinalen Skalenniveaus nicht durch die Berechnung des Nettoeffektes erfolgen. Um einen ersten Überblick für die folgende, statistisch exakte Überprüfung zu gewinnen, stellt die Tabelle die relativen Häufigkeitsverteilungen des Prä- und Posttests II sowie der Experimental- und Kontrollgruppe dar:

Prätest				
	Stufe I	Stufe II	Stufe III	Stufe IV
Experimentalgruppe	1,2	76,3	20,2	0
Kontrollgruppe	10,4	62,0	22,8	4,8
Posttest II				
	Stufe I	Stufe II	Stufe III	Stufe IV
Experimentalgruppe	2,8	16,7	25,0	55,5
Kontrollgruppe	11,8	63,3	22,0	2,9

Tabelle 3.2.3.1.7: Allgemeines Experimentierverhalten; Posttest II;
- relative Häufigkeiten in den unterschiedlichen Gruppen
Der Verfasser

Die Werte legen die Wirksamkeit des treatments nahe. Im Posttest II sind die Schülerinnen und Schüler experimentell deutlich kompetenter. Dieser Unterschied besteht in der Kontrollgruppe nicht. Durch die Verwendung einer Konfigurationsfrequenzanalyse wurden die Unterschiede statistisch exakt überprüft. Sie ergibt in Bezug auf den Unterschied zwischen den Prä- und Posttest II und den zwischen der Experimental- und Kontrollgruppe ein hoch signifikantes Ergebnis. Mit einem Wert von $\epsilon = 1,3$ (vgl. Bortz 1995, S. 150 ff.) ergibt sich eine sehr hohe Effektstärke.

Nach dem Nachweis der grundsätzlichen Wirksamkeit des Projektes sollen nun einige differentialpsychologische Merkmale geklärt werden. Um einen ersten Überblick über die Datentendenzen zu gewinnen, stellt die folgende Tabelle einen Zusammenhang zwischen den Häufigkeitsverteilungen und den genannten Merkmalen her:

⁵² Ich verweise an dieser Stelle zunächst auf den Abschnitt 2.5 in dieser Arbeit. Hier leite ich ausführlich die Ebenen der Wirksamkeit aus der Methode des Experimentes her und erläutere wie die Wirksamkeit der MINIPHÄNOMENTA auf den unterschiedlichen Ebenen gemessen werden kann. Zusammenfassend kann hier gesagt werden, dass die Ebene II der Wirksamkeit im Vergleich zur Ebene I nicht nur auf der Analyse einer Prä-Posttest-Differenz beruht, sondern dass auch Lerneffekte, Reifungseffekte, systematische Wissenszuwächse durch den Einbezug der Ergebnisse einer Kontrollgruppe weitgehend relativiert werden. An dieser Stelle muss betont werden, dass der Begriff der Ebene sich nicht auf den Umfang der Wirksamkeit bezieht. Ist folglich eine Variable auf der Ebene II wirksam, so ist sie damit nicht wirksamer als auf der Ebene I. Vielmehr wurde die Wirksamkeit durch den Einbezug einer Kontrollgruppe genauer gemessen.

Prätest				
Experimentalgruppe	Stufe I	Stufe II	Stufe III	Stufe IV
Männlich	8,6	62,9	25,7	2,8
Weiblich	6,9	65,5	24,1	3,4
Dritte Klasse	12,9	71,0	16,1	0,0
Vierte Klasse	3,0	66,7	24,2	6,1
Kontrollgruppe	Stufe I	Stufe II	Stufe III	Stufe IV
Männlich	9,8	70,4	16,1	3,7
Weiblich	5,6	70,8	19,4	4,2
Dritte Klasse	13,9	63,3	22,8	0,0
Vierte Klasse	2,4	67,1	23,2	7,3
Posttest II				
Experimentalgruppe	Stufe I	Stufe II	Stufe III	Stufe IV
Männlich	2,5	17,5	27,5	52,5
Weiblich	3,1	18,8	21,9	56,5
Dritte Klasse	2,9	17,1	25,7	54,3
Vierte Klasse	2,7	16,2	24,4	56,7
Kontrollgruppe	Stufe I	Stufe II	Stufe III	Stufe IV
Männlich	12,1	60,6	24,2	3,1
Weiblich	11,5	65,8	20,1	2,6
Dritte Klasse	17,2	62,8	20,0	0,0
Vierte Klasse	6,1	63,6	24,2	6,1

Tabelle 3.2.3.1.8: Allgemeines Experimentierverhalten; Posttest II;
- relative Häufigkeiten in den Subpopulationen
Der Verfasser

In Bezug auf die Wirksamkeit der MINIPHÄNOMENTA lassen sich der Tabelle zwei zentrale Ergebnisse entnehmen. Zunächst kann festgestellt werden, dass die Geschlechtszugehörigkeit sowohl in beiden Gruppen als auch zu beiden Messzeitpunkten keine Rolle für das Erreichen einer bestimmten Lerntiefestufe spielt. Ganz anders bei der Zugehörigkeit zu einer bestimmten Klassenstufe: Im Prätest wurde festgestellt, dass die Klassenstufe das Auftreten einer extremen Verhaltenskategorie entscheidend beeinflusst. Schülerinnen und Schüler der dritten Klasse mussten häufiger der Verhaltenskategorie I (,negative Reaktionen') zugeordnet werden. Kinder der Klassenstufe vier wiesen hingegen häufiger ein ,elaboriertes Experimentierverhalten' (Stufe IV) auf. Dieser Unterschied besteht in der Kontrollgruppe konsistent über den gesamten Untersuchungszeitraum. Bei den Messergebnissen der Experimentalgruppe besteht dieser Unterschied nach der Teilnahme an der MINIPHÄNOMENTA nicht mehr. Das Projekt war in der Klassenstufe drei besonders wirksam.

Vor der Teilnahme an der MINIPHÄNOMENTA experimentierten 57% der Schülerinnen und Schüler in der Gruppe. Es bestand ein ausgeprägter Zusammenhang zwischen einem positiven [allgemeinen Experimentierverhalten] und dem Arbeiten in der Gruppe. Im Posttest II hat sich dieses Bild in der Kontrollgruppe kaum verändert. Hier arbeiteten 56% der Kinder mit mindestens einer weiteren Schülerin oder einem weiteren Schüler gemeinsam. Ganz anders im Posttest II in der Experimentalgruppe: Gruppenarbeit war hier der Regelfall.

Insgesamt ist das treatment MINIPHÄNOMENTA in Bezug auf die Variable [allgemeines Experimentierverhalten]

HOCHWIRKSAM (Ebene II)

- Die Berechnung der Effektstärke ergab mit einem Wert von $\varepsilon = 1,3$ einen sehr hohen Effekt.
- Die MINIPHÄNOMENTA wirkte auf alle Subpopulationen der Experimentalgruppe (Umfang der Wirkung = 100%). Durch die Teilnahme am treatment konnten insbesondere die bestehenden Unterschiede zwischen den Klassenstufen nivelliert werden.
- Infolge der Teilnahme am Projekt nahm der Anteil der Gruppenarbeit an den Experimenten deutlich zu.

Die Variable Einstellung gegenüber Phänomenen und ihrer Bearbeitung

Die folgende Tabelle fasst zentrale deskriptive Werte der Posttests I und II zusammen, um den Verlauf der Wirksamkeit über den Untersuchungszeitraum beschreiben zu können:

	Posttest I	Posttest II
Arithmetisches Mittel	49,2	48,1
Konfidenzintervall (95%)	48,1 – 50,2	47,1 – 49,0
Modalwert	56	48
Standardabweichung	8,1	6,8
range	39	33

Tabelle 3.2.3.1.9: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Posttest II;
- deskriptive Werte
Der Verfasser

Der Mittelwertsunterschied zwischen beiden Werten ist mit einem t-Wert von -1,8 und 212 Freiheitsgraden nicht signifikant. Die Einstellungsänderung ist somit ein zeitlich stabiles Merkmal. Der durchschnittlich erreichte kumulierte Rohpunktwert hat zwar leicht abgenommen. Dieser Unterschied ist jedoch statistisch nicht bedeutsam.

Zur vertiefenden Analyse der beiden Verteilungen werden die erreichten kumulierten Punktskategorien zugewiesen, deren cut-off-scores dem Abschnitt 3.2.1.2 zu entnehmen sind. Aufgrund des leicht unterschiedlichen Stichprobenumfanges verwende ich in der folgenden Grafik zur besseren Vergleichbarkeit der Datenreihen relative Werte:

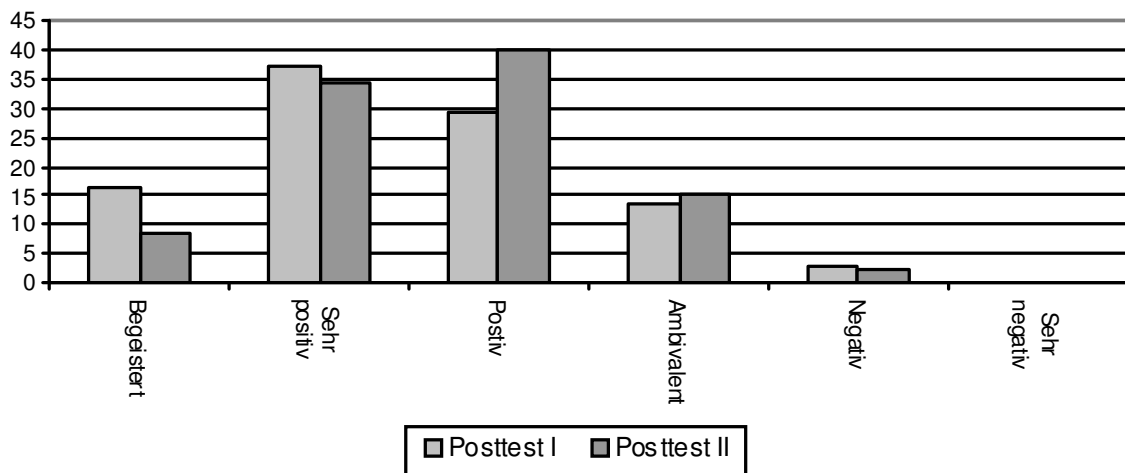


Diagramm 3.2.3.1.3: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Posttest II;
- Einstellungskategorien in Posttest I- und II
Der Verfasser

Beide Verteilungen sind einander ähnlich, auch wenn sich der Schwerpunkt im Posttest II in Bezug auf das obige Diagramm leicht nach rechts verschoben hat. Dies ist vor allem dadurch bedingt, dass der Kategorie ‚begeistert‘ im Posttest II deutlich weniger Versuchspersonen zugeordnet werden konnten. Die Kategorie ‚positiv‘ hingegen ist nun die modale Kategorie. Insgesamt stabilisiert sich die [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] im ‚sehr positiven‘ bis ‚positiven‘ Bereich.

Die [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] stabilisiert sich im Posttest II im ‚sehr positiven‘ bis ‚positiven‘ Bereich.

Der Nachweis der Wirksamkeit der MINIPHÄNOMENTA auf diese Variable unter der Berücksichtigung der beiden Faktoren Zeit und Gruppe geschieht durch die Berechnung des Nettoeffektes, dessen Signifikanz durch den Interaktionseffekt einer zweifaktoriellen Varianzanalyse überprüft wird⁵³:

	AM Prätest	AM Posttest II	Differenz
Experimentalgruppe	43,0	48,1	-5,1
Kontrollgruppe	41,9	40,8	1,1
		Nettoeffekt:	-6,2

Tabelle 3.2.3.1.10: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Posttest II;
- Berechnung des Nettoeffekts
Der Verfasser

Dieser Nettoeffekt ist, dies ergab die Pillai-Spur⁵⁴ mit einem F-Wert von 33,4, hoch signifikant. Die Effektgröße⁵⁵ liegt mit einem Wert von $\eta^2_{\text{part.}} = 0,12$ im mittleren Bereich (vgl. Bortz, Döring 2006, S. 606).

Nachdem die groben Datentendenzen offen gelegt sind, soll nun die Frage beantwortet werden, wie sich die Zeit- und Gruppenunterschiede in den einzelnen Dimensionen der [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] niederschlagen. Die folgende Tabelle listet die arithmetischen Mittel der Dimensionen sowie die Nettoeffekte auf:

⁵³ Ich verweise an dieser Stelle erneut auf den Abschnitt 2.5 in dieser Arbeit, indem die Berechnung der Wirksamkeit auf der Ebene II ausführlich erläutert wird. Zusammenfassend kann hier festgestellt werden, dass ich mich in der Berechnungsvorschrift des Nettoeffektes an Bortz und Döring (1995, S. 521) orientiere. Ein negativer Nettoeffekt kann als ein Indiz für die Wirksamkeit des treatments gewertet werden. Seine Signifikanzüberprüfung erfolgt, wie bereits im obigen Textverlauf erläutert durch den Interaktionseffekt einer zweifaktoriellen Varianzanalyse. In Bezug auf einen positiven Nettoeffekt können zwei Fälle unterschieden werden:

1. Das treatment ist nicht wirksam: Die Mittelwertsdifferenz in der Kontrollgruppe ist größer als in der Experimentalgruppe.
2. Das treatment ist negativ wirksam: Dieser Fall tritt ein, wenn der Mittelwert des Prätests in der Experimentalgruppe größer ist als der Mittelwert des Posttests. Damit hat das treatment einen der gerichteten Forschungshypothese genau entgegenlaufenden Effekt.

Der Nachweis der Wirksamkeit auf der Ebene II beginnt, bei einem metrischen Skalenniveau, mit der Berechnung des Nettoeffektes. Ist dieser negativ, so wird er durch die genannte zweifaktorielle Varianzanalyse auf seine statistische Bedeutsamkeit hin überprüft.

⁵⁴ Die Berechnung erfolgt auf der Basis des allgemeinen linearen Modells (vgl. Abschnitt 2.5 in dieser Arbeit). Die Pillai-Spur ist nach der Angabe von Bühl (2006) der robusteste Test und wird daher im Folgenden für die Signifikanzüberprüfung verwendet.

⁵⁵ Die Effektgröße darf hier nicht mit dem Nettoeffekt verwechselt werden. Sie stellt im Gegensatz zum Nettoeffekt ein standardisiertes Maß dar, das den Vergleich von mehreren Variablen ermöglicht.

Affektiv			
	AM Prätest	AM Posttest II	Differenz
Experimentalgruppe	18,1	20,0	-1,9
Kontrollgruppe	18,2	17,6	0,6
		Nettoeffekt:	-2,5
Kognitiv			
	AM Prätest	AM Posttest II	Differenz
Experimentalgruppe	15,0	16,0	-1,0
Kontrollgruppe	17,6	15,0	2,6
		Nettoeffekt:	-3,6
Verhaltensbezogen			
	AM Prätest	AM Posttest II	Differenz
Experimentalgruppe	9,9	12,1	-2,2
Kontrollgruppe	8,7	8,2	0,5
		Nettoeffekt:	-2,7

Tabelle 3.2.3.1.11: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Posttest II;
- Berechnung der Nettoeffekte der einzelnen Dimensionen
Der Verfasser

Sämtliche Nettoeffekte sind hoch signifikant. Die Schülerinnen und Schüler konnten durch die Teilnahme an der MINIPHÄNOMENTA ihre [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] nachhaltig positiv verändern. Der Vergleich der Dimensionen wird dadurch erschwert, dass sie durch unterschiedlich viele Items operationalisiert werden. Die folgende Grafik zeigt daher die durchschnittlich erreichten relativen kumulierten Rohpunktwerte der Experimentalgruppe:

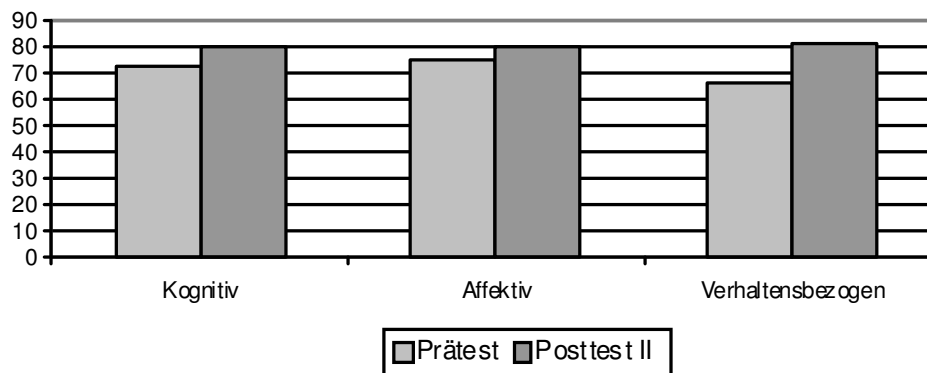


Diagramm 3.2.3.1.4: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Posttest II;
- Komponenten der Einstellung
Der Verfasser

Besonders deutlich ist die Zunahme in der verhaltensbezogenen Dimension. Dieser Umstand ist als überaus erfreulich zu bewerten, zeigt er doch, dass die Schülerinnen und Schüler das Experimentieren nach der Teilnahme an der MINIPHÄNOMENTA als eine aktive und planvolle Tätigkeit sehen.

Abschließend soll der Einfluss differentialpsychologischer Faktoren auf die Wirksamkeit des treatments untersucht werden. Es wird also der Frage nachgegangen, ob die MINIPHÄNOMENTA in Bezug auf Jungen und Mädchen und unterschiedliche Altersgruppen gleich wirksam war, oder ob sich ihre Wirksamkeit auf bestimmte Stichprobensubpopulationen beschränkte. Die folgende Tabelle stellt die Nettoeffekte für Jungen und Mädchen sowie für unterschiedliche Altersgruppen getrennt dar:

Merkmal	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Männlich	-6,4	27,1	0,00
Weiblich	-5,2	4,4	0,03
7 bis 8 Jahre	-0,9	3,3	0,07
9 Jahre	-3,5	1,6	0,19
10 bis 11 Jahre	-7,3	18,6	0,00

Tabelle 3.2.3.1.12: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Posttest II;
- Nettoeffekte der Subpopulationen
Der Verfasser

Die MINIPHÄNOMENTA ist sowohl auf die [Einstellung] der Jungen als auch auf die der Mädchen wirksam. Beide Gruppen wiesen nach der Teilnahme an der MINIPHÄNOMENTA eine statistisch bedeutsam positivere [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] auf. Dabei ist der Nettoeffekt bei den Jungen jedoch höher, so dass ihre Einstellung durch die Teilnahme an der MINIPHÄNOMENTA umfassender gefördert wurde.

Betrachtet man die Wirksamkeit der MINIPHÄNOMENTA auf die unterschiedlichen Altersgruppen so ergibt sich ein heterogenes Bild. Zunächst zur Gruppe der zehn- bis elfjährigen: Der ermittelte Nettoeffekt ist hoch signifikant. Die MINIPHÄNOMENTA förderte folglich die [Einstellung gegenüber Phänomenen und ihrer Bearbeitung] dieser Altersgruppe. Dieser Umstand ist als erfreulich zu bewerten, da, dies zeigen die Ergebnisse des Prätests, die Einstellung dieser Subpopulation vor der Teilnahme an der MINIPHÄNOMENTA sehr negativ war. In Bezug auf die beiden anderen Altersklassen liegt kein signifikanter Nettoeffekt vor. Damit kann festgestellt werden, dass das treatment der MINIPHÄNOMENTA in Bezug auf diese Altersgruppen nicht wirksam war. Dieses Ergebnis wird dadurch relativiert, dass die [Einstellung] in diesen Altersgruppen, dies zeigen die Ergebnisse des Prätests (vgl. Abschnitt 3.2.1.2 in dieser Arbeit) insgesamt im oberen Bereich liegen.

Insgesamt ist das treatment MINIPHÄNOMENTA in Bezug auf die [Einstellung gegenüber Phänomenen und ihrer Bearbeitung]

WIRKSAM (Ebene II)

- Es liegt eine mittlere Effektgröße von $\eta^2_{\text{part.}} = 0,12$ vor.
- Die [Einstellung] der Schülerinnen und Schüler wird in allen Dimensionen gefördert. Besonders deutlich ist die Wirkung in der ‚verhaltensbezogenen Dimension‘.
- Die differentialpsychologischen Merkmale moderieren entscheidend die Wirksamkeit der MINIPHÄNOMENTA. Hochwirksam ist sie bei den Jungen und den zehn- bis elfjährigen und wirksam bei den Mädchen. Bei den Altersgruppen der sieben- bis acht- und der neunjährigen konnte hingegen keine Wirksamkeit festgestellt werden. Der umfang der Wirkung beträgt 60%.

Die Variable formale Kompetenz

Zunächst zum Vergleich der beiden Posttestwerte. Die folgende Grafik stellt die arithmetischen Mittel der Experimentalgruppe zu beiden Messzeitpunkten gegenüber:

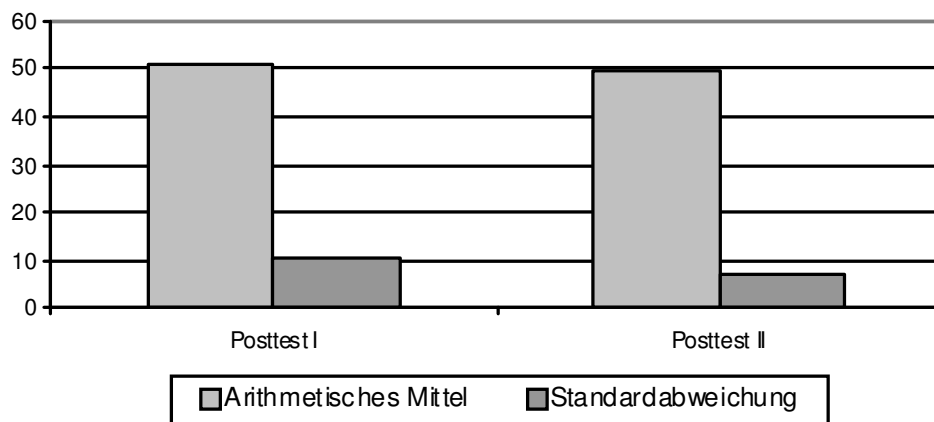


Diagramm 3.2.3.1.5: Formale Kompetenz; Posttest II;
- deskriptive Werte des Posttest I und II
Der Verfasser

In der ersten orientierenden Ansicht kann festgestellt werden, dass das arithmetische Mittel der [formalen Kompetenz] mit zunehmendem zeitlichem Abstand zum treatment sinkt. Auch bei der Standardabweichung zeigt sich eine Abnahme. Damit sinkt die Wahrscheinlichkeit für extreme Werte. Zur Überprüfung der statistischen Bedeutsamkeit dieses Unterschiedes wird ein t-Test für unabhängige Stichproben verwendet, dessen zentrale Werte der folgenden Tabelle zu entnehmen sind:

Größe	ermittelter Wert
df	41
t-Wert	0,47

Tabelle 3.2.3.1.13: Formale Kompetenz; Posttest II;
- Signifikanzüberprüfung der Mittelwerte der Posttests I und II
Der Verfasser

Die Mittelwertsdifferenz ist folglich nicht signifikant. Damit können beide Verteilungen als ähnlich klassifiziert werden, so dass festgestellt werden kann, dass es sich bei der [formalen Kompetenz] um ein zeitlich relativ stabiles Merkmal handelt.

In der folgenden Tabelle sind die Nettoeffekte der drei Basisvariablen der [formalen Kompetenz] (‚Bearbeitungszeit‘, ‚Gruppenklima beim Experimentieren‘ und ‚Konzentration auf den Versuch‘) sowie deren Signifikanzüberprüfung dargestellt:

Merkmal	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Bearbeitungszeit	-2,7	58,6	0,00
Gruppenklima	-0,8	22,6	0,00
Konzentration	-0,7	21,3	0,00

Tabelle 3.2.3.1.14: Formale Kompetenz; Posttest II;
- Nettoeffekte der Basisvariablen (Prätest-Posttest II)
Der Verfasser

Die Ergebnisse zeigen eindrucksvoll den großen Einfluss des treatments. Die MINIPHÄNOMENTA veränderte alle drei Basisvariablen und damit die Rahmendaten des Experimentierens entscheidend: Die ‚Zeit der Bearbeitung‘ stieg, die ‚Konzentration auf die Aufgabe‘ wurde verbessert und das ‚Gruppenklima‘ wurde positiver.

Noch beeindruckender ist der Einfluss des treatments auf die eigentliche [formale Kompetenz]. Zeigten sich die Schülerinnen und Schüler im Prätest experimentell unerfahren, besaßen folglich in weiten Bereichen keine altersangemessenen Konzepte über den Prozess des Experimentierens und die Reflektion über eben diesen, so konnte dieser Bereich durch die Teilnahme an der MINIPHÄNOMENTA entscheidende gefördert werden. Der Nettoeffekt von -13,5 muss als extrem hoch bewertet werden. Die Pillai-Spur ergab mit einem F-Wert von 42,4 ein hoch signifikantes Ergebnis. Auf dieser Basis wurde die Effektgröße ermittelt. Sie ist mit einem Wert von $\eta^2_{\text{part.}} = 0,32$ als hoch zu klassifizieren (vgl. Bortz, Döring 2006, S. 606).

Bei der [formalen Kompetenz] handelt es sich um ein heterogenes Merkmal, dass in dieser Arbeit durch drei Dimensionen operationalisiert wurde. Zur vertiefenden Analyse soll der Einfluss des treatments auf die genannten Dimensionen untersucht werden. Die folgende Tabelle listet die Nettoeffekte sowie die Ergebnisse ihrer Signifikanzüberprüfung auf:

Dimension	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Logische Kompetenzen	-5,6	50,3	0,00
Experimentelle Kompetenzen	-4,6	19,3	0,00
Spezifisch naturwissenschaftliche Denkweisen	-4,5	83,2	0,00

Tabelle 3.2.3.1.15: Formale Kompetenz; Posttest II;
- Nettoeffekte der Dimensionen
Der Verfasser

Die Teilnahme an der MINIPHÄNOMENTA hatte damit Einfluss auf alle drei Dimensionen und förderte die Schülerinnen und Schüler in Bezug auf die [formale Kompetenz] umfassend. Die einzelnen Dimensionen wurden durch jeweils unterschiedlich viele Kategorien operationalisiert. Um zu einer genaueren, vergleichenden Aussage zu gelangen, werden in der folgenden Grafik die durchschnittlichen relativen Punktsummen des Prätestes und des Posttests II gegenübergestellt:

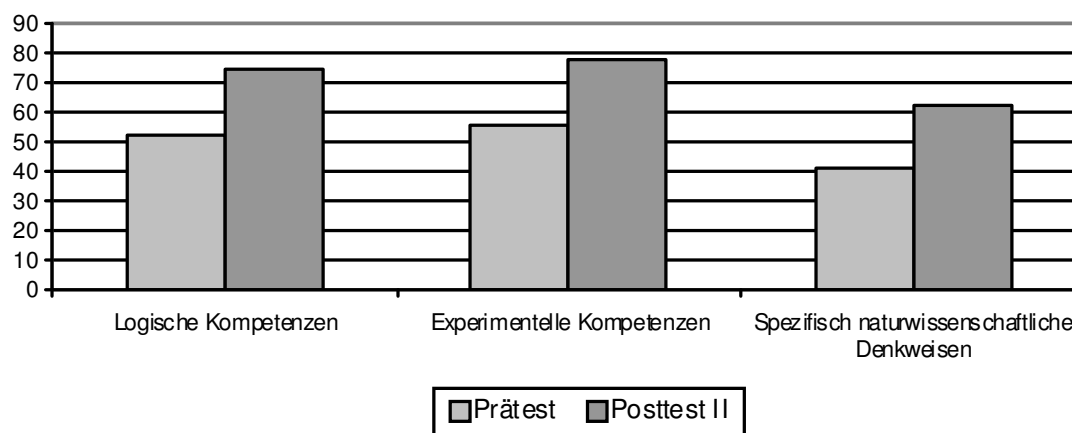


Diagramm 3.2.3.1.6: Formale Kompetenz; Posttest II;
- durchschnittliche relative Punktsummen der Dimensionen der formalen Kompetenz
Der Verfasser

Die grundsätzliche Struktur der Daten aus dem Prätest in Bezug auf die Experimentalgruppe ist auch nach der Teilnahme an der MINIPHÄNOMENTA erhalten geblieben. Bei den ‚experimentellen Kompetenzen‘ erreichten die Schülerinnen und die Schüler den höchsten Punktwert. Nur knapp darunter liegen die ‚logischen Kompetenzen‘. Die ‚spezifisch naturwissenschaftlichen Denkweisen‘, wie zum Beispiel das gezielte Manipulieren der Variablen oder erste Einblicke in das Prinzip der Kausalität, fallen den Schülerinnen und Schülern deutlich schwerer. Die MINIPHÄNOMENTA war auf alle Dimensionen der [formalen Kompetenz] wirksam: Die erzielten Werte liegen in allen Dimensionen deutlich über dem Prätestniveau.

Abschließend sei auf differentialpsychologische Fragestellungen eingegangen. Die Ergebnisse der Analyse sind in der folgenden Tabelle zusammengefasst:

Merkmal	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Jungen	-21,5	40,1	0,00
Mädchen	-7,9	13,8	0,00
Dritte Klasse	-13,2	19,6	0,00
Vierte Klasse	-14,5	34,5	0,00

Tabelle 3.2.3.1.16: Formale Kompetenz; Posttest II;
- Überprüfung differentialpsychologischer Fragestellungen
Der Verfasser

Diese Daten unterstreichen den in Bezug auf diese Variable bereits festgestellten umfassenden Effekt von ASIP. Das treatment war auf alle Subpopulationen wirksam.

Insgesamt ist das treatment MINIPHÄNOMENTA im Hinblick auf die Variable [formale Kompetenz]

HOCHWIRKSAM (Ebene II)

- Mit einem Wert von $\eta^2_{part.} = 0,32$ wurde eine sehr hohe Effektgröße ermittelt.
- Das treatment fördert gleichmäßig alle definierten Dimensionen der [formalen Kompetenz].
- Die Teilnahme an der MINIPHÄNOMENTA führt bei allen Subpopulationen der Stichprobe zu einem weit reichenden Ausbau der [formalen Kompetenz]. Der Umfang der Wirkung beträgt damit 100%.

Die Variable aktuelle Motivation

Im Posttest I konnte die Wirksamkeit des treatments auf die Variable nachgewiesen werden. Hier ist zunächst nach dem zeitlichen Verlauf eben dieser Wirksamkeit zu fragen. Die folgende Tabelle stellt daher die zentralen deskriptiven Werte der beiden Posttestmessungen gegenüber:

	Posttest I	Posttest II
Arithmetisches Mittel	53,5	50,2
Konfidenzintervall (95%)	52,4 – 54,7	48,5 – 51,9
Modalwert	52	49
Standardabweichung	6,1	8,7
range	28	43

Tabelle 3.2.3.1.17: Aktuelle Motivation; Posttest II;
- deskriptive Kennwerte der Posttests I und II
Der Verfasser

Der Unterschied ist mit einem t-Wert von -3,17 und 107 Freiheitsgraden signifikant. Folglich sinkt die durch das treatment verursachte Steigerung der [Motivation] statistisch bedeutsam ab, das heißt die [Motivation] der Schülerinnen und Schüler wird leicht negativer. Dieses Absinken der Ergebnisse der [Motivation] vom Posttest I zum Posttest II soll in der folgenden Grafik durch die Zuweisung der relativen individuell erreichten Rohpunktsummen zu Kategorien verdeutlicht werden:

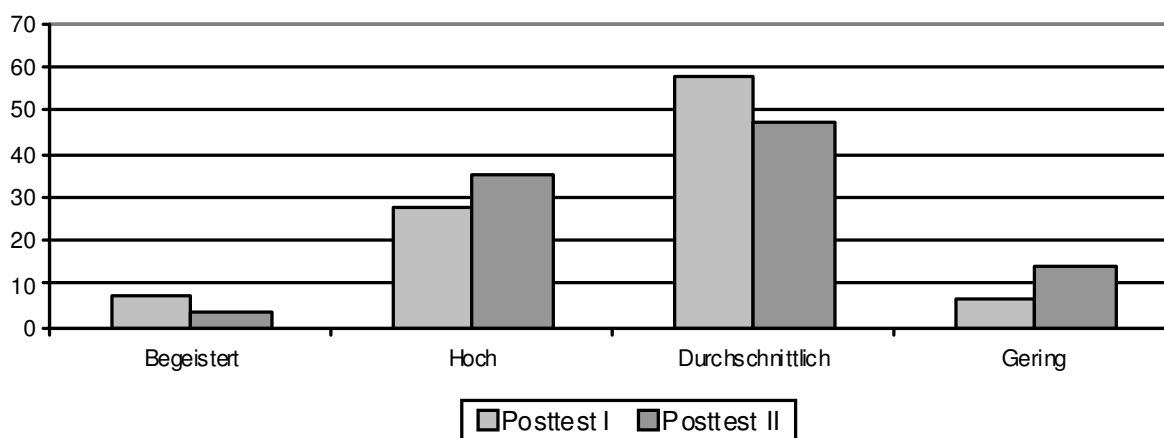


Diagramme 3.2.3.1.7: Aktuelle Motivation; Posttest II;
- relative Kategorienhäufigkeit in Posttest I und II
Der Verfasser

Der Anteil der Schülerinnen und Schüler mit einer ‚begeisterten und durchschnittlichen Motivation‘ ist gesunken. Die Anzahl der Kinder mit einer ‚geringen Motivation‘ und - überraschenderweise - auch diejenige mit einer ‚hohen Motivation‘ ist gestiegen. Die Daten legen die hier nicht weiter zu überprüfende These nahe, dass ein Teil der Schülerinnen und Schüler mit ‚begeisterter Motivation‘ im Posttest I nun nur noch eine ‚hohe Motivation‘ aufweisen. Dies gilt analog für die Kategorien ‚durchschnittlich‘ und ‚gering‘.

Die [Motivation] sinkt zwischen dem Posttest I und II signifikant ab.

Mit dem Nachweis des zeitlichen Verlaufes ist jedoch keineswegs die Wirksamkeit der MINIPHÄNOMENTA in Frage gestellt, denn sie kann immer nur in Relation zum Prätest ermittelt werden. Ich beginne mit der Analyse des Einflusses der MINIPHÄNOMENTA auf die gesamte Variable [aktuelle Motivation]. Die folgende Tabelle stellt zentrale Werte zusammen:

Merkmal	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Motivation	-4,2	6,4	0,01

Tabelle 3.2.3.1.18: Aktuelle Motivation; Posttest II;
- Nettoeffekt der Motivation
Der Verfasser

Die Tabelle zeigt einen negativen signifikanten Nettoeffekt, der die Wirksamkeit des treatments der MINIPHÄNOMENTA nachweist. Zusätzlich wurde die Effektgröße bestimmt. Sie liegt mit einem Wert von $\eta^2_{\text{part.}} = 0,10$ im mittleren Bereich (Bortz, Döring 2006, S. 606). Nach der Teilnahme am Projekt wiesen die Schülerinnen und Schüler damit eine leicht höhere [Motivation] auf, sich mit physikalischen Fragestellungen zu beschäftigen.

Wie bereits bei den vorherigen Variablen beschrieben ist auch die [Motivation] als komplexes Merkmal, das heißt auf mehreren Dimensionen operationalisiert. In der folgenden Tabelle wird das Ergebnis der Überprüfung des Einflusses der MINIPHÄNOMENTA auf die einzelnen Dimensionen der Motivation angeführt:

Dimension	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Kompetenzerwartungen	4,6	54,8	0,00
Misserfolgsbefürchtungen	-13,9	208,3	0,00
Interesse	0,4	1,0	0,32

Tabelle 3.2.3.1.18: Aktuelle Motivation; Posttest II;
- Überprüfung der Dimensionen
Der Verfasser

Die Daten zeigen ein erstaunlich heterogenes Bild. Die MINIPHÄNOMENTA beeinflusst die ‚Misserfolgsbefürchtungen‘ der Schülerinnen und Schüler entscheidend. Nach der Teilnahme am Projekt äußern sie deutlich weniger ‚Misserfolgsbefürchtungen‘. Die subjektiven ‚Kompetenzerwartungen‘ sind hingegen durch das treatment signifikant gefallen. Dies zeigt der positive und hoch signifikante Nettoeffekt. Das ‚Interesse‘ an der Fragestellung konnte durch das treatment nicht verändert werden. Jene Interpretation wird durch den leicht positiven aber nicht signifikanten Nettoeffekt evident.

Wie kann diese Datenlage interpretiert werden? Das ‚Interesse‘ der Schülerinnen und Schüler eine derartige Fragestellung zu bearbeiten ist, dies zeigen die Prätestergebnisse, sehr hoch und wird durch die MINIPHÄNOMENTA nicht weiter gefördert. Die geringeren ‚Kompetenzerwartungen‘ lassen sich durch die offene Gestaltung der Experimentierstationen erklären. Durch die Abwesenheit externer Vorgaben, das völlig selbstgesteuerte Vorgehen der Schülerinnen und Schüler an den Stationen, kommt es zwangsläufig, gerade wenn man die geringen Erfahrungen der Schülerinnen und Schüler mit diesem Lernbereich bedenkt, zu Misserfolgen, die das subjektive ‚Kompetenzgefühl‘ schmälern. Diese Misserfolge werden aber von den Kindern nicht negativ gedeutet: Durch die Beschaffenheit des interaktiven Lernprozesses resultieren aus den Misserfolgen keine negative Konsequenzen wie es im setting Schule sonst allzu oft üblich ist. Insofern kann das signifikante Absinken der subjektiven ‚Kompetenzerwartung‘ als die wachsende Fähigkeit der Schülerinnen und Schüler beschrieben werden, ihre eigenen Fähigkeiten realistisch einzuschätzen.

Abschließend soll der Frage nach dem Einfluss der Geschlechteszugehörigkeit auf die [aktuelle Motivation] nachgegangen werden. Es ergibt sich erneut ein heterogenes Bild. Die [Motivation] der Mädchen wurde durch das treatment umfassend gefördert. Der Nettoeffekt von -11,2 ist mit einem F-Wert von 24,4 hoch signifikant. Durch die Teilnahme an der MINIPHÄNOMENTA sank die [Motivation] der Jungen mit einem Nettoeffekt von 4,5 und einem F-Wert von 6,7 signifikant ab. Relativierend muss hier eingewandt werden, dass der signifikante Nettoeffekt durch einen Zuwachs der [aktuellen Motivation] in der Kontrollgruppe und ein leichtes Absinken in der Experimentalgruppe begründet werden kann. Es ist folglich hier problematisch von einer negativen Wirksamkeit der MINIPHÄNOMENTA zu sprechen. Vielmehr legt die genauere Betrachtung der Daten den Schluss nahe, dass die MINIPHÄNOMENTA auf diese Gruppe nicht wirksam war.

Die MINIPHÄNOMENTA war in Bezug auf die Variable [aktuelle Motivation]

TEILWEISE WIRKSAM (Ebene II)

- Mittlerer Nettoeffekt der Gesamtvariablen [aktuelle Motivation] von $\eta^2_{\text{part.}} = 0,10$
- Das treatment ist nur auf der Dimension ‚Misserfolgsbefürchtung‘ hochwirksam. In Bezug auf das ‚Interesse‘ ist ASIP unwirksam, die subjektiven ‚Kompetenzerwartungen‘ sanken infolge der Teilnahme an der MINIPHÄNOMENTA.
- Das Projekt war nur in Bezug auf die [Motivation] der Mädchen wirksam. Der Umfang der Wirkung beträgt somit 50%.

Die Variable Wissen

Die folgende Tabelle stellt die arithmetischen Mittel und die Standardabweichungen der beiden Indizes aus dem Posttest I und II vergleichend gegenüber, um einen ersten Überblick über den zeitlichen Verlauf dieser Variablen zu erhalten:

	Arithmetisches Mittel	Standardabweichung
Posttest I		
Index I (,Qualität')	4,1	1,4
Index II (,Quantität')	5,2	0,3
Posttest II		
Index I (,Qualität')	4,2	1,8
Index II (,Quantität')	5,1	0,7

Tabelle 3.2.3.1.19: Wissen; Posttest II;
- deskriptive Kennwerte des Posttests I und II
Der Verfasser

Die Werte sind einander derartig ähnlich, dass eine Signifikanzüberprüfung für das Vorliegen von Mittelwertsdifferenzen unterbleiben kann. Die arithmetischen Mittel beider Indizes (,Qualität' und ,Umfang') zeigen deutlich, dass es sich bei der Variablen [Wissen] um ein überaus stabiles Merkmal handelt, das in der Experimentalgruppe nach der Teilnahme an der MINIPHÄNOMENTA konstant bleibt.

Betrachten wir nun die Nettoeffekte beider Indizes. In der unten stehenden Tabelle werden sie, zusammen mit den zentralen Daten ihrer Signifikanzüberprüfung aufgelistet:

Index	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Index I (,Qualität')	-2,2	85,8	0,00
Index II (,Quantität')	-0,9	9,1	0,00

Tabelle 3.2.3.1.20: Wissen; Posttest II;
- Nettoeffekte der beiden Indizes
Der Verfasser

Beide Nettoeffekte, wenn auch unterschiedlich stark ausgeprägt, sind statistisch hoch bedeutsam. Für den Index I (,Qualität') konnte mit einem Wert von $\eta^2_{\text{part.}} = 0,36$ eine extrem hohe Effektgröße ermittelt werden. Nach der Teilnahme an der MINIPHÄNOMENTA waren die Schülerinnen und Schüler statistisch hoch bedeutsam besser in der Lage einen Versuch zu beobachten und auf der Basis der empirischen Ergebnisse eigene Denkbemühungen über den Versuch anzustellen. Die Effektgröße des Index II (,Quantität') ist deutlich kleiner. Sie liegt mit einem Wert von $\eta^2_{\text{part.}} = 0,13$ im mittleren Bereich. Die Kinder verwendeten in Folge des treatments mehr Karten und bilden mehr Inseln. Zusammenfassend lässt sich damit feststellen, dass das [Wissen] der Schülerinnen und Schüler durch die Teilnahme am Projekt entscheidend gefördert wurde.

Abschließend sei nach dem Einfluss differentialpsychologischer Faktoren, hier die Klassenstufe und die Geschlechtszugehörigkeit, auf das [Wissen] der Kinder gefragt. Die folgenden Tabellen geben einen orientierenden Überblick:

Merkmal	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Jungen	-4,8	91,2	0,00
Mädchen	-3,3	36,7	0,00
Dritte Klasse	-2,2	45,4	0,00
Vierte Klasse	-2,1	36,1	0,00

Tabelle 3.2.3.1.21: Wissen; Posttest II;
- Überprüfung differentialpsychologischer Fragestellungen für den Index I („Qualität“)
Der Verfasser

Merkmal	Nettoeffekt	F-Wert der Pillai-Spur	Signifikanzniveau
Jungen	-1,4	13,5	0,00
Mädchen	-0,2	0,4	0,49
Dritte Klasse	-1,1	9,1	0,03
Vierte Klasse	-0,5	1,4	0,23

Tabelle 3.2.3.1.21: Wissen; Posttest II;
- Überprüfung differentialpsychologischer Fragestellungen für den Index II („Quantität“)
Der Verfasser

Durch die Teilnahme am treatment stieg der Index I („Qualität“) in allen Subgruppen der Stichprobe hoch signifikant. Die MINIPHÄNOMENTA war folglich wirksam.

In Bezug auf den Index II („Quantität“) ergibt ein heterogenes Bild: Die Jungen verwendeten nach der Teilnahme an ASIP statistisch hoch bedeutsam mehr Karten und bildeten mehr Inseln, ebenso die Schülerinnen und Schülern der Klassenstufe drei. Bei den Mädchen und den Kindern der vierten Klassenstufe stieg der Wert in der Experimentalgruppe im Posttest II in Relation zu der Kontrollgruppe. Dieser Unterschied ist jedoch in beiden Fällen nicht signifikant. Das treatment förderte damit selektiv diejenigen Gruppen, bei denen im Prätest niedrige Werte für diesen Index vorlagen.

Die MINIPHÄNOMENTA war in Bezug auf die Variable [Wissen]

HOCHWIRKSAM (Ebene II)

- Hohe Effektgröße der Gesamtvariablen [Wissen] mit einem mittleren partiellem η^2 von 0,245.
- Umfassende Förderung der durch den Index I beschriebenen ‚Qualität‘ der Beantwortung der Aufgabe. In allen Subpopulationen lassen sich statistisch hoch bedeutsame Nettoeffekte nachweisen. Der Umfang der Wirkung beträgt auf diesem Index 100%.
- Die ‚Quantität‘ der Bearbeitung der Aufgabe wurde nicht in allen Elementen der Stichprobe signifikant gefördert. Der Umfang der Wirkung beträgt hier lediglich 50%.

3.2.3.2 Zusammenfassung der Ergebnisse des Posttest II

Die folgende Tabelle stellt die zentralen Ergebnisse des Posttest II zusammen:

Variable	Wirksamkeitsstufe	Zusammenfassung der Ergebnisse
Deskriptive Aspekte	Erinnerung vorhanden: Weitere Analyse sinnvoll	<p>Die Erinnerung an die Stationen der MINIPHÄNOMENTA ist ein stabiles Merkmal: Auch im Posttest II erinnerten sich die Schülerinnen und Schüler an den größten Teil der Versuche.</p> <p>Es gelang ihnen erneut einen Versuch umfassend zu beschreiben. Im Durchschnitt der Gruppe wurden Aufbau, Funktion und Effekt der Station deutlich.</p> <p>Die Erinnerung an die Stationen ist affektiv gefärbt und wird in der Form konkreter Vorstellungsbilder gespeichert.</p>
Allgemeines Experimentierverhalten	HOCHWIRKSAM	<p>Im Posttest II wurde ein elaborierteres allgemeines Experimentierverhalten gemessen als im Posttest I. Eine mögliche Interpretation bestünde in der Veränderung des Unterrichtes durch die MINIPHÄNOMENTA.</p> <p>Durch die Verwendung einer Konfigurationsfrequenzanalyse konnte ein signifikanter Zuwachs vom Prätest zum Posttest II bestätigt werden.</p> <p>ASIP war auf alle Stichprobensubpopulation wirksam, insbesondere jedoch auf das allgemeine Experimentierverhalten der Schülerinnen und Schüler der Klassenstufe 3.</p> <p>Das treatment förderte entscheidend die Gruppenarbeit an den Versuchen.</p>
Einstellung	WIRKSAM	<p>Die Einstellung gegenüber Phänomenen und ihrer Bearbeitung ist über den Zeitraum der Untersuchungsstufe I ein stabiles Merkmal.</p> <p>Es konnte eine mittlere Effektgröße ermittelt werden.</p> <p>Das treatment war in allen Dimension wirksam. Besonders deutlich zeigten sich die positiven Effekte in der verhaltensbezogenen Dimension.</p> <p>In Bezug auf die differentialpsychologischen Merkmale ergibt sich ein heterogenes Bild. Signifikante beziehungsweise hoch signifikante Nettoeffekte gibt es in Bezug auf die Einstellung der Jungen, der Mädchen und der zehn- bis elfjährigen Schülerinnen und Schüler. Die Einstellung der beiden anderen Altersgruppen wurde durch die MINIPHÄNOMENTA nicht gefördert.</p>
Formale Kompetenz	HOCHWIRKSAM	<p>Es konnte eine hohe Effektgröße ermittelt werden.</p> <p>Die MINIPHÄNOMENTA war sowohl auf die Rahmenaspekte des Experimentierens als auch auf die Dimensionen der formalen Kompetenz wirksam.</p> <p>Die Wirksamkeit besteht in Bezug auf alle Subpopulationen der Stichprobe.</p>

Aktuelle Motivation	TEILWEISE WIRKSAM	<p>Im Posttest II sank die Motivation in Relation zum Posttest I signifikant ab. Die Wirksamkeit des treatments hat folglich abgenommen.</p> <p>Es wurde eine mittlere Effektgröße ermittelt.</p> <p>In Bezug auf die Ergebnisse der einzelnen Dimensionen ergibt sich ein heterogenes Bild. Während die Misserfolgsbefürchtungen in Folge der MINIPHÄNOMENTA abnahmen, sank die Erwartung an die eigene Kompetenz. Auf das Interesse an physikalischen Fragestellungen hatte die MINIPHÄNOMENTA keinen Einfluss</p> <p>Weiterhin war die Teilnahme an ASIP nur für die Mädchen mit einer Steigerung der Motivation verbunden.</p>
Wissen	HOCHWIRKSAM	<p>Das Wissen ist ein stabiles Merkmal. Die Ergebnisse des Posttest I und II ähneln sich stark.</p> <p>Differentialpsychologische Merkmale haben auf den Index I keine Wirkung. Ganz anders beim Index II: Nur die Jungen und die Schülerinnen und Schüler der dritten Klasse verwendeten im Posttest mehr Karten und Inseln.</p>

Insgesamt war damit das treatment **WIRKSAM**.

Tabelle 3.2.3.2.1: Zusammenfassung; Posttest II;
- Zusammenfassung der Ergebnisse des Posttest II
Der Verfasser

3.2.4 Zusammenfassende Bewertung der Ergebnisse der Untersuchungsstufe I

Im Rahmen dieser Untersuchungsstufe ist es gelungen die Wirksamkeit des treatments über einen Zeitraum von sechs Monaten auf zwei Ebenen differenziert zu beschreiben. Dabei konnte gezeigt werden, dass durch die intensive Beschäftigung mit den Exponaten eine hohe Wirksamkeit erzielt wird: Die Schülerinnen und Schüler erinnerten sich differenziert an die Versuche, verbesserten ihre [Einstellung] gegenüber physikalischen Aufgabenstellungen und lernten vor allem das Experimentieren als einen aktiven und spannungsgeladen Prozess kennen. Waren die Kinder im Vorfeld der MINIPHÄNOMENTA experimentell extrem unerfahren, so bauten sie im Zuge der Beschäftigung mit den Stationen eine beachtliche [formale Kompetenz] auf, die es ihnen ermöglichen gezielt zu experimentieren und auf der Basis der empirischen Ergebnisse zu eigenen subjektiv tragfähigen Erklärungen zu kommen.

Abschließend sollen hier die Ergebnisse vor dem Hintergrund der im Abschnitt 1.6 in dieser Arbeit eingeführten Hypothesen zusammenfassend diskutiert werden:

Hypothese:

Die Ergebnisse der Kontroll- und Experimentalgruppe unterscheiden sich im Prätest nicht.

Die Hypothese kann insgesamt angenommen werden. Das niedrige Niveau des Prätests war sowohl in der Experimental- als auch in der Kontrollgruppe gegeben. Dieses Ergebnis deutet darauf hin, dass die Physikalische Bildung im Primarbereich insgesamt zu wenig gefördert wird.

Hypothese:

In der Experimentalgruppe zeigt sich ein signifikanter Unterschied zwischen Prä- und Posttest I.

In Bezug auf alle Variablen konnte im Posttest I ein Einfluss der MINIPHÄNOMENTA nachgewiesen werden.

Hypothese:

Dieser Unterschied bleibt bis zum Posttest II erhalten.

Diese Hypothese kann nicht vollständig angenommen werden. In Bezug auf die Wirksamkeitsebene II müssen im Bereich einiger Subdimensionen und Gruppen der Stichprobe Einschränkungen gemacht werden. Dies gilt insbesondere für die Variable [Motivation].

Hypothese:

Die Wirksamkeit der MINIPHÄNOMENTA zeigt sich in allen sechs erfassten Variablen.

Diese Hypothese kann angenommen werden. Die MINIPHÄNOMENTA hat, wenn auch unterschiedlich ausgeprägt, Einfluss auf alle erfassten abhängigen Variablen.

Hypothese:

Es ist anzunehmen, dass die MINIPHÄNOMENTA auf einige Variablen besonders großen Einfluss haben wird.

Besonders wirksam ist die MINIPHÄNOMENTA, dies wurde bereits in der Einleitung dieses Abschnittes ausgeführt, auf die [formale Kompetenz] und das [Wissen] der Schülerinnen und Schüler, insbesondere auf dessen ‚Qualität‘. Durch die Beschäftigung mit den Versuchen lernen sie gezielt zu experimentieren und kontrolliert über einen Versuch nachzudenken.

Hypothese:

Die Wirksamkeit von ASIP hängt nicht vom Geschlecht der Schülerinnen und Schüler und nicht von der besuchten Klassenstufe ab. Damit ist die MINIPHÄNOMENTA in allen Subgruppen der Stichprobe wirksam.

Diese Hypothese kann nur teilweise angenommen werden. Es gelten folgende Einschränkungen:

- [Einstellung]: Keine Förderung der [Einstellung] der jüngeren Schülerinnen und Schüler
- [Motivation]: Keine Erhöhung der [Motivation] der Jungen infolge der Teilnahme an der MINIPHÄNOMENTA.
- [Wissen]: Die ‚Quantität‘ der Bearbeitung der Aufgabe stieg nach der Teilnahme am Projekt nicht bei den Schülerinnen und Schülern der vierten Klasse und auch nicht bei den Mädchen.

3.3 Ergebnisse der Untersuchungsstufe II

3.3.1 Die Wirksamkeit von ASIP zweieinviertel Jahre nach der Teilnahme

Ich verweise an dieser Stelle zunächst auf den Abschnitt 2.2 in dieser Arbeit, indem der explorative Charakter dieser Untersuchungsstufe herausgearbeitet wurde. Der nun folgende Abschnitt soll im Zusammenhang dieser Arbeit lediglich einen Ausblick bieten. Dem Charakter dieses Abschnittes entsprechend, wurde ein einfaches unvollständiges Experiment (ohne Kontrollgruppe) gewählt, dessen Validität zusätzlich durch eine relativ geringe Versuchspersonenzahl eingeschränkt wird.

Welche Konsequenzen hat die beschriebene Beschaffenheit nun für die vorzunehmende Auswertung? Aus der Sicht des Verfassers können hier lediglich die groben deskriptiven Tendenzen des Datenmaterials, vergleichend zu denen des Prätests der Untersuchungsstufe I, analysiert werden. Einzelaspekte, zum Beispiel differentialpsychologische Fragestellungen, sollen aufgrund der methodischen Beschaffenheit der Untersuchungsstufe nicht ausgewertet werden. Die Ergebnisse würden eine zu geringe Validität aufweisen.

Die Variable deskriptive Aspekte

Ist die Voraussetzung zur Analyse von Wirksamkeit, das heißt die Erinnerung an die Stationen, auch nach zweieinviertel Jahren noch gegeben? Erinnern sich die Schülerinnen und Schüler nach diesem Zeitraum überhaupt noch an die Stationen? Und wenn dies gegeben ist, wie umfangreich sind die Erinnerungen noch? Diese Fragen gilt es vor der Analyse der Wirksamkeit zu klären.

Hier konnten insgesamt 80 Fragebögen ausgewertet werden: 10 Bögen (11%) mussten ausgesondert werden, da sich bei ihnen eine zu ausgeprägte Ja-sage-Tendenz (vgl. Abschnitt 2.3.3 in dieser Arbeit) zeigte. Dieser Wert ist als erfreulich niedrig zu bewerten: Die Schülerinnen und Schüler bearbeiteten die Fragestellung mit umfassender Gründlichkeit und Konzentration. Dieses Ergebnis untermauert die in der Auswertung des Posttest II der Untersuchungsstufe I vorgetragene These, dass der dort vorhandene hohe Ausschuss der Fragebögen auf die Wiederholung der niedrigkomplexen Aufgabe zurückzuführen ist.

Die folgende Tabelle fasst einige zentrale deskriptive Werte der Variablen des ersten Teils des Fragebogens zusammen:

Bezeichnung	Wert
Arithmetisches Mittel	6,6
Konfidenzintervall (95%)	6,0 – 7,1
Modalwert	6,0 (Es existieren mehrere Modalwerte!)
Standardabweichung	2,5
range	12,0

Tabelle 3.3.1.1: Deskriptive Aspekte; Untersuchungsstufe II;
- deskriptive Werte zum ersten Teil des Fragebogens
Der Verfasser

Damit erinnern sich die Schülerinnen und Schüler nach der genannten Zeitspanne noch an 50,7 Prozent der Stationen der MINIPHÄNOMENTA. Die Anzahl der erinnerten Stationen liegt damit niedriger als in der Untersuchungsstufe I. Bedenkt man aber die Zeitspanne sowie die Tatsache, dass die untersuchten Schülerinnen und Schüler zum Zeitpunkt der MINIPHÄNOMENTA an der Schule in der ersten beziehungsweise in der zweiten Klasse⁵⁶ waren, so ist die Anzahl der erinnerten Stationen als extrem hoch zu beurteilen. Weiterhin positiv ist die geringe Dispersion der Werte zu bewerten. Die Erinnerung an die MINIPHÄNOMENTA scheint damit ein relativ stabiles Merkmal über die einzelnen Gruppen und Mitglieder der Stichprobe zu sein.

Auch nach einer großen Zeitspanne erinnern sich die Schülerinnen und Schülern noch an die Stationen der MINIPHÄNOMENTA.

Damit ist bewiesen, dass sich die Schülerinnen und Schüler nach der genannten Zeitspanne noch an die Stationen erinnern. Fraglich ist jedoch noch, wie diese Erinnerung beschaffen ist. Dies wird im zweiten Teil des Fragebogens erhoben. Die Auswertung wird anhand einer quantitativen Inhaltsanalyse vorgenommen, deren Ergebnisse in der folgenden Tabelle wiedergegeben sind:

⁵⁶ In Bezug auf diese Untersuchungsstufe ergeben sich zwei forschungsmethodische Probleme, die die Gültigkeit der Ergebnisse einschränken (vgl. Abschnitt 2.2 in dieser Arbeit):

1. Geringe interne Validität in Bezug auf den zentralen Kausalitätsschluss
2. Im Gegensatz zu den Schülerinnen und Schülern der Untersuchungsstufe I, die zum Zeitpunkt des treatments in der dritten beziehungsweise vierten Klassenstufe waren, waren die Untersuchungsteilnehmerinnen und Untersuchungsteilnehmer zum Zeitpunkt des treatments in der ersten beziehungsweise zweiten Klassenstufe. Dies schränkt die Gültigkeit der Ergebnisse ein.

	Arithmetisches Mittel	Standardabweichung
Anzahl der Stichpunkte	5,3	1,3
Anzahl der richtigen Stichpunkte	4,2	1,3
Aufbau des Versuches	3,2	0,9
Funktion des Versuches	3,0	1,0
Effekt des Versuches	3,1	1,0
Erklärungsansätze für den Versuch	1,4	0,8
Affektive Äußerungen zum Versuch	3,2	0,8
Eidetischer Charakter der Äußerungen	3,5	0,8

Tabelle 3.3.1.2: Deskriptive Aspekte; Untersuchungsstufe II;
- Ausprägungen in den Analysekatgorien
Der Verfasser

Die Werte belegen die nachhaltige [Erinnerung] der Schülerinnen und Schüler an die Stationen der MINIPHÄNOMENTA: ‚Aufbau‘, ‚Funktion‘ und ‚Effekt‘ des Versuches können größtenteils richtig beschrieben werden. Die Erinnerungen der Kinder sind ‚affektiv‘ gefärbt und haben ‚eidetischen Charakter‘. Der ausgewählte Versuch wurde im Durchschnitt mit 5,3 Stichpunkten beschrieben: Auch wenn dieser Wert im Vergleich zur Untersuchungsstufe I geringfügig gesunken ist, so ist auch hier positiv hervorzuheben, dass die Schülerinnen und Schüler kaum unbedeutende oder falsche Stichpunkte nannten.

Für eine vertiefende inhaltliche Interpretation kategorisiere ich die individuell erreichten kumulierten Werte des zweiten Teils des Fragebogens entsprechend dem Abschnitt 3.2.2.1. Die folgende Grafik stellt die relativen und absoluten Häufigkeiten der Kategorien dar:

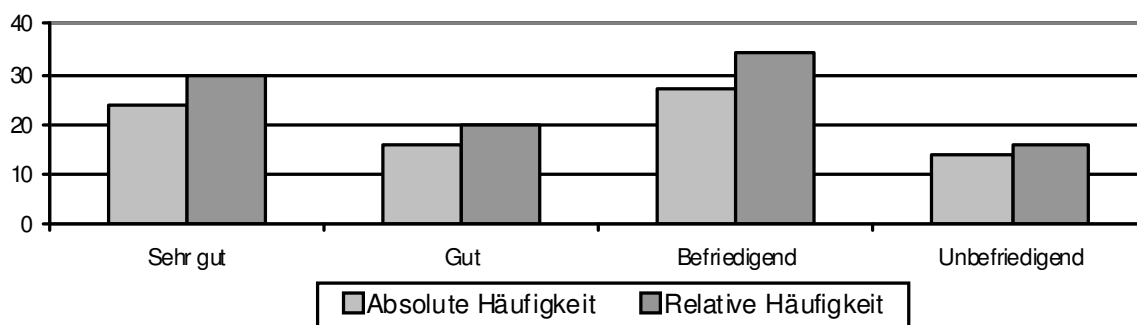


Diagramm 3.3.1.1: Deskriptive Aspekte; Untersuchungsstufe II;
- relative und absolute Häufigkeiten
Der Verfasser

Die Grafik verdeutlicht die hohe Qualität der Erinnerung der Schülerinnen und Schüler an die Stationen der MINIPHÄNOMENTA. 67 Kinder und damit 83,3 Prozent der untersuchten Versuchspersonen konnten den Kategorien ‚sehr gut‘, ‚gut‘ oder ‚befriedigend‘ zugeordnet werden.

Die Schülerinnen und Schüler erinnern sich größtenteils umfassend und korrekt an die Stationen der MINIPHÄNOMENTA. Damit ist die Voraussetzung für die Analyse der Wirksamkeit gegeben.

Die Variable allgemeines Experimentierverhalten

Das folgende Histogramm zeigt die Verteilung der Schülerinnen und Schüler auf die Verhaltenskategorien⁵⁷, für den Prätest und die Untersuchungsstufe II. Aufgrund des unterschiedlichen Stichprobenumfanges ist eine Darstellung in Relativwerten erforderlich:



Diagramm 3.3.1.2: Allgemeines Experimentierverhalten; Untersuchungsstufe II - relative Häufigkeiten des Prätests und der Untersuchungsstufe II
Der Verfasser

Beide Verteilungen sind deutlich unterschiedlich: Damit scheint ASIP in Bezug auf das [allgemeine Experimentierverhalten] der Schülerinnen und Schüler auch nach der genannten Zeitspanne noch wirksam zu sein.

Bevor ich mit der inhaltlichen Interpretation des Unterschiedes der beiden Verteilungen fortfahre, soll überprüft werden, ob sich die Ergebnisse des Prä- und Posttests statistisch bedeutsam unterscheiden. Dies geschieht durch einen k*1-Chi-Quadrat-Test, dessen Ergebnisse in der folgenden Tabelle zusammengefasst sind:

⁵⁷ Die folgenden Auswertungsschritte für die Variable [allgemeines Experimentierverhalten] beziehen sich auf die insgesamt 44 in der Experimentalgruppe II durch Verhaltensbeobachtung und Interview erhobenen Datensätze.

Bezeichnung	Wert
df	3
χ^2	65,6

Tabelle 3.3.1.3: Allgemeines Experimentierverhalten; Untersuchungsstufe II
- Ergebnisse des Chi-Quadrat-Testes
Der Verfasser

Beide Verteilungen unterscheiden sich hoch signifikant. Auf dieser Basis konnte mit einem Wert von $\epsilon = 2,4$ eine hohe Effektstärke ermittelt werden (vgl. Bortz 1995, S. 150 ff.). Damit zeigen die Schülerinnen und Schüler nach der Teilnahme an der MINIPHÄNOMENTA ein deutlich kompetenteres [allgemeines Experimentierverhalten]. ‚Ablehnende Reaktionen‘ gegenüber dem Experimentieren kommen nicht mehr vor. Erreichten im Vorfeld von ASIP etwa zwei Drittel der Schülerinnen und Schüler lediglich die Verhaltenskategorie II, so ist im Posttest das ‚elaborierte Experimentierverhalten‘ zum Regelfall geworden.

Die MINIPHÄNOMENTA ist in Bezug auf das [allgemeine Experimentierverhalten]:

HOCHWIRKSAM (Ebene I)

Nach der Teilnahme waren die Schülerinnen und Schüler experimentell deutlich kompetenter. Dies zeigt die hohe Effektstärke von $\epsilon = 2,4$.

Die Variable Einstellung gegenüber Phänomen und ihrer Bearbeitung

In der Untersuchungsstufe II konnten 90 Fragebögen zur [Einstellung der Schülerinnen und Schüler gegenüber Phänomen und ihrer Bearbeitung] ausgewertet werden. Die folgende Tabelle fasst die wichtigsten deskriptiven Werte der Referenzgruppe aus der Untersuchungsstufe I und der Untersuchungsstufe II zusammen:

	Prätest	Untersuchungsstufe II
Arithmetisches Mittel	42,4	45,2
Konfidenzintervall (95%)	41,5 – 43,3	43,4 – 47,0
Modalwert	37	52
Standardabweichung	9,3	8,5
range	44	32

Tabelle 3.3.1.4: Einstellung gegenüber Phänomen und ihrer Bearbeitung; Untersuchungsstufe II;
- deskriptive Kennwerte
Der Verfasser

Die Werte zeigen, dass im Durchschnitt die Einstellung der Schülerinnen und Schüler auch nach einer langen Zeitspanne nach der Teilnahme an der MINIPHÄNOMENTA noch positiver ist. In der folgenden Tabelle wird die Punktwertdifferenz auf ihre statistische Bedeutsamkeit unter der Verwendung eines t-Testes für unabhängige Stichproben untersucht:

Bezeichnung	Wert
df	89
t-Wert	5,0

Tabelle 3.3.1.5: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Untersuchungsstufe II
- Ergebnisse des t-Testes
Der Verfasser

Der Unterschied zwischen beiden Mittelwerten ist hoch signifikant. Es ergibt sich mit einer Überlappung der Prä- und der Posttestverteilung von 69% eine mittlere Effektgröße (vgl. Bortz, Döring 2006, S. 607 f.). Nach der Teilnahme an ASIP haben die Kinder damit eine leicht positivere [Einstellung gegenüber Phänomen und ihrer Bearbeitung].

Macht sich dieser Unterschied auch in den in Abschnitt 3.2.1.2 definierten Einstellungskategorien bemerkbar? Dies sei mit Hilfe der folgenden Grafik überprüft, die die relativen Häufigkeiten der Verhaltenskategorien des Prätests der Untersuchungsstufe I und die der Experimentalgruppe II gegenüberstellt:

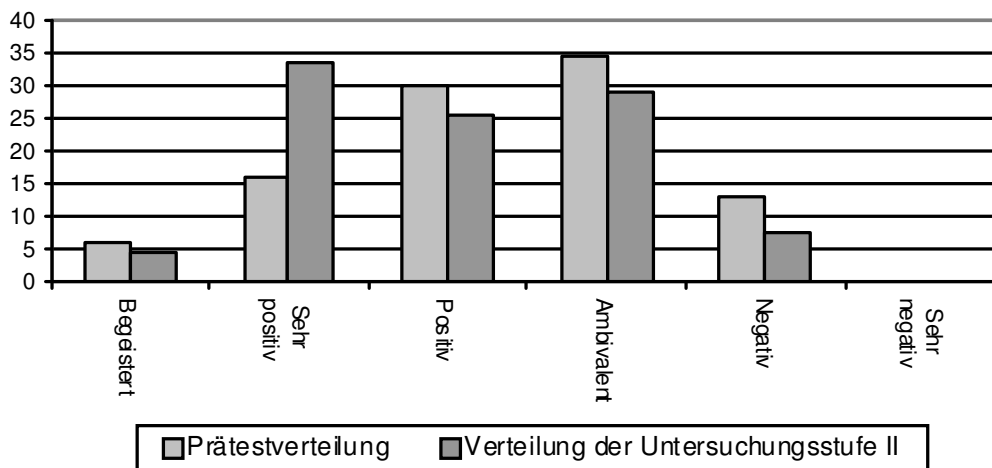


Diagramm 3.3.1.3: Einstellung gegenüber Phänomenen und ihrer Bearbeitung; Untersuchungsstufe II
- relative Häufigkeit im Prätest und in der Untersuchungsstufe II
Der Verfasser

Beide Verteilungen zeigen deutliche Unterschiede, so dass eine Signifikanzüberprüfung aufgrund der Evidenz der Datenlage unterbleiben kann. Durch den Einfluss des Projektes wiesen insbesondere deutlich mehr Schülerinnen und Schüler eine ‚sehr positive Einstellung‘ auf. Weiterhin ging die Anzahl derjenigen Kinder mit einer ‚ambivalenten und negativen Einstellung‘ deutlich zurück.

ASIP ist in Bezug auf die Variable [Einstellung gegenüber Phänomenen und ihrer Bearbeitung]:

WIRKSAM (Ebene I)

Es liegt eine mittlere Effektgröße vor mit einer Überlappung der Prä-Posttestverteilung von 69%.

Die Variable formale Kompetenz

Ich beginne mit der Auswertung der Rahmendaten⁵⁸: In der folgenden Tabelle stelle ich die Ergebnisse des Prätests der Untersuchungsstufe I und die der Untersuchungsstufe II gegenüber:

	Prätest			Untersuchungsstufe II		
	Dauer des Experimentierens	Klima beim Experimentieren	Konzentration beim Experimentieren	Dauer des Experimentierens	Klima beim Experimentieren	Konzentration beim Experimentieren
Arithmetisches Mittel	4,4	2,5	2,7	7,9	3,0	3,2
Standardabweichung	2,0	0,9	1,0	2,1	0,9	0,9

Tabelle 3.3.1.6: formale Kompetenz; Untersuchungsstufe II - deskriptive Werte der Rahmendaten
Der Verfasser

In der Tabelle 3.3.1.7 sind die Ergebnisse der Überprüfung der Mittelwertsdifferenzen zwischen den Prätestergebnissen und den Ergebnissen der Untersuchungsstufe II wiedergegeben:

	Dauer des Experimentierens	Klima beim Experimentieren	Konzentration beim Experimentieren
df	27	27	27
t-Wert	13,1	5,5	4,1

Tabelle 3.3.1.7: formale Kompetenz; Untersuchungsstufe II - Ergebnisse des t-Tests
Der Verfasser

Damit kann in allen Fällen die Forschungshypothese angenommen werden. Die Schülerinnen und Schüler arbeiteten nach der Teilnahme an der MINIPHÄNOMENTA länger an einem Versuch, waren konzentrierten und das Klima bei der Gruppenarbeit war besser.

⁵⁸ Die folgenden Auswertungsschritte für die Variable [formale Kompetenz] basieren auf der Analyse von 39 Untersuchungen, die in der Untersuchungsstufe II durchgeführt wurden.

Durch die MINIPHÄNOMENTA konnten die Rahmenaspekte des Experimentierens entscheidend verbessert werden.

Damit ist der Einfluss von ASIP auf die Rahmenaspekte nachgewiesen. Doch wie sieht es mit der tatsächlichen [formalen Kompetenz] der Schülerinnen und Schüler aus? Die folgende Tabelle stellt einige wichtige deskriptive Daten des Prätests der Untersuchungsstufe I und der Untersuchungsstufe II dar:

	Prätest	Untersuchungsstufe II
Arithmetisches Mittel	33,4	46,7
Konfidenzintervall (95%)	31,9 – 34,8	42,4 – 50,9
Modalwert	29	64
Standardabweichung	9,2	11,0
range	39	37

Tabelle 3.3.1.8: formale Kompetenz; Untersuchungsstufe II
- deskriptive Kennwerte des Prätests und der Untersuchungsstufe II
Der Verfasser

Die beträchtliche Differenz der Mittelwerte wurde durch einen t-Test für unabhängige Stichproben untersucht. Mit einem t-Wert von -8,5 und 27 Freiheitsgraden ergibt sich ein hoch signifikanter Unterschied. Darauf aufbauend konnte mit einer Überlappung der Prä- und der Posttestverteilung von 42% eine hohe Effektstärke ermittelt werden (vgl. Bortz, Döring 2006, S. 607 f.). Die Schülerinnen und Schüler sind damit nach der Teilnahme formal deutlich kompetenter als solche, die an dem Projekt nicht teilgenommen haben. Besonders erfreulich ist, dies zeigen die Ergebnisse, dass das höhere Maß an [formaler Kompetenz] überaus langzeitwirksam ist. Selbst nach dem hier vorliegenden extrem langen Zeitraum zwischen der Teilnahme an der MINIPHÄNOIMENTA und der Untersuchung ist die [formale Kompetenz] im Vergleich zum Posttest I (vgl. Abschnitt 3.2.2.1 in dieser Arbeit) nur geringfügig gesunken.

Die MINIPHÄNOMENTA ist in Bezug auf die [formale Kompetenz] der Schülerinnen und Schüler in der Untersuchungsstufe II

HOCHWIRKSAM (Ebene I)

Durch die Teilnahme an der MINIPHÄNOMENTA verbesserten sich die ‚Rahmenaspekte‘ des Experimentierens. Die Schülerinnen und Schüler arbeiten hoch signifikant länger und konzentrierter an einem Experiment. Auch das ‚Gruppenklima‘ ist während des Experimentierens hoch signifikant positiver als im Prätest der Untersuchungsstufe I.

Die eigentlichen [formalen Kompetenz] der Schülerinnen und Schüler sind nach der MINIPHÄNOMENTA enorm gestiegen. Nach 27 Monaten gibt es immer noch eine hohe Effektgröße und damit eine Überlappung der Prä- und Posttestverteilung von 42%.

Die Variable aktuelle Motivation

Im Abschnitt 2.3.6 konnte nachgewiesen werden, dass die Motivationsstruktur von Kindern in der dritten Klasse noch nicht ausreichend differenziert ist, um durch die hier vorgeschlagene Methode gemessen zu werden. Daher beziehen sich die folgenden Ausführungen, wie bereits für Untersuchungsstufe I (vgl. Abschnitt 3.2 in dieser Arbeit), lediglich auf die vierte Klassenstufe.

In der folgenden Tabelle werden einige wichtige deskriptive Kennwerte der Ergebnisse⁵⁹ des Prätests aus der Untersuchungsstufe I und der Untersuchungsstufe II gegenübergestellt:

	Prätest	Untersuchungsstufe II
Arithmetisches Mittel	45,6	47,3
Konfidenzintervall (95%)	44,4 – 46,8	45,5 – 49,1
Modalwert	41,0	50
Standardabweichung	9,0	6,0
range	41	25

Tabelle 3.3.1.9: Aktuelle Motivation; Untersuchungsstufe II
- deskriptive Kennwerte
Der Verfasser

⁵⁹ Die folgende Analyse zur Variablen [aktuelle Motivation] basiert auf insgesamt 46 ausgewerteten Fragebögen.

In Bezug auf diese Variable zeigt sich folglich nur ein geringer Mittelwertsunterschied von 1,7 Rohtestpunkten. Dieser wird überprüft durch einen t-Test für unabhängige Stichproben. Mit einem t-Wert von 1,4 und 45 Freiheitsgraden liegt ein nicht signifikanter Unterschied vor. Insofern ist die Bestimmung der Effektstärke hier nicht sinnvoll. Damit ist die [Motivation] der Schülerinnen und Schüler ein naturwissenschaftliches Phänomen zu bearbeiten nach der Teilnahme an der MINIPHÄNOMENTA höher. Dieser Unterschied ist jedoch nicht signifikant, kann folglich zufällig bedingt sein.

Die MINIPHÄNOMENTA war in Bezug auf die [aktuelle Motivation] der Schülerinnen und Schüler in der Untersuchungsstufe II:

UNWIRKSAM (Ebene I)

Die Variable Wissen

Die folgende Tabelle fasst die Ergebnisse des Prätests der Untersuchungsstufe I und der Untersuchungsstufe II⁶⁰ für beide Indices (vgl. Abschnitt 3.2.1.2 in dieser Arbeit) zusammen:

	Index I (Qualität)	Index II (Quantität)
Prätest		
Arithmetisches Mittel	1,76	4,80
Standardabweichung	0,6	1,2
Untersuchungsstufe II		
Arithmetisches Mittel	4,5	5,4
Standardabweichung	1,2	0,2

Tabelle 3.3.1.10: Wissen; Untersuchungsstufe II;
- deskriptive Werte des Prätests und der Untersuchungsstufe II
Der Verfasser

Die Mittelwertsunterschiede beider Indices werden durch einen t-Test für unabhängige Stichproben untersucht. In beiden Fällen kann die Forschungshypothese angenommen werden (Index I: $t = 14,8 / df = 38$; Index II: $t = 5,2 / df = 38$). Auf dieser Basis wurden die Effektgrößen ermittelt. Für den Index I („Qualität“) wurde mit einem Überlappungsbereich der Prä- und der Posttestverteilung von 5% ein sehr hoher, für den Index II („Quantität“) mit einem Überlappungsbereich beider Verteilungen von 55% ein hoher Effekt ermittelt (vgl. Bortz, Döring 2006, S. 607 f.).

⁶⁰ Insgesamt konnten für die Analyse der Variablen [Wissen] 39 Gruppenarbeiten der Schülerinnen und Schüler untersucht werden.

Zur vertiefenden inhaltlichen Interpretation sollen die individuell erreichten Werte Kategorien zugewiesen werden. Ich orientiere mich in Bezug auf die cut-off-scores an den Erörterungen im Abschnitt 3.2.1.2. Die folgenden Diagramme stellen die relativen kumulierten Kategorienhäufigkeiten der Prätestergebnisse und der Untersuchungsstufe II für beide Indices dar:

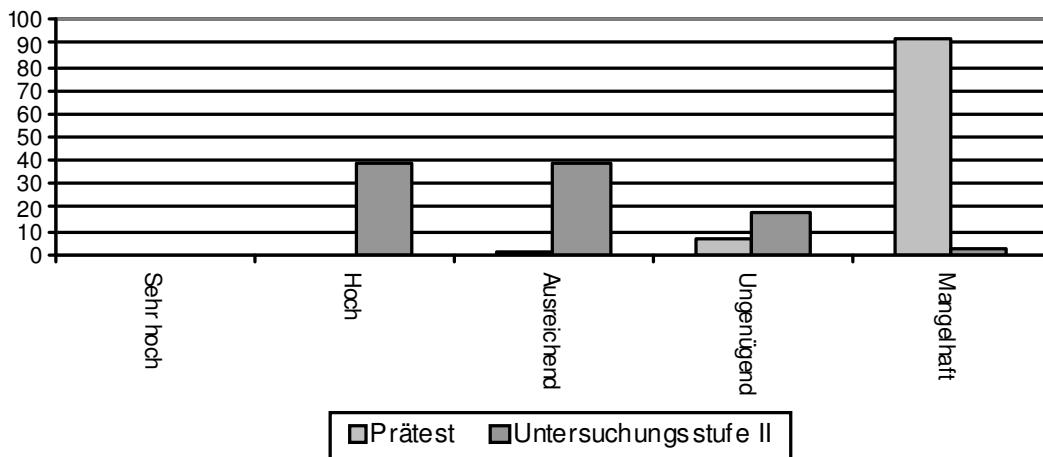


Diagramm 3.3.1.4: Wissen; Untersuchungsstufe II;
- Kategorienhäufigkeiten des Prätests und der Untersuchungsstufe II für den Index I („Qualität“)
Der Verfasser

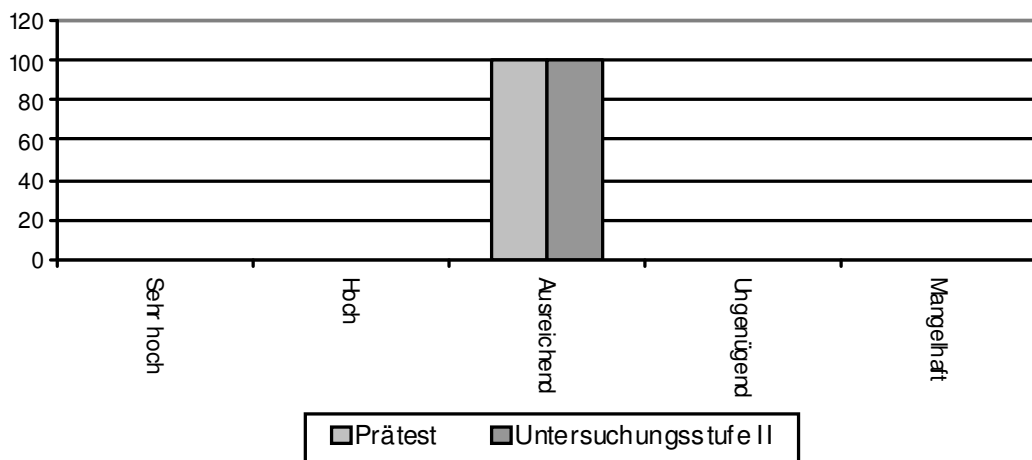


Diagramme 3.3.1.5: Wissen; Untersuchungsstufe II;
- Kategorienhäufigkeiten des Prätests und der Untersuchungsstufe II für den Index II („Quantität“)
Der Verfasser

Die Ergebnisse unterstreichen die Langzeitwirksamkeit der MINIPHÄNOMENTA. Ihre Effekte zeigen sich auch in Bezug auf die Variable noch in ähnlich starker Form, wie bereits im Posttest I. Die Schülerinnen und Schüler richten nach der Teilnahme an ASIP ihr Augenmerk verstärkt auf die inhaltliche ‚Qualität der Lösung‘ und bemühen sich nicht mehr vorrangig um eine möglichst vollständige Bearbeitung der Aufgabe. Sie wird jedoch nicht außer Acht gelassen, ihre ‚Quantität‘ ist immer noch befriedigend.

Die MINIPHÄNOMENTA war bei den Schülerinnen und Schülern in Bezug auf die Variable [Wissen] in der Untersuchungsstufe II:

HOCHWIRKSAM (Ebene I)

- Beide Indizes sind durch die Teilnahme hoch signifikant gestiegen. Für beide Indices konnte auf dieser Basis eine hohe Effektstärke ermittelt werden mit einem Überlappungsbereich der Prä- Posttestverteilungen von 5% (Index I) bzw. 55% (Index II).
- Die Schülerinnen und Schüler konzentrieren sich nach dem Projekt vermehrt auf die inhaltliche ‚Qualität‘ ihrer Arbeiten, ohne dabei den Umfang und die Vollständigkeit zu vernachlässigen.

3.3.2 Zusammenfassung der Ergebnisse: Auswertung im Hinblick auf die Hypothesen

Die folgende Tabelle fasst die Wirksamkeitsstufen der einzelnen Variablen sowie einige zentrale Forschungsbefunde zusammen:

Variable	Wirksamkeitsstufe	Zusammenfassung der Ergebnisse
Deskriptive Aspekte	Erinnerung vorhanden: Weitere Analyse sinnvoll	Nach dem genannten Zeitraum erinnerten sich die Schülerinnen und Schüler noch an etwa die Hälfte der gezeigten Stationen. Sie können Aufbau, Funktion und Effekt einer Station gut und umfassend beschreiben. Ihre Erinnerungen sind affektiv gefärbt und haben eidetischen Charakter.
Allgemeines Experimentierverhalten	HOCHWIRKSAM	Nach der Teilnahme an der MINIPHÄNOMENTA zeigten die Schülerinnen und Schüler ein hoch signifikant besseres allgemeines Experimentierverhalten. Ablehnende Reaktionen gegenüber dem Experimentieren kommen nicht mehr vor. Erreichten im Vorfeld von ASIP etwa zwei Drittel der Schülerinnen und Schüler lediglich die Verhaltenskategorie II, so ist in der Untersuchungsstufe II das elaborierte Experimentierverhalten (Stufe IV) das am häufigsten Vorkommende.
Einstellung	WIRKSAM	Es konnte eine mittlere Effektgröße ermittelt werden.
Formale Kompetenz	HOCHWIRKSAM	Die Schülerinnen und Schüler arbeiteten länger an einer Station, waren konzentriert und es gab produktive Gruppenarbeitsprozesse. Auch die formale Kompetenz stieg beträchtlich. Dies drückt sich in der hohen Effektgröße aus.
Aktuelle Motivation	UNWIRKSAM	Die aktuelle Motivation war nach der Teilnahme an der MINIPHÄNOMENTA nicht signifikant höher als vor der Teilnahme.
Wissen	HOCHWIRKSAM	Für beide Indices konnte eine hohe Effektgröße ermittelt werden. Durch die Teilnahme verlagerte sich der Schwerpunkt bei der Bearbeitung der Aufgabe von dem Versuch einer möglichst vollständigen Lösung zu dem Bemühen um eine hohe inhaltliche Qualität, ohne dabei den Umfang der Lösung zu vernachlässigen.

Insgesamt war damit das treatment WIRKSAM.

Tabelle 3.3.2.1: Zusammenfassung; Untersuchungsstufe II;
- Zusammenfassung der Ergebnisse der Untersuchungsstufe II
Der Verfasser

Abschließend sollen an dieser Stelle die für diese Untersuchungsstufe relevanten Hypothesen diskutiert werden.

Hypothese:

Die MINIPHÄNOMENTA ist auch nach dieser langen Zeitspanne noch wirksam.

Die Hypothese kann angenommen werden. Die Wirksamkeit von ASIP konnte auf vielen Variablen belegt werden. Am deutlichsten ist der Effekt auf die [formale Kompetenz] der Schülerinnen und Schüler. Durch die Teilnahme an dem Programm lernten sie nachhaltig eigenständig zu experimentieren und kontrolliert über den Versuch nachzudenken. Diese Annahme wird durch die Ergebnisse auf den Variablen [allgemeines Experimentierverhalten] und [Wissen] unterstützt. Auch die Grundlage sich mit physikalischen Fragestellungen auseinander zu setzen verbesserte sich. So wiesen die Schülerinnen und Schüler nach der Teilnahme an ASIP eine positivere [Einstellung gegenüber Phänomen und ihrer Bearbeitung] auf. Lediglich auf die [Motivation] der Schülerinnen und Schüler zur tiefergehenden kognitiven Bearbeitung physikalischer Sachverhalte hatte die MINIPHÄNOMENTA, dies deutete sich bereits im Posttest II der Untersuchungsstufe I an, keinen Einfluss.

Hypothese:

Es ist anzunehmen, dass ASIP auf einige Variablen besonders wirksam ist.

Besonders wirksam war die MINIPHÄNOMENTA auf folgende Variablen: [allgemeines Experimentierverhalten], [Einstellung], [formale Kompetenz] und [Wissen]. In dieser Aufzählung sticht die [formale Kompetenz] deutlich heraus. Die Differenz zwischen dem Prätest der Untersuchungsstufe I und der Experimentalgruppe II ist immens.

3.4 Erweiterung des Wirksamkeitsbegriffes

Die Wirksamkeit der MINIPHÄNOMENTA wurde in dieser Arbeit bislang lediglich im Hinblick auf die Schülerinnen und Schüler der dritten und vierten Klassenstufe analysiert. Das treatment wurde als wirksam erachtet, sofern sich bei den Kindern auf den abhängigen Variablen Prä-Posttest-Unterschiede nachweisen ließen. Aus der Projektstruktur (vgl. insbesondere Abschnitt 1.3.1 dieser Arbeit) ergeben sich vier weitere Wirksamkeitsebenen:

1. Zusammenarbeit Eltern-Lehrerschaft: Dieses Verhältnis wird von beiden Seiten häufig als schwierig beschrieben. Die MINIPHÄNOMENTA wäre hier wirksam, sofern es durch die gemeinsame Projektarbeit der genannten Personengruppen gelänge dieses Verhältnis im Sinne einer positiven Zusammenarbeit auch über die MINIPHÄNOMENTA hinaus zielorientiert und belastbar zu gestalten.
2. Auseinandersetzung in der Familie mit physikalischen Inhalten: Derzeit liegen keine Studien zur Auseinandersetzung von Familien mit physikalischen Inhalten vor. Zu beantworten wären zum Beispiel die folgenden Fragen: Ist die Diskussion über physikalische Fragen Teil der Alltagskommunikation der Familien? Wie häufig und intensiv experimentieren Eltern mit ihren Kindern? Wie gestaltet sich die Informationssuche in Bezug auf physikalische Fragestellungen? Die MINIPHÄNOMENTA wäre wirksam, sofern sie positiven Einfluss auf die Auseinandersetzung der Familien mit physikalischen Inhalten hat. Mögliche Indikatoren wären hier häufigere Diskussionen über physikalische Fragestellungen, regelmäßiger gemeinsamer Besuch naturwissenschaftlicher Museen usw.
3. Einstellung der Eltern gegenüber der Physik: Die Einstellung gegenüber Naturwissenschaften und insbesondere der Physik ist, dies konnte ich in den Abschnitten 1.4.3 und 2.4.2 dieser Arbeit zeigen, häufig negativ geprägt. Befragte Personen betonen zwar die Notwendigkeit physikalischen Denkens und Forschens, können sich jedoch für derartige Fragestellungen in keinsten Weise emotional begeistern. Durch die intensive Einbeziehung der Eltern in das Projekt MINIPHÄNOMENTA könnte die Einstellung nachhaltig positiv beeinflusst werden. ASIP wäre wirksam, sofern die Eltern durch das gemeinsame Experimentieren mit ihren Kindern und durch die Arbeit im Projekt ihre Einstellung gegenüber der Physik positiv verändern könnten.

4. Einfluss der MINIPHÄNOMENTA auf den Unterricht: Im Abschnitt 1.3.3 dieser Arbeit zitierte ich die Studie von Sauer (2005). Dieser konnte in einem qualitativen und einzelfallorientierten Vorgehen die Einflüsse der MINIPHÄNOMENTA auf den schulischen Unterricht zeigen. Dieser erste Zugang zum Feld müsste durch ein breiter angelegtes, standardisiertes Vorgehen weiter validiert werden. Die MINIPHÄNOMENTA wäre wirksam auf den Unterricht in der Schule, sofern Fragen der physikalischen Elementarbildung in experimenteller, das heißt handlungsorientierter Art und Weise, ein stärkeres Gewicht eingeräumt würde. Ein weiterer Anhaltspunkt der Wirksamkeit des treatments wäre der weitgehende Verzicht auf lehrerorientierte Erklärungen. Vielmehr sollten die Schülerinnen und Schüler auf der Basis des Experimentes eigenständige Erklärungen erarbeiten und in der Gruppe diskutieren.

Die genannten Wirksamkeitsebenen transzendieren die in dieser Studie gewählte schülerorientierte Wirksamkeitsdefinition in vielfacher Hinsicht. Für die erste der genannten Wirksamkeitsebenen wurden Daten erhoben, die im Folgenden kurz interpretiert werden sollen. Es muss an dieser Stelle betont werden, dass diese Daten lediglich einen ersten Zugang zum Feld darstellen, einen explorativen Charakter aufweisen. Insofern kommt den gewonnenen Ergebnissen eine geringere Validität zu als denen der vorangegangenen Studie. Diese Untersuchungen sollen lediglich den Blick dafür öffnen, dass die Wirksamkeit der MINIPHÄNOMENTA durch die Untersuchung der Schülerinnen und Schüler nicht abschließend geklärt ist. Sie stellt einen umfassenden und methodisch kontrollierten Zugang dar, der aber durch weitere Untersuchungen untermauert werden muss.

Ad 1: Zusammenarbeit Eltern-Lehrerschaft: Dieser Frage wurde im informellen Gespräch mit den Lehrerinnen und Lehrern im Rahmen der Untersuchung der einzelnen Schulklassen nachgegangen. Ich befragte hier ausgewählte Lehrerinnen und Lehrer in der Pause vor oder nach der Erhebung in ihrer jeweiligen Schulklasse. Dabei orientierte ich mich an den folgenden Kernideen:

- Die Elterninformationsveranstaltung zur MINIPHÄNOMENTA: Beteiligung der Eltern, Atmosphäre wären des Experimentierens, gegenseitiges Feedback
- Der Nachbau der Stationen: Aktivierung der Eltern, Hintergrund der teilnehmenden Elternteile, Anzahl der Versuche, Probleme und Lösungsansätze beim Bau der Exponate
- Der Einfluss des Projektes auf das allgemeine Verhältnis der beiden genannten Gruppen zueinander

Die Beteiligung der Eltern an der Informationsveranstaltung (vgl. Abschnitt 1.3.1 in dieser Arbeit) war in den betreffenden Schulen erstaunlich hoch. Von etwa 80% der Schülerinnen und Schüler erschien mindestens ein Elternteil. Dieses Ergebnis muss als extrem hoch bewertet werden, bedenkt man, dass beispielsweise die Teilnahme bei einem Elternabend nach der Aussage der betreffenden Lehrerinnen und Lehrer bei etwa 50% liegt. Die Gruppe der Eltern, der Schülerinnen und Schüler und der Lehrerinnen und Lehrer experimentierte begeistert bis zu zwei Stunden an den Stationen der MINIPHÄNOMENTA. Alle Seiten betonten, dass die gemeinsame Arbeit an den Stationen ihnen viel Freude bereitet hat, besonders positiv wird die Möglichkeit eines neuen Zuganges zu der oft angstbesetzten Thematik der Physik bewertet.

Die Aktivierung der Eltern für den Nachbau der Stationen war im Allgemeinen leicht und geschah zumeist im Rahmen der beschriebenen Informationsveranstaltung. Die Elternteile unterstützten den Aufbau eines eigenen Experimentierfeldes an der Schule ihrer Kinder auf ganz unterschiedliche Art und Weise. Einige bevorzugten die Unterstützung des Projektes durch Sach- oder Geldspenden, andere bauten in einer Arbeitsgruppe die Stationen nach. In allen Schulen gelang es den Schülerinnen und Schülern einige Experimente dauerhaft zur Verfügung zu stellen. Auftretende Probleme konnten durch die vielfältigen beruflichen Orientierungen der Eltern effizient gelöst werden. Insgesamt kann festgestellt werden, dass an dem Projekt insbesondere Eltern beteiligt waren, die intensiv Anteil an der schulischen Laufbahn ihres Kindes nehmen.

Durch die Teilnahme am Projekt entstand zwischen den Eltern und den Lehrerinnen und Lehrer ein intensiver Austauschprozess, der die Kommunikation über andere Sachverhalte förderte.

3.5 Ausblick

Die zentrale Untersuchungsfrage ist beantwortet, das heißt die Langzeitwirksamkeit der MINIPHÄNOMENTA wurde durch die Verwendung von zwei Untersuchungsstufen auf ebenfalls zwei unterschiedlich exakten Ebenen nachgewiesen.

Diese Feststellung bedeutet keinesfalls, dass damit die Begleitforschung zur Langzeitwirksamkeit von ASIP abgeschlossen ist. Vielmehr gilt, dies zeigen unterschiedliche wissenschaftstheoretische Vorstellungen über das Fortschreiten wissenschaftlicher Disziplinen, dass der Prozess des Forschens praktisch und theoretisch nicht abschließbar ist. Wissenschaft beginnt und endet mit offenen Fragen. Dabei erhellt sie sukzessive ihr Gegenstandsfeld, erweitert nach methodischen Prinzipien ihr Wissen (vgl. Chalmers 2001, S 53 ff.). Wissenschaftlicher Fortschritt beruht damit auf zwei Prinzipien: Zunächst kommt keiner Aussage abschließende Gültigkeit zu, weiterhin ergeben sich aus den bestehenden Ergebnissen differenziertere Fragestellungen.

In Bezug auf die hier verhandelte Forschungsfrage gilt es für spätere Vorhaben aus der Sicht des Autors vor allem zwei Aspekte zu klären:

1. Wie lässt sich die hohe Langzeitwirksamkeit der MINIPHÄNOMENTA erklären?
2. Sind Effekte der MINIPHÄNOMENTA im Physikunterricht der Sekundarstufe I und II nachweisbar?

Ad 1: Durch das hier angewandte Design konnte die Frage geklärt werden, ob die MINIPHÄNOMENTA wirksam ist und wo, das heißt in Bezug auf welche Aspekte sie wirksam ist. Ungeklärt blieb bislang warum das Projekt wirksam ist. Welche Faktoren und Prozesse führen also zu der hier festgestellten hohen Wirksamkeit von ASIP? Theoretisch sollte sich die Beantwortung dieser Frage an den vorgestellten Bausteinen des interaktiven Lernens (vgl. Abschnitt 1.4.3 in dieser Arbeit) orientieren. Es gilt die zentralen Definitionselemente des beschriebenen hochkomplexen Lernprozess herauszuarbeiten und diese als unabhängige Variablen in einem Experiment zu operationalisieren. Es ergibt sich ein längsschnittlicher mehrfaktorieller Experimentalplan, dessen abhängige Variable, die Langzeitwirksamkeit durch die hier eingeführten Messinstrumente beschrieben werden könnte.

Das vorgeschlagene Verfahren sei hier beispielhaft für die beiden unabhängigen Variablen „Freiwilligkeit“ und „Art der Stationen“ beschrieben: Zunächst werden eine Experimentalgruppe und eine Kontrollgruppe gebildet. Beide Gruppen nehmen an der MINIPHÄNOMENTA teil, jedoch unter verschiedenen Bedingungen: Während die Kontrollgruppe in der in Abschnitt 1.4.2 in dieser Arbeit beschriebenen Art und Weise an dem Projekt teilnimmt, werden in der Experimentalgruppe die beiden unabhängigen Variablen manipuliert. Hier wird zunächst das Merkmal der „Freiwilligkeit“ beseitigt. Die Schülerinnen und Schüler beschäftigen sich in dieser Gruppe mit den Stationen in einer fest vorgeschriebenen Art und Weise im Unterricht. Weiterhin wird in dieser Gruppe die „Art der Stationen“ verändert: Sie werden optisch aufwendig und insgesamt intransparenter gestaltet. Nach bestimmten Zeiträumen werden die abhängigen Variablen erneut erhoben. Durch die Verwendung einer Varianzanalyse wird festgestellt, in welcher Gruppe eine höhere Langzeitwirksamkeit vorhanden war. Durch die Interpretation der Haupteffekte und des Interaktionseffektes können so die zentralen Bedingungen des Lernens an den Stationen identifiziert werden.

Ad 2: Sind Schülerinnen und Schüler, die in der Grundschule an dem Projekt MINIPHÄNOMENTA teilgenommen haben im Physikunterricht in der Sekundarstufe I und II kompetenter. Die Beantwortung dieser Frage ist genauso interessant wie schwierig. Es ergeben sich zwei zentrale Hindernisse: Zunächst der extreme zeitliche Abstand zwischen Prä- und Posttest. Dieser führt zu einer Vielzahl von forschungspraktischen Problemen, wobei der zentralste sicherlich die enorme Stichprobenmortalität sein dürfte. Das zweite Problem ist deutlich gravierender. Es betrifft die interne Validität einer solchen Aussage. Angenommen man untersucht Schülerinnen und Schüler der vierten Klassenstufe im Prätest und wiederholt die Messung im Rahmen des Physikunterrichtes in der siebten Klassenstufe. Im Durchschnitt liegen damit folglich drei Jahre zwischen der Prätest- und der Posttestmessung. Die in diesem Zeitraum auftretenden Störfaktoren sind durch die vorliegende sozialwissenschaftliche Methodik nur unzureichend kontrollierbar. Der angenommene Effekt in der Experimentalgruppe ist damit nicht mehr sicher auf das treatment, die MINIPHÄNOMENTA zurückführbar.

Beide Probleme können aus der Sicht des Autors am besten durch die Verwendung eines qualitativen Verfahrens gelöst werden. In relativ kurzen Abständen (ca. sechs Monate) wird eine kleine Anzahl von Schülerinnen und Schülern sehr intensiv beobachtet und befragt. So ist zum einen ein intensiver Kontakt gewährleistet, der die Stichprobenmortalität begrenzt, zum anderen können die Leistungen, im Vergleich zu einer Kontrollgruppe, differenziert erfasst und Störfaktoren weitgehend kontrolliert werden.