

**Die Ergebnisqualität von Testkunden
aus unterschiedlichen soziodemografischen
Gruppen beim Mystery Shopping**

DISSERTATION

zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)
an der Universität Flensburg

Eingereicht von
Gunnar Grieger
aus Hamburg

Erstversion Juni 2008 | Überarbeitete Version Dezember 2008
(*Tag der Disputation: 28.11.2008*)

Inhaltsübersicht

1 Einführung

- 1.1 *Problemstellung*
- 1.2 *Relevanz des Themas*
- 1.3 *Zielsetzung und Gang der Untersuchung*

2 Mystery Shopping: Bewertete Darstellung der Grundlagen

- 2.1 *Mystery Shopping aus theoretischer Perspektive*
- 2.2 *Mystery Shopping aus praktischer Perspektive*
- 2.3 *Kritische Würdigung von Mystery Shopping*

3 Konzeption einer empirischen Untersuchung zur Ermittlung der Ergebnisqualität von Mystery Shoppern

- 3.1 *Identifikation von Komponenten zur Beurteilung der Ergebnisqualität von Mystery Shoppern*
- 3.2 *Konzeption eines Testszenarios zur Gegenüberstellung der Ergebnisqualität von Mystery Shoppern*
- 3.3 *Vorgehensweise bei der empirischen Untersuchung*

4 Die Ergebnisqualität von Mystery Shoppern im Spiegel empirischer Befunde

- 4.1 *Teilnehmerverhalten der Erhebung*
- 4.2 *Ergebnisse der Komponenten zur Bestimmung der Ergebnisqualität von Mystery Shoppern*
- 4.3 *Kernaussagen der ermittelten Ergebnisse bezogen auf die Merkmalsausprägungen von Mystery Shoppern*

5 Schlussbetrachtung

- 5.1 *Zusammenfassung der Ergebnisse und Implikationen für die Praxis*
- 5.2 *Restriktionen der Untersuchung*
- 5.3 *Weitere Ansatzpunkte für zukünftige Forschungsbemühungen zur Mystery Shopping-Thematik*

Inhaltsverzeichnis

Inhaltsübersicht.....	I
Abbildungsverzeichnis.....	VI
Tabellenverzeichnis.....	VIII
Abkürzungsverzeichnis	XI
1 Einführung.....	1
1.1 <i>Problemstellung</i>	1
1.2 <i>Relevanz des Themas</i>	2
1.2.1 Bedeutung der Dienstleistungsqualität für den Unternehmenserfolg und Defizite von Kundenbefragungen	2
1.2.2 Stellenwert von Mystery Shopping als Alternative zur Evaluation der Dienstleistungsqualität	3
1.3 <i>Zielsetzung und Gang der Untersuchung</i>	8
2 Mystery Shopping: Bewertete Darstellung der Grundlagen	10
2.1 <i>Mystery Shopping aus theoretischer Perspektive</i>	10
2.1.1 Begriffsbestimmung Mystery Shopping.....	10
2.1.1.1 Terminologie	10
2.1.1.2 Definitionen.....	11
2.1.2 Einordnung von Mystery Shopping als Verfahren zur Erhebung von Dienstleistungsqualität	13
2.1.2.1 Begriffsbestimmung Dienstleistung und Dienstleistungsqualität	13
2.1.2.2 Überblick der Verfahren zur Erhebung der Dienstleistungsqualität	14
2.1.2.3 Subjektive Verfahren zur Erhebung der Dienstleistungsqualität	16
2.1.2.4 Objektive Verfahren zur Erhebung der Dienstleistungsqualität	22
2.1.3 Entwicklung von Mystery Shopping im Zeitverlauf.....	27
2.1.3.1 Ursprung und Wurzeln von Mystery Shopping	27
2.1.3.2 Mystery Shopping von den 1950er bis zu den 2000er Jahren	28
2.1.3.3 Zukünftige Entwicklung von Mystery Shopping.....	32
2.1.4 Würdigung vergangener Forschungsbemühungen zur Mystery Shopping-Thematik	33
2.1.4.1 Überblick der Beiträge.....	33
2.1.4.2 Erkenntnisstand vergangener Forschungsbemühungen	43
2.2 <i>Mystery Shopping aus praktischer Perspektive</i>	45
2.2.1 Marktgröße und Marktumfeld.....	45
2.2.2 Zielsetzungen, Einsatzmöglichkeiten und Vorgehen bei Mystery Shopping	47
2.2.3 Konzeption einer Mystery Shopping-Erhebung anhand der einzelnen Phasen	
eines Projektverlaufs	52

2.2.3.1	Die Zielsetzungsphase einer Mystery Shopping-Erhebung	52
2.2.3.2	Operative Vorbereitung und Erhebungsphase einer Mystery Shopping-Erhebung	53
2.2.3.3	Auswertungsphase und Verwendung der Daten einer Mystery Shopping-Erhebung	68
2.3	<i>Kritische Würdigung von Mystery Shopping</i>	72
2.3.1	Nutzen von Mystery Shopping	72
2.3.2	Exkurs: Make or Buy bei Mystery Shopping?	74
2.3.2.1	Grundlagen der Transaktionskostentheorie	75
2.3.2.2	Mystery Shopping aus transaktionskostentheoretischer Perspektive	79
2.3.3	Ethische Aspekte von Mystery Shopping	84
2.3.4	Erfolgsfaktor Testerstab	87
2.3.4.1	Schulungsgrad der Mystery Shopper	87
2.3.4.2	Soziodemografische und einstellungsbezogene Merkmale der Mystery Shopper	89
3	Konzeption einer empirischen Untersuchung zur Ermittlung der Ergebnisqualität von Mystery Shoppnern	92
3.1	<i>Identifikation von Komponenten zur Beurteilung der Ergebnisqualität von Mystery Shoppnern</i>	92
3.1.1	Urteilstendenzen und Beurteilungssakkurtheit	93
3.1.2	Zuverlässigkeit: Beurteilerübereinstimmung und Beurteilerreliabilität	98
3.1.3	Ergänzende pragmatische Überlegungen zur Bestimmung der Ergebnisqualität	99
3.2	<i>Konzeption eines Testszenarios zur Gegenüberstellung der Ergebnisqualität von</i>	
	<i>Mystery Shoppnern</i>	102
3.2.1	Auswahl der Mystery Shopper und eines geeigneten Testobjektes	102
3.2.2	Konzeption des Beobachtungsbogens zur Ermittlung der Ergebnisqualität von	
	Mystery Shoppnern	104
3.2.2.1	Statische Messung der Dienstleistungsqualität mittels SERVPERF	105
3.2.2.2	Dynamisierung der Messung der Dienstleistungsqualität mit der Kontaktpunkt-Analyse	111
3.3	<i>Vorgehensweise bei der empirischen Untersuchung</i>	114
3.3.1	Erfassung der Komponenten der Ergebnisqualität	114
3.3.1.1	Erfassung der Urteilstendenzen	114
3.3.1.2	Erfassung der Zuverlässigkeit	118
3.3.1.3	Erfassung der Sorgfaltstendenz	120
3.3.2	Bewertung der Komponenten der Ergebnisqualität	121
3.3.3	Aufbau und Ablauf der Erhebung	123
4	Die Ergebnisqualität von Mystery Shoppnern im Spiegel empirischer Befunde	128
4.1	<i>Teilnehmerverhalten der Erhebung</i>	128
4.2	<i>Ergebnisse der Komponenten zur Bestimmung der Ergebnisqualität von Mystery Shoppnern</i>	130
4.2.1	Mittelwertstendenz von Mystery Shoppnern	130
4.2.1.1	Mittelwertstendenz bezüglich des Geschlechtes der Mystery Shopper	130
4.2.1.2	Mittelwertstendenz bezüglich des Alters der Mystery Shopper	131
4.2.1.3	Mittelwertstendenz bezüglich des höchsten Bildungsabschlusses der Mystery Shopper ..	134
4.2.1.4	Mittelwertstendenz bezüglich Tätigkeit der Mystery Shopper	135

4.2.1.5	Mittelwertstendenz bezüglich Erfahrung als Mystery Shopper	139
4.2.1.6	Mittelwertstendenz bezüglich der Fast Food-Affinität der Mystery Shopper.....	140
4.2.1.7	Mittelwertstendenz bezüglich der Restaurant-Affinität der Mystery Shopper	141
4.2.1.8	Mittelwertstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs	142
4.2.1.9	Mittelwertstendenz bezüglich der spezifischen Restaurant-Besuchshäufigkeit.....	143
4.2.1.10	Zusammenfassung der Ergebnisse der Mittelwertstendenz	144
4.2.2	Streuungstendenz von Mystery Shoppnern	145
4.2.2.1	Streuungstendenz bezüglich des Geschlechtes der Mystery Shopper.....	145
4.2.2.2	Streuungstendenz bezüglich des Alters der Mystery Shopper	146
4.2.2.3	Streuungstendenz bezüglich des höchsten Bildungsabschlusses der Mystery Shopper.....	147
4.2.2.4	Streuungstendenz bezüglich der Tätigkeit der Mystery Shopper.....	148
4.2.2.5	Streuungstendenz bezüglich der Erfahrung als Mystery Shopper	149
4.2.2.6	Streuungstendenz bezüglich der Fast Food-Affinität der Mystery Shopper	150
4.2.2.7	Streuungstendenz bezüglich der Restaurant-Affinität der Mystery Shopper.....	150
4.2.2.8	Streuungstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs.....	151
4.2.2.9	Streuungstendenz bezüglich der spezifischen Restaurant-Besuchshäufigkeit	152
4.2.2.10	Zusammenfassung der Ergebnisse der Streuungstendenz.....	154
4.2.3	Korrelationstendenz von Mystery Shoppnern.....	154
4.2.3.1	Korrelationstendenz bezüglich des Geschlechtes der Mystery Shopper.....	154
4.2.3.2	Korrelationstendenz bezüglich des Alters der Mystery Shopper	156
4.2.3.3	Korrelationstendenz bezüglich des höchsten Bildungsabschlusses der Mystery Shopper ..	157
4.2.3.4	Korrelationstendenz bezüglich der Tätigkeit der Mystery Shopper.....	159
4.2.3.5	Korrelationstendenz bezüglich der Erfahrung als Mystery Shopper	161
4.2.3.6	Korrelationstendenz bezüglich der Fast Food-Affinität der Mystery Shopper	162
4.2.3.7	Korrelationstendenz bezüglich der Restaurant-Affinität der Mystery Shopper	164
4.2.3.8	Korrelationstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs der Mystery Shopper.....	166
4.2.3.9	Korrelationstendenz bezüglich der spezifischen Restaurant-Besuchshäufigkeit	168
4.2.3.10	Zusammenfassung der Ergebnisse der Korrelationstendenz.....	169
4.2.4	Zuverlässigkeit von Mystery Shoppnern.....	170
4.2.4.1	Zuverlässigkeit bezüglich des Geschlechts.....	171
4.2.4.2	Zuverlässigkeit bezüglich der Altersklasse.....	172
4.2.4.3	Zuverlässigkeit bezüglich des höchsten Bildungsabschlusses	174
4.2.4.4	Zuverlässigkeit bezüglich der Tätigkeit.....	176
4.2.4.5	Zuverlässigkeit bezüglich der spezifischen Restaurant-Affinität.....	178
4.2.4.6	Zuverlässigkeit bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs	180
4.2.4.7	Zusammenfassung der Ergebnisse der Komponente Zuverlässigkeit.....	182
4.2.5	Sorgfältstendenz bei Mystery Shoppnern.....	184
4.2.5.1	Gewissenhaftigkeit bei Mystery Shoppnern	184
4.2.5.1.1	Gewissenhaftigkeit bezüglich des Geschlechtes der Mystery Shopper.....	185
4.2.5.1.2	Gewissenhaftigkeit bezüglich des Alters der Mystery Shopper	185
4.2.5.1.3	Gewissenhaftigkeit bezüglich des höchsten Bildungsabschlusses der	
	Mystery Shopper	187
4.2.5.1.4	Gewissenhaftigkeit bezüglich der Tätigkeit der Mystery Shopper	187
4.2.5.1.5	Gewissenhaftigkeit bezüglich der Erfahrung als Mystery Shopper	188
4.2.5.1.6	Gewissenhaftigkeit bezüglich der Fast Food-Affinität der Mystery Shopper	189
4.2.5.1.7	Gewissenhaftigkeit bezüglich der Restaurant-Affinität der Mystery Shopper.....	190
4.2.5.1.8	Gewissenhaftigkeit bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs.....	190

4.2.5.1.9	Gewissenhaftigkeit bezüglich der spezifischen Restaurant-Besuchshäufigkeit	191
4.2.5.1.10	Zusammenfassung der Ergebnisse der Gewissenhaftigkeit	192
4.2.5.2	Vollständigkeit des Ausfüllens bei Mystery Shoppem	192
4.2.5.2.1	Vollständigkeit bezüglich des Geschlechtes der Mystery Shopper	193
4.2.5.2.2	Vollständigkeit bezüglich des Alters der Mystery Shopper	193
4.2.5.2.3	Vollständigkeit bezüglich des höchsten Bildungsabschlusses der Mystery Shopper	194
4.2.5.2.4	Vollständigkeit bezüglich der Tätigkeit der Mystery Shopper	195
4.2.5.2.5	Vollständigkeit bezüglich der Erfahrung als Mystery Shopper	196
4.2.5.2.6	Vollständigkeit bezüglich der Fast Food-Affinität der Mystery Shopper	196
4.2.5.2.7	Vollständigkeit bezüglich der Restaurant-Affinität der Mystery Shopper	197
4.2.5.2.8	Vollständigkeit bezüglich der Häufigkeit eines Fast Food-Restauranbesuchs	198
4.2.5.2.9	Vollständigkeit bezüglich der spezifischen Restaurant-Besuchshäufigkeit	199
4.2.5.2.10	Zusammenfassung der Ergebnisse der Vollständigkeit	200
4.3	<i>Kernaussagen der ermittelten Ergebnisse bezogen auf die Merkmalsausprägungen von Mystery Shoppem</i>	201
4.3.1	Ergebnisqualität bezogen auf das Geschlecht der Mystery Shopper	201
4.3.2	Ergebnisqualität bezogen auf das Alter der Mystery Shopper	201
4.3.3	Ergebnisqualität bezogen auf den höchsten Bildungsabschluss der Mystery Shopper	201
4.3.4	Ergebnisqualität bezogen auf die Tätigkeit der Mystery Shopper	202
4.3.5	Ergebnisqualität bezogen auf die Erfahrung als Mystery Shopper	202
4.3.6	Ergebnisqualität bezogen auf die Fast Food-Affinität der Mystery Shopper	203
4.3.7	Ergebnisqualität bezogen auf die Restaurant-Affinität der Mystery Shopper	203
4.3.8	Ergebnisqualität bezogen auf die Fast Food-Häufigkeit der Mystery Shopper	204
4.3.9	Ergebnisqualität bezogen auf die spezifische Restaurant-Besuchshäufigkeit	204
5	Schlussbetrachtung	208
5.1	<i>Zusammenfassung der Ergebnisse und Implikationen für die Praxis</i>	208
5.2	<i>Restriktionen der Untersuchung</i>	211
5.3	<i>Weitere Ansatzpunkte für zukünftige Forschungsbemühungen zur Mystery Shopping-Thematik</i>	213
Anhang 1 – Ermittlung der Mittelwerts- und Streuungstendenz		215
Anhang 2 – Mittelwerte, Standardabweichungen der Gruppen		220
Anhang 3 – Abgrenzung zu der Arbeit von Schmidt (2007)		244
Literaturverzeichnis		245

Abbildungsverzeichnis

Abbildung 1-1: Aufbau der Untersuchung	9
Abbildung 2-1-1: Systematisierung der kundenorientierten Verfahren zur Erhebung der Dienstleistungsqualität	15
Abbildung 2-1-2: Hauptvarianten multiattributiver Messung von Dienstleistungsqualität	17
Abbildung 2-1-2: Mystery Shopping – Klassifikationsschema vergangener Forschungsbemühungen	34
Abbildung 2-2-1: Verpasste Umsatzchancen.....	48
Abbildung 2-2-2: Einsatz von Mystery Shopping und Kundenbefragung im Branchenvergleich.....	49
Abbildung 2-2-3: Mystery Shopping – Grundsätzlicher Bezugsrahmen	50
Abbildung 2-2-4: Vorgehensmodell von Mystery Shopping als Projektverlauf.....	51
Abbildung 2-2-5: Aggregationsgrad gewonnener Informationen	69
Abbildung 2-2-6: Handlungsrelevanzmatrix auf Basis subjektiver und objektiver Erfüllung.....	70
Abbildung 2-3-1: Möglicher konzeptioneller Rahmen eines Return on Mystery Shopping.....	74
Abbildung 2-3-2: Schematische Darstellung der Transaktionskostentheorie	76
Abbildung 2-3-3: Kostenwirkungen c.p. zunehmender Ausprägungen der Transaktionsmerkmale.....	78
Abbildung 2-3-4: Kostenwirkungen c.p. zunehmender Ausprägungen der Transaktionsmerkmale.....	79
Abbildung 2-3-5: Ausgewählte Schulungsmaßnahmen und Schulungsgrad bei Mystery Shopping	88
Abbildung 3-1-1: Cronbachs Aspekte der Beurteilungsakkuratheit	97
Abbildung 3-2-1: GAP-Modell der Dienstleistungsqualität	106
Abbildung 3-2-2: Ergebnisvarianten des SERVQUAL-Vergleichsprozesses	108
Abbildung 3-2-3: Service Blueprint der Line of Visibility eines Fast Food-Restaurantbesuchs	112
Abbildung 3-3-1: Pfaddiagramm eines Strukturgleichungsmodells	115
Abbildung 3-3-2: Komponenten der Ergebnisqualität von Mystery Shoppern	121
Abbildung 3-3-3: Arbeitsqualität und Schulungsbedarf	122
Abbildung 3-3-4: Orte der Testdurchführung	124
Abbildung 4-1-1: Verteilung der paarweise durchgeführten Testkäufe auf die Kalenderwochen 2008.....	128
Abbildung 4-1-2: Verteilung der paarweise durchgeführten Testkäufe auf Tageszeiten.....	129
Abbildung 4-2-1-1: Komponente Mittelwertstendenz bezüglich Geschlecht.....	131
Abbildung 4-2-1-2: Komponente Mittelwertstendenz bezüglich Altersklassen	132
Abbildung 4-2-1-3: Komponente Mittelwertstendenz bezüglich des höchsten Bildungsabschlusses	135
Abbildung 4-2-1-4: Komponente Mittelwertstendenz bezüglich Tätigkeit	136
Abbildung 4-2-1-5: Komponente Mittelwertstendenz bezüglich Erfahrung als Mystery Shopper.....	139
Abbildung 4-2-1-6: Komponente Mittelwertstendenz bezüglich Fast Food-Affinität.....	140

Abbildung 4-2-1-7: Komponente Mittelwertstendenz bezüglich Restaurant-Affinität.....	141
Abbildung 4-2-1-8: Komponente Mittelwertstendenz bezüglich Häufigkeit eines Fast Food-Restaurantbesuchs	142
Abbildung 4-2-1-9: Komponente Mittelwertstendenz bezüglich spezifischer Restaurant-Besuchshäufigkeit ..	143
Abbildung 4-2-2-1: Durchschnittliche Standardabweichung (Geschlecht)	145
Abbildung 4-2-2-2: Durchschnittliche Standardabweichung (Altersklassen).....	146
Abbildung 4-2-2-3: Durchschnittliche Standardabweichung (höchster Bildungsabschluss).....	147
Abbildung 4-2-2-4: Durchschnittliche Standardabweichung (Tätigkeit).....	148
Abbildung 4-2-2-5: Durchschnittliche Standardabweichung (Erfahrung als Mystery Shopper)	149
Abbildung 4-2-2-6: Durchschnittliche Standardabweichung (Fast Food-Affinität)	150
Abbildung 4-2-2-7: Durchschnittliche Standardabweichung (Restaurant-Affinität)	151
Abbildung 4-2-2-8: Durchschnittliche Standardabweichung (Häufigkeit eines Fast Food-Restaurantbesuchs) ..	152
Abbildung 4-2-2-9: Durchschnittliche Standardabweichung (Restaurant-Besuchshäufigkeit)	153
Abbildung 4-2-4-1: Relativer Anteil zuverlässig beurteilter Items.....	183
Abbildung A-1: Exemplarischer grafischer Vergleich von Mittelwertprofilen	216
Abbildung A-2: Exemplarischer grafischer Vergleich von aggregierten Variationskoeffizienten	219

Tabellenverzeichnis

Tabelle 1-1: Empirische Studien zur Mystery Shopping-Methodik.....	6
Tabelle 2-1-1: Ausgewählte Begriffe für das Verfahren Mystery Shopping in der Literatur.....	10
Tabelle 2-1-2: Vor- und Nachteile ausgewählter Methoden zur Datengewinnung.....	21
Tabelle 2-1-3: Klassifizierung von Expertenbeobachtung und Mystery Shopping.....	24
Tabelle 2-1-4: Kundenbefragung vs. Mystery Shopping.....	26
Tabelle 2-1-5: Wissenschaftliche Beiträge.....	35
Tabelle 2-2-1: Einfluss kognitiver Prozesse auf Reliabilität und Validität.....	62
Tabelle 2-2-2: Anforderungen an und Voraussetzungen für den erfolgreichen Einsatz von Mystery Shoppern.....	67
Tabelle 2-3-1: Konstellationen der Kosteneinflussfaktoren.....	81
Tabelle 2-3-2: Relative Vorteilhaftigkeit bei unterschiedlichen Konstellationen.....	82
Tabelle 2-3-3: Forschungshypothesen.....	91
Tabelle 3-1-1: Urteilstendenzen.....	95
Tabelle 3-1-2: Komponenten zur Abbildung der Ergebnisqualität von Mystery Shoppern.....	101
Tabelle 3-2-1: Festlegung der soziodemografischen Merkmale und Ausprägungen der Mystery Shopper.....	102
Tabelle 3-2-2: Festlegung der Erfahrung und einstellungsbezogener Merkmale der Mystery Shopper.....	103
Tabelle 3-2-3: Operationalisierung der SERVPERF-Dimensionen am Beispiel von McDonald's.....	110
Tabelle 3-2-4: Operationalisierung der Kontaktpunkte am Beispiel von McDonald's.....	113
Tabelle 3-3-1: PLS vs. kovarianzbasierte Strukturgleichungsmodelle.....	116
Tabelle 3-3-2: Gütemaße zur Beurteilung von PLS-Modellen.....	117
Tabelle 3-3-3: Bewertung der Arbeitsqualität.....	123
Tabelle 3-3-4: Zusammensetzung der Stichprobe.....	126
Tabelle 3-3-5: Voraussetzungen für den erfolgreichen Einsatz von Mystery Shoppern und ihre Erfüllung.....	127
Tabelle 4-2-1-2-1: SERVPERF-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der Altersklassen.....	133
Tabelle 4-2-1-2-2: Kontaktpunkt-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der Altersklassen.....	134
Tabelle 4-2-1-4-1: SERVPERF-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der Tätigkeit.....	137
Tabelle 4-2-1-4-2: Kontaktpunkt-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der Tätigkeit.....	138
Tabelle 4-2-1-9: SERVPERF-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der spezifischen Restaurant-Besuchshäufigkeit.....	144
Tabelle 4-2-3-1-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Geschlecht).....	155
Tabelle 4-2-3-1-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Geschlecht).....	155

Tabelle 4-2-3-2-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Altersklassen)	156
Tabelle 4-2-3-2-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Altersklassen)	157
Tabelle 4-2-3-3-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Bildungsabschluss)	158
Tabelle 4-2-3-3-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Bildungsabschluss)	159
Tabelle 4-2-3-4-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Tätigkeit)	160
Tabelle 4-2-3-4-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Tätigkeit)	160
Tabelle 4-2-3-5-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Testerfahrung)	161
Tabelle 4-2-3-5-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Testerfahrung)	162
Tabelle 4-2-3-6-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Fast Food-Affinität)	163
Tabelle 4-2-3-6-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Fast Food-Affinität)	164
Tabelle 4-2-3-7-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Restaurant-Affinität)	165
Tabelle 4-2-3-7-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Restaurant-Affinität)	166
Tabelle 4-2-3-8-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Fast Food-Häufigkeit)	167
Tabelle 4-2-3-8-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Fast Food-Häufigkeit)	167
Tabelle 4-2-3-9-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Restaurant-Besuchshäufigkeit)	168
Tabelle 4-2-3-9-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Restaurant-Besuchshäufigkeit)	169
Tabelle 4-2-4-1-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Geschlecht	171
Tabelle 4-2-4-1-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Geschlecht	172
Tabelle 4-2-4-2-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Altersklasse	173
Tabelle 4-2-4-2-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Altersklasse	174
Tabelle 4-2-4-3-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Bildungsabschluss	175
Tabelle 4-2-4-1-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Bildungsabschluss	176
Tabelle 4-2-4-4-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Tätigkeit	177
Tabelle 4-2-4-4-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Tätigkeit	178
Tabelle 4-2-4-5-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Restaurant-Affinität	179
Tabelle 4-2-4-5-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Restaurant-Affinität	180
Tabelle 4-2-4-6-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Fast Food-Häufigkeit	181
Tabelle 4-2-4-6-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Fast Food-Häufigkeit	182
Tabelle 4-2-5-1-1: Komponente Gewissenhaftigkeit bezüglich Geschlecht	185
Tabelle 4-2-5-1-2-1: Komponente Gewissenhaftigkeit bezüglich Altersklasse	186
Tabelle 4-2-5-1-2-2: Korrelation Alter/Item gleich oder ungleich beantwortet	186
Tabelle 4-2-5-1-3: Komponente Gewissenhaftigkeit bezüglich höchstem Bildungsabschluss	187
Tabelle 4-2-5-1-4: Komponente Gewissenhaftigkeit bezüglich Tätigkeit	188
Tabelle 4-2-5-1-5: Komponente Gewissenhaftigkeit bezüglich Mystery Shopping-Erfahrung	188

Tabelle 4-2-5-1-6: Komponente Gewissenhaftigkeit bezüglich Fast Food-Affinität	189
Tabelle 4-2-5-1-7: Komponente Gewissenhaftigkeit bezüglich Restaurant-Affinität.....	190
Tabelle 4-2-5-1-8: Komponente Gewissenhaftigkeit bezüglich Häufigkeit eines Fast Food-Restaurantbesuchs	191
Tabelle 4-2-5-1-9: Komponente Gewissenhaftigkeit bezüglich der spezifischen Restaurant-Besuchshäufigkeit	192
Tabelle 4-2-5-2-1: Komponente Vollständigkeit bezüglich Geschlecht	193
Tabelle 4-2-5-2-1-1: Komponente Vollständigkeit bezüglich Altersklasse	194
Tabelle 4-2-5-2-2-2: Korrelation Alter/Items ausgelassen oder vollständig ausgefüllt.....	194
Tabelle 4-2-5-2-3: Komponente Vollständigkeit bezüglich höchstem Bildungsabschluss	195
Tabelle 4-2-5-2-4: Komponente Vollständigkeit bezüglich Tätigkeit.....	195
Tabelle 4-2-5-2-5: Komponente Vollständigkeit bezüglich Mystery Shopping-Erfahrung	196
Tabelle 4-2-5-2-6: Komponente Vollständigkeit bezüglich Fast Food-Affinität	197
Tabelle 4-2-5-2-7: Komponente Vollständigkeit bezüglich Restaurant-Affinität.....	198
Tabelle 4-2-5-2-8: Komponente Vollständigkeit bezüglich Häufigkeit eines Fast Food-Restaurantbesuchs	199
Tabelle 4-2-5-2-9: Komponente Vollständigkeit bezüglich der spezifischen Restaurant-Besuchshäufigkeit....	200
Tabelle 4-3-1: Übersicht der Kernaussagen der empirischen Befunde	206
Tabelle 5-1: Zusammenfassende Übersicht der Untersuchungshypothesen.....	209
Tabelle 5-2: Eignungseinstufung als Mystery Shopper.....	210

Abkürzungsverzeichnis

Abb.	Abbildung
B2C	Business to Consumer
BetrVG	Betriebsverfassungsgesetz
BVM	Bundesverband Deutscher Markt- und Sozialforscher
bzw.	beziehungsweise
c.p.	ceteris paribus
et al.	et alii
etc.	et cetera
f.	folgende
ff.	fort folgende
Hrsg.	Herausgeber
GfK	Gesellschaft für Konsumforschung
Jg.	Jahrgang
MSPA	Mystery Shopping Providers Association
S.	Seite
SERVPERF	Service Performance
SERVQUAL	Service Quality
SPSS	Statistical Product and Service Solution
Tab.	Tabelle
u.a.	unter anderem
usw.	und so weiter
vgl.	vergleiche
vs.	versus
WSS	Wilmark Service System, Inc.
z.B.	zum Beispiel

1 Einführung

1.1 Problemstellung

Unter *Mystery Shopping* werden im allgemeinen Verfahren zur Erhebung von Dienstleistungsqualität subsumiert, bei denen geschulte Beobachter als *normale* Kunden auftreten und *reale* Kundensituationen wahrnehmen. Das Dienstleistungsgeschehen wird dabei nach einem zuvor festgelegten Kriterienkatalog bewertet. Nicht die subjektive Wahrnehmung, sondern eine möglichst objektive Beurteilung von Qualitätsaspekten ist zentraler Gegenstand des Verfahrens.¹

Im Gegensatz zu den wenigen wissenschaftlichen Beiträgen zur Mystery Shopping-Methodik² als Instrument zur Erhebung der Dienstleistungsqualität existiert eine Vielzahl an praktischen Publikationen³, die in einem fragmentierten, theorielosen Entwicklungsprozess einen Wissenszuwachs durch Versuch und Irrtum anstreben.⁴ Das Wissen über ein methodisches Vorgehen bei der Konzeption einer Mystery Shopping-Untersuchung ist daher nur als wenig befriedigend einzustufen. Insofern stellt sich die Frage: Wie ist eine Mystery Shopping-Untersuchung zu konzipieren, damit Dienstleistungsqualität erfolgreich erhoben werden kann?

Während vergangene Forschungsbemühungen sich eingehend dem Nachweis der Gütekriterien Reliabilität und Validität von Mystery Shopping⁵ gewidmet haben, blieb eine Betrachtung der als Mystery Shopper eingesetzten Personen aus. Unternehmen, die Mystery Shopping einsetzen, bedienen sich in der Regel Agenturen, die ihren Testerstab über das Internet rekrutieren.⁶ Häufig sind diesen Agenturen lediglich soziodemografische Merkmale der registrierten Testkunden bekannt. Die Qualität der Arbeit und die Güte der Ergebnisse können erst im Nachhinein beurteilt werden. Basierend auf diesen Überlegungen stellt sich daher die Frage: Wie unterscheiden sich Mystery Shopper aus unterschiedlichen soziodemografischen Gruppen in der Qualität der abgelieferten Ergebnisse?

¹ Vgl. Matzler/Kittinger-Rosanelli (2000: 222), Bruhn (2006: 87)

² Vgl. z.B. die Beiträge von Wilson (1998: 148), Deckers (1999), Finn/Kayandé (1999), Dawes/Sharp (2000), Matzler/Kittinger-Rosanelli (2000: 221), Haas (2003), Lingenfelder/Schmidt/Wieseke (2005: 183), Schmidt (2007: 3)

³ z.B. die Beiträge von Drees/Schiller (2001), Beckmann (2003), Beister (2003), Guber/Kerper (2003), Pergandé/Stücken (2004), Böhm/Fischl/Rennhak (2006), Dobbstein/Windbacher (2007), May (2007)

⁴ Vgl. hierzu auch Schmidt (2007: 6f.)

⁵ Vgl. Finn/Kayandé (1999), Dawes/Sharp (2000), Schmidt (2007)

⁶ z.B. MSM Germany (<http://www.msmiq.net/bewerbung>; Stand 2.6.2008), BARE International (<https://www.sassieshop.com/sassie/SassieShopperSignup/Signup.php>; Stand 2.6.2008)

Im folgenden Abschnitt soll die Relevanz dieser Fragestellungen verdeutlicht werden. Dafür wird zunächst die Bedeutung der Dienstleistungsqualität für den Unternehmenserfolg erläutert und Defizite zur traditionell eingesetzten Kundenbefragung aufgezeigt. Dann wird der Stellenwert von Mystery Shopping als Instrument zur Erhebung von Dienstleistungsqualität in Theorie und Praxis dargelegt. Der letzte Abschnitt detailliert die Zielsetzung der Arbeit und schildert den Gang der Untersuchung

1.2 Relevanz des Themas

Die Relevanz der Fragestellungen der vorliegenden Arbeit ergibt sich einerseits aus Defiziten traditioneller Ansätze zur Erhebung der Dienstleistungsqualität und andererseits aus der in den vergangenen Jahren zunehmend gewachsenen praktischen Bedeutung von Mystery Shopping, der seitens der Forschung nur wenig Beachtung geschenkt wurde.

1.2.1 Bedeutung der Dienstleistungsqualität für den Unternehmenserfolg und Defizite von Kundenbefragungen

In Theorie und Praxis herrscht weitgehend Einigkeit über den Einfluss der Qualität des Leistungsangebots und der Kundenzufriedenheit auf den nachhaltigen Unternehmenserfolg.⁷ Die Wirkungskette Dienstleistungsqualität, Kundenzufriedenheit, Kundenbindung und Unternehmenserfolg wurde in zahlreichen Studien empirisch bestätigt.⁸ Besondere Bedeutung gewinnt dieser Zusammenhang vor dem Hintergrund, dass das Ertragspotenzial eines Kunden erst langfristig herangebildet wird und ausgeschöpft werden kann. Zudem sind die Kosten für die Akquisition eines Neukunden fünf- bis sechsmal höher als die Pflege eines Bestandskunden.⁹ Daher ist es für einen Anbieter von Dienstleistungen von Bedeutung, die Qualität der erbrachten Leistungen regelmäßig zu erheben und zu überprüfen.

Es existiert eine Vielzahl wissenschaftlicher Beiträge, die sich mit der Operationalisierung des Dienstleistungsqualitätskonstrukts und der Güteprüfung der Ansätze auseinandersetzen.¹⁰ Dabei wird die Dienstleistungsqualität mit schriftlichen oder mündlichen Kundenbefragungen erhoben. Es lassen sich jedoch gravierende Nachteile beim Einsatz der Kundenbefragungstechniken zur Erhebung der Dienstleistungsqualität anführen:

- Zwischen tatsächlicher und beschriebener bzw. erlebter Situation wird häufig eine Abweichung festgestellt. Dieses als „Bias“ bezeichnete Phänomen kann z.B. auf

⁷ Vgl. Matzler/Pechlaner/Kohl (2000: 157), Lingenfelder/Wieseke/Schmidt (2003: 284)

⁸ für einen Überblick siehe z.B. Homburg/Bucerius (2006: 64ff.)

⁹ Vgl. Rust/Zahorik (1993), Siefke (1998: 4)

¹⁰ Vgl. z.B. Voss/Aleda/Roth/Rosenzweig/Richard (2004: 212ff.), Schmidt (2007: 2)

Vergessen, falsche Angaben wegen angenommener sozialer Erwünschtheit oder kognitive Dissonanzen zurückzuführen sein.¹¹

- Die Kosten sind hoch und die Ergebnisse häufig wenig aussagekräftig und detailliert.¹²
- Detaillierte Berichte, die möglichst wenig durch Vergessen verzerrt werden, erfordern eine zeitnahe Befragung, die bei einer schriftlichen Befragung nur schwer erfolgen kann. Eine Befragung am Point of Sale erfordert hingegen einen erhöhten Zeitaufwand und geht mit hohen Kosten einher. Mündliche Befragungen lassen häufig nur eingeschränkt eine Informationsgewinnung zu, da Auskunftspersonen sprachlich begrenzt sein können.¹³
- Denkbar ist auch eine Einschränkung der Beurteilungsfähigkeit der Dienstleistungsqualität durch den Kunden, z.B. durch ein langfristiges Unternehmensimage, Kontakt zu Mitarbeitern sowie vergangene Erfahrungen mit dem Unternehmen.
- Ferner könnte sich während des Erhebungszeitraums das Verhalten der Mitarbeiter aufgrund der angekündigten Maßnahme ändern (Beobachtungseffekt), so dass es nicht repräsentativ ist.
- Schließlich können Kundenbefragungen zur Erhebung der Dienstleistungsqualität in ihrer Interpretierbarkeit durch Deckeneffekte und Anspruchsinflation eingeschränkt werden.¹⁴

1.2.2 Stellenwert von Mystery Shopping als Alternative zur Evaluation der Dienstleistungsqualität

Vor dem Hintergrund der geschilderten Kritik an Kundenbefragungen verwundert es nicht, dass in der Praxis andere Instrumente eingesetzt werden. Mystery Shopping hat sich in den vergangenen zehn Jahren als leistungsfähiges Instrument zur Analyse von Dienstleistungen herausgestellt.¹⁵ Auch in der Praxis erfreut sich das Instrument zunehmender Beliebtheit. Es wird branchenübergreifend überall dort eingesetzt, wo Unternehmen ihren Kunden begegnen (B2C), wie z.B. in Einzelhandelsgeschäften, Gaststätten und Hotels, Call Centern, Krankenhäusern oder Werkstätten.¹⁶ Betrachtet man die Anzahl der Mitglieder, die beim

¹¹ Vgl. Morrison/Colman/Preston (1997: 352f.), Matzler/Kittinger-Rosanelli (2000: 225), Finn (2001: 310), Wilson (2001: 722)

¹² Vgl. Wieseke/Lingenfelder/Schmidt (2006: 42ff.)

¹³ Vgl. Matzler/Kittinger-Rosanelli (2000: 225), Wilson (2001: 722), Schmidt (2007: 2)

¹⁴ Vgl. Schmidt (2007: 197)

¹⁵ Vgl. Haas (2002: 277), Lingensfelder/Schmidt/Wieseke (2005: 183)

¹⁶ Vgl. Deckers (1999: 12f.)

Bundesverband Deutscher Markt- und Sozialforscher (BVM) als Anbieter von Mystery Shopping-Dienstleistungen registriert sind, so ist zwischen 1997 und 2007 ein Anstieg von drei auf 20 Unternehmen zu verzeichnen.¹⁷ Während das deutsche Marktvolumen für Mystery Shopping 2003 auf 20-30 Mio. € geschätzt wurde, wurde es 2005 auf einer Bandbreite von 30-50 Mio. € eingestuft.¹⁸ Der europäische Verband, die Mystery Shopping Providers Association (MSPA), geht für Deutschland von einem jährlichen durchschnittlichen Wachstum von 8% bis 2010 aus.¹⁹ Das Umsatzvolumen für Mystery Shopping-Dienstleistungen in Europa wird von der MSPA für 2005 auf 210 Mio. € und für Nordamerika auf rund 600 Mio. € geschätzt.

Trotz der Präsenz in der Praxis wurde das Thema Mystery Shopping in der wissenschaftlichen Literatur bisher nur wenig untersucht.²⁰ In der Literatur existieren einige wenige empirische Studien, die sich mit der Mystery Shopping-Methodik auseinandersetzen.²¹

Aus der Übersicht in Tabelle 1-1 wird deutlich, dass sich bisherige Forschungsbemühungen insbesondere mit Gütekriterien auseinandersetzen.²² Es ist zu konstatieren, dass Reliabilität und Validität der Methode nachgewiesen werden konnten. Ferner wurde belegt, dass sich die Methode eignet, um theoretische und wahrnehmungsrelevante Aspekte der Dienstleistungsqualität zu erfassen und zu analysieren.²³ Im Vergleich zu Kundenbefragungen hat sich Mystery Shopping hinsichtlich Ergebnisqualität, Leistungsfähigkeit und anfallender Kosten als vorteilhafter erwiesen.²⁴

Die Forderung nach Erbringung eines Nachweises von Reliabilität und Validität von Mystery Shopping-Untersuchungen ist nicht neu und wurde bereits 1997 von Morrison et al. angeregt.²⁵ Während Forschungsbemühungen dieser Anregung nachgekommen sind, blieb eine detailliertere Betrachtung aus, die sich mit einer Ergänzung der Anregung, nämlich der Untersuchung der „effects of gender, age, and other demographic factors“,²⁶ auseinandersetzt.

¹⁷ Daten für 1997 siehe Deckers (1999: 3), Daten für 2007 siehe www.bvm-net.de (Stand: 13.9.2007); Die Entwicklung ist ein Indikator, da nicht alle Anbieter beim BVM als Mitglied oder Anbieter registriert sind. So ergab eine 2005 von diesem Autor durchgeführte Internetrecherche eine Anzahl von 44 Unternehmen, die Mystery Shopping-Dienstleistungen in Deutschland anbieten.

¹⁸ Vgl. Geißler (2004), Dostert (2005)

¹⁹ Vgl. die Pressemitteilung der MSPA vom 19.5.2006

²⁰ Vgl. z.B. Wilson (1998: 148), Deckers (1999), Finn/Kayandé (1999), Dawes/Sharp (2000), Matzler/Kittinger-Rosanelli (2000: 221), Haas (2003), Lingenfelder/Schmidt/Wieseke (2005: 183), Schmidt (2007: 3)

²¹ Eine Übersicht empirischer Beiträge bietet die Tabelle A-2-1

²² nämlich Finn/Kayandé (1999), Dawes/Sharp (2000), Haas (2003) und Schmidt (2007)

²³ Vgl. Semel (2006: 98)

²⁴ Vgl. Finn/Kayandé (1999), Schmidt (2007)

²⁵ Vgl. Morrison/Colman/Preston (1997: 359)

²⁶ Morrison/Colman/Preston (1997: 359)

Begründet wird diese Anregung durch unterschiedliche Erkenntnisse, die in der Dienstleistungsforschung dokumentiert sind.²⁷

Beispielsweise wurde in Warenhäusern festgestellt, dass Männern eine höhere Servicepriorität als Frauen zukommt.²⁸ In einer Servicestudie in Restaurants wurde ermittelt, dass weibliche und männliche Servicekräfte Männer lieber bedienen als Frauen, in der Summe aber weibliche Servicekräfte besser als ihre männlichen Kollegen bewertet werden.²⁹ Dies lässt die Vermutung zu, dass weibliche und männliche Mystery Shopper in derselben Dienstleistungssituation unterschiedliche Wahrnehmungen haben und diese möglicherweise unterschiedlich interpretieren und bewerten. Es ist darüber hinaus denkbar, dass sich neben geschlechtsspezifischen Eigenschaften weitere soziodemografische Faktoren von Mystery Shoppern auf die Güte der Ergebnisse auswirken und sich spezifische Merkmalsausprägungen möglicherweise als vorteilhaft erweisen könnten. Für Unternehmen, die Mystery Shopper einsetzen, würde die Klärung Aufschluss für die Zusammensetzung einer geeigneten Feldmannschaft geben.³⁰

²⁷ Vgl. hierzu und nachfolgend Morrison/Colman/Preston (1997: 355f.)

²⁸ Vgl. Zinkhan/Stoiadin (1984) und Stead/Zinkhan (1986)

²⁹ Vgl. Hall (1993)

³⁰ Dies ist insbesondere vor dem bereits angesprochenen Hintergrund interessant, dass Testkunden von Mystery Shopping-Anbietern häufig über das Internet rekrutiert werden und somit lediglich soziodemografische Merkmale der Tester bekannt sind.

Tabelle 1-1: Empirische Studien zur Mystery Shopping-Methodik

Autoren	Branche	Untersuchungsschwerpunkt	Charakteristika der Empirie	Empirische Befunde
Finn, Kayandé (1999)	Einzelhandel	Vergleich von Mystery Shopping und Kundenbefragung unter psychometrischen Gesichtspunkten	n=6 Mystery Shopper führten 96 Testkäufe in 16 Filialen durch	<ul style="list-style-type: none"> -Die mit Mystery Shopping erhobenen Daten erwiesen sich als reliabel -Die erhobenen Daten haben eine höhere Reliabilität als die von Kundenbefragungen -Die Reliabilität der Daten ist bei der Bewertung des Ladengeschäfts (etwa „hard facts“) höher als die bei der Bewertung von Serviceaspekten (etwa „soft facts“) -Reliable Beurteilungen, die mit Hilfe von Mystery Shopping generiert werden, sind kostengünstiger als die von Kundenbefragungen zu erheben, selbst wenn 10 mal höhere Kosten pro Datensatz angenommen werden -Zur Beurteilung des Ladengeschäfts sind 12 Datensätze, für die Beurteilung von Servicequalität 40-60 Datensätze erforderlich
Dawes, Sharp (2000)	Einzelhandel	Reliabilität und Validität von Mystery Shopping-Erhebungen	n=60 Mystery Shopper 450 Filialen wurden je drei Mal von unterschiedlichen Paaren besucht. (Wiederholung der Welle vier Mal binnen eines Jahres)	<ul style="list-style-type: none"> -Die erhobenen Daten erwiesen sich als reliabel -Kriteriums- und Konvergenzvalidität der Daten kann belegt werden -Es wird eine hohe Stabilität der Komponenten, die eine objektive Dienstleistungsqualität bestimmen, festgestellt -Es wird ein hoher Stichprobenfehler für Mystery Shopping festgestellt
Finn (2001)	Einzelhandel	Betrachtung der Verkaufsfähigkeit und des Erscheinungsbildes von Ladenlokalen bei Gebrauchsgüter-Filialisten und Einzelgeschäften im Leistungsvergleich mit dem Wettbewerb	n=13 Mystery Shopper 17 Filialen Anzahl der Testfälle nicht bekannt	<ul style="list-style-type: none"> -Die Methode ist wirtschaftlicher bei Einzelgeschäften als bei Filialisten umzusetzen -Testen von Verkaufsfähigkeit benötigt doppelt so viele Tests wie die des Erscheinungsbildes bei Einzelgeschäften und fünf mal so viele bei Filialisten -Im Vergleich zu Finn/Kayandé (1999) wird festgestellt, dass zur Beurteilung der Servicequalität für Konsumgütergeschäfte mehr Tests benötigt werden als zur Beurteilung der Verkaufsfähigkeit für Gebrauchsgütergeschäfte -Mystery Shopping wird als „brauchbare Alternative zu Kundenbefragungen im Leistungsvergleich“ angesehen

Haas (2003)	Einzelhandel	Untersuchung der Validität der Urteile von Mystery Shoppem anhand eines Vergleichs mit Kundenbefragungen	n=60 Testverkaufsgespräche und n=202 Datensätze aus Kundenbefragungen	<ul style="list-style-type: none"> -Urteile von Mystery Shoppem lassen sich auch bei komplexeren Sachverhalten mit denen von echten Kunden vergleichen -Die auf die Beratungszufriedenheit von Kunden wirkenden Einflüsse wirken auch im Falle der Mystery Shopper -Testkunden sind in der Lage, einen deutlich höheren Teil der Varianz ihres Beratungszufriedenheitsurteils zu erklären als reale Kunden
Schmidt (2007)	Reisebüros	Reliabilität und Validität des Dienstleistungsqualitätsurteils von Mystery Shoppem	<p>Studie 1: n=16 Mystery Shopper führten paarweise in 86 Reisebüros Testberatungen durch (2 Paare/Reisebüro)</p> <p>Studie 2: n=16 Mystery Shopper führten paarweise in 71 Reisebüros Testberatungen durch (2 Paare/Reisebüro)</p> <p>Studie 3: n=21 Mystery Shopper führten paarweise in 76 Reisebüros Testberatungen durch (2 Paare/Reisebüro mit 1-wöchiger Verzögerung)</p>	<ul style="list-style-type: none"> -Die Methode erwies sich als ein reliables Instrument (Interne Konsistenz, Interrater-Reliabilität, Retest-Reliabilität). Dabei ist insbesondere zu konstatieren, dass verschiedene Mystery Shopper eine gemeinsam erlebte Dienstleistungssituation ähnlich bzw. gleich beurteilen, und dass ihr Urteil zeitliche Konsistenz besitzt. -Die Methode erwies sich als valides Instrument (Inhaltsvalidität, Konstruktvalidität bzgl. des Operationalisierungsansatzes SERVPERF und des Phasenmodells der Dienstleistungsqualität) -Die Methode erwies sich als leistungsfähiger im Vergleich zur Befragung von Kunden, welche nicht in der Lage sind, ein prognosevalides Gesamtzufriedenheitsurteil über die wahrgenommene Dienstleistungsqualität abzugeben

Quelle: Eigene Darstellung

1.3 Zielsetzung und Gang der Untersuchung

Vor dem Hintergrund der bisher nur in wenigen Beiträgen theoriegeleiteten Entwicklung zur Mystery Shopping-Thematik besteht die generelle Zielsetzung in der Konzeption einer Mystery Shopping-Methodik. In einer Felduntersuchung soll dann der Einfluss soziodemografischer Faktoren auf die Ergebnisqualität der Testkaufdaten überprüft werden. Hieraus lassen sich unterschiedliche Teilziele ableiten.

Ein *erstes Teilziel* besteht in der Darstellung der historischen Entwicklung des Phänomens Mystery Shopping in Theorie und Praxis von den Ursprüngen bis zum Status Quo. Dabei erfolgt eine Abgrenzung zu anderen Verfahren der Evaluierung der Dienstleistungsqualität sowie eine Einordnung des Verfahrens.

Das *zweite Teilziel* ist die Konzeption einer Mystery Shopping-Untersuchung, mit der Dienstleistungsqualität abgebildet werden kann. Hierbei werden Anforderungen und Voraussetzungen herausgearbeitet, damit handlungsrelevante Informationen der Dienstleistungsqualität gewonnen werden können.

Das *dritte Teilziel* liegt in der Identifikation von Determinanten der Ergebnisqualität von Mystery Shoppern.

Ein *viertes Teilziel* schließlich stellt die Ergebnisqualität von Mystery Shoppern aus unterschiedlichen soziodemografischen Gruppen gegenüber.

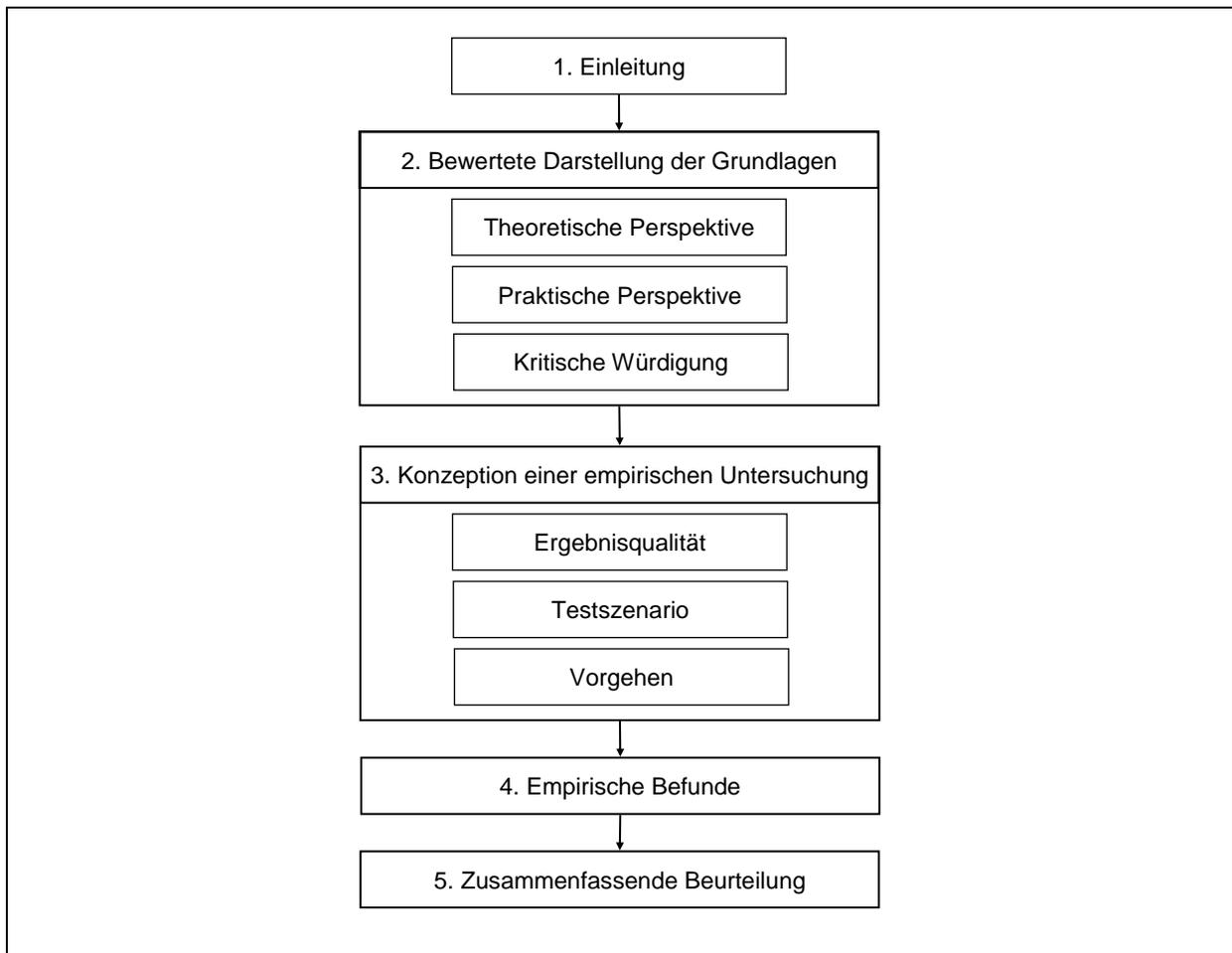
Der Gang der Untersuchung ist mit dieser Zielsetzung bereits in weiten Teilen vorgezeichnet. In Kapitel 2 wird die Methode Mystery Shopping aus praktischer und theoretischer Perspektive erläutert. Nach einer Begriffbestimmung und Definition wird das Verfahren eingeordnet. Dann wird ein Überblick der Ursprünge und Entwicklung sowie vergangener Forschungsbemühungen skizziert. Nach einem Marktüberblick werden mögliche Zielsetzungen und Einsatzfelder skizziert. Bei der Betrachtung der Konzeption einer Mystery Shopping-Untersuchung werden theoretische Erkenntnisse und praktische Beiträge um eigene Überlegungen ergänzt. Das Kapitel schließt mit einer kritischen Würdigung, in deren Rahmen die zu untersuchenden Forschungshypothesen dargelegt werden.

In Kapitel 3 wird eine empirische Untersuchung konzipiert, mit der die Forschungshypothesen überprüft werden können. Hierfür wird spezifiziert, wie die Ergebnisqualität von Mystery Shoppern mit Hilfe unterschiedlicher Komponenten beurteilt werden kann. Ferner werden ein Testszenario in der Fast Food-Branche sowie ein Beobachtungsbogen entwickelt. Schließlich wird die Erfassung der einzelnen Komponenten der Ergebnisqualität und der Gang der Felduntersuchung beschrieben.

In Kapitel 4 wird die Ergebnisqualität der unterschiedlichen untersuchten Gruppen von Mystery Shoppern dargestellt. Dies erfolgt anhand der im vorangegangenen Kapitel konzipierten Untersuchung.

Die Arbeit schließt mit dem zusammenfassenden Kapitel 5. Neben einer Synopse der Untersuchungsergebnisse werden die Limitationen dieser Arbeit aufgezeigt. Ferner werden die aus der Untersuchung resultierenden Implikationen für Unternehmen, die Mystery Shopper einsetzen, genannt. Zudem werden Hinweise für zukünftige Forschungsbemühungen gegeben. Abbildung 1-1 liefert einen Überblick des Aufbaus der Untersuchung.

Abbildung 1-1: Aufbau der Untersuchung



Quelle: Eigene Darstellung

2 Mystery Shopping: Bewertete Darstellung der Grundlagen

Ziel dieses Kapitels ist es, die Methode Mystery Shopping aus theoretischer und praktischer Perspektive zu erläutern. Bei Einnahme der theoretischen Perspektive wird zunächst auf die Terminologie und Definitionen eingegangen, gefolgt von einer Einordnung des Verfahrens. Nach einem Überblick der Ursprünge und der Entwicklung wird ein Überblick vergangener Forschungsbemühungen präsentiert. Bei Einnahme der praktischen Perspektive werden nach Darstellung eines Marktüberblicks mögliche Zielsetzungen und Einsatzfelder erläutert. Dann wird die Konzeption einer Mystery Shopping-Untersuchung betrachtet. Dabei werden theoretische Erkenntnisse und Beiträge aus der Praxis um eigene Überlegungen ergänzt. Das Kapitel schließt mit einer kritischen Würdigung und der Ableitung der im nächsten Kapitel zu untersuchenden Forschungshypothesen.

2.1 Mystery Shopping aus theoretischer Perspektive

2.1.1 Begriffsbestimmung Mystery Shopping

2.1.1.1 Terminologie

Betrachtet man im Rahmen einer Begriffsbestimmung die Literatur, so lassen sich fast 50 unterschiedliche Benennungen finden. DECKERS kategorisiert die unterschiedlichen Benennungen in vier Gruppen (siehe Tabelle 2-1-1).³¹

Tabelle 2-1-1: Ausgewählte Begriffe für das Verfahren Mystery Shopping in der Literatur

Kategorie	Bezeichnungen
<i>Testdurchführung</i>	Outside Shoppers, Professional Shoppers, Trained Customers, Anonyme Kunden, Geheimkäufer, Geheimtester, Phantom-Kunden, Scheinkäufer, Verdeckte Einkäufer, Mystery Men, Mysterious Shoppers, Mystery Shoppers, Shadow Shoppers, Silent Shoppers, Virtual Customers
<i>Methode</i>	Dummy Calls, Telefontests, Beobachtungstest, Branch Survey Shoppers, Mystery Research, Mystery Shopping Surveys, Testkauf-Interviews
<i>Einsatzgebiete</i>	Buchtests, Test-Tanken, Pseudo Patients, Hotel-Tester, Silent Guests, Ladentests, Schaltertests, Dienstleistungstests, Werkstatttests
<i>Zielsetzung</i>	Diebstahltests, Kassentests, Kontrollkäufe, Servicetests

Quelle: Eigene Darstellung in Anlehnung an DECKERS (1999: 4ff) und die dort zitierte Literatur

³¹ Vgl. Deckers (1999: 4ff.)

Die erste Kategorie „Testdurchführung“ zielt auf Aspekte der eigentlichen Tätigkeit ab, etwa die Entsendung der Tester durch ein externes Unternehmen, die professionelle Sichtweise oder die verdeckte Identität des Testers. In der Kategorie „Methode“ werden Hinweise auf die zugrunde liegende Methode geliefert, etwa Beobachtung oder Interview. Die dritte Kategorie „Einsatzgebiete“ beschreibt wiederum Branchen, in denen die Methode eingesetzt wird, etwa Tankstellen oder Hotels. Die letzte Kategorie „Zielsetzung“ betont schließlich das Ziel des Verfahrens, etwa die Kassiergenauigkeit zu prüfen, die Aufmerksamkeit des Personals bei Ladendieben zu testen oder das Dienstleistungsgeschehen zu evaluieren. Aus den unterschiedlichen Begriffen wird deutlich, dass die Anwendungsbereiche von Mystery Shopping äußerst vielfältig sind. In den folgenden Ausführungen wird Mystery Shopping insbesondere als Methode zur Evaluierung der wahrgenommenen Dienstleistungsqualität behandelt.

Trotz des breiten Spektrums werden am häufigsten die Bezeichnungen „Mystery Shopper“ und „Testkäufer“ verwendet³² bzw. „Mystery Shopping“ und „Testkäufe“ für das Verfahren. DECKERS argumentiert, dass der Begriff „Mystery Customer“ (Testkunde) treffender sei, da er nicht ausschließlich einen Kauf impliziert und auf Dienstleistungen erweitert werden kann. Spätere Veröffentlichungen in Deutschland verwenden jedoch hauptsächlich den Begriff „Mystery Shopping“³³. Gleichwohl ist die Verwendung des Begriffs „Mystery“ als unglücklich anzusehen, da bei der Übersetzung möglicherweise negativ belegte Assoziationen wie „Mysterium“ hervorgerufen werden. Im Folgenden wird dennoch weiterhin der Begriff Mystery Shopping verwendet, da er sich in der deutschen wie internationalen Literatur offenbar durchgesetzt hat.

2.1.1.2 Definitionen

Eine Vielzahl an Definitionen für Mystery Shopping ist in der Literatur zu finden, die sich inhaltlich zum Teil sehr unterscheiden. Im Folgenden wird eine Auswahl präsentiert.

KELLER setzte sich insbesondere mit in Kreditinstituten durchgeführten Schaltertests auseinander und definiert:

„Unter dem Begriff Schaltertest fassen wir all jene Testanlagen zusammen, bei denen ein speziell geschulter Tester mit einem zu testenden Bankangestellten ein für diesen nicht als außergewöhnlich erkennbares Beratungsgespräch führt und bei denen der Gesprächsverlauf und –inhalt auf eine dem Bankangestellten nicht ersichtliche Art aufgezeichnet wird.“³⁴

³² Vgl. ebenda (1999:6)

³³ Vgl. z.B. Matzler/Kittinger-Rosanelli (2000), Lingenfelder/Wieseke/Schmidt (2003 & 2005), Haas (2002), Wieseke/Schmidt/Lingenfelder (2006)

³⁴ Keller (1982: 97)

DECKERS geht in seiner Definition viel weiter als KELLER und beschreibt die von ihm als „Testkundenverfahren“ benannte Methode als

„...verdeckt teilnehmende Beobachtung, die unter Einsatz eines Gesprächsleitfadens, im Zusammenhang des Kaufens und Konsumierens und mit der Zielrichtung der Datenerhebung und Qualitätssteuerung zum Einsatz gebracht wird.“³⁵

Neben Einordnung in die Marktforschung und Benennung des methodischen Vorgehens weist er dabei insbesondere auf das Ziel des Verfahrens hin.

DREES und SCHILLER liefern ebenfalls einen ergebnisorientierten Ansatz, der insbesondere die durch Schulung angestrebte Objektivität hervorhebt:

„Mystery Shopping liefert durch den Einsatz speziell geschulter, systematisch eingesetzter und genau gebriefteter Testpersonen (Mystery Shopper), die verdeckt als ‚normale‘ Kunden Dienstleistungs- und Kaufsituationen simulieren, objektive Antworten auf Fragen zu Dienstleistungsqualität und Qualitätsmerkmalen eines Unternehmens.“³⁶

HAAS relativiert die Möglichkeit, objektive Ergebnisse zu erzielen, weist auf das methodische Vorgehen hin, unterlässt aber einen Hinweis auf die Zielsetzung:

„Unter Mystery Shopping versteht man eine verdeckt teilnehmende Beobachtung, durch die subjektiv wahrgenommene Sachverhalte möglichst objektiv erfasst werden.“³⁷

Um eine zufrieden stellende Definition abzugeben, welche die methodische Vorgehensweise, das angestrebte Ergebnis sowie Folgeprozesse beinhaltet, definiert der Autor der vorliegenden Arbeit wie folgt:

Mystery Shopping ist der Einsatz geschulter Beobachter, die als *normale* Kunden auftreten und *reale* Kundensituationen wahrnehmen, mit dem Ziel, das Dienstleistungsgeschehen nach einem zuvor festgelegten Kriterienkatalog zu bewerten, um ein objektiviertes Bild der Leistungsqualität des Unternehmens zu gewinnen, welches als Grundlage für strategische Entscheidungen und operative Maßnahmen zur Qualitätsverbesserung dient, um die objektive und wahrgenommene Dienstleistungsqualität zu verbessern.

Die Gültigkeit dieser Definition kann nur vor dem Hintergrund einer Abgrenzung zu anderen Methoden der Evaluation von Dienstleistungsgeschehen überprüft werden. Voraussetzung

³⁵ Deckers (1999: 29)

³⁶ Drees/Schiller (2003: 162)

³⁷ Haas (2002: 279)

dafür ist eine Einordnung von Mystery Shopping in die Marktforschung, das Aufzeigen der Wurzeln und der Entwicklung sowie eine Betrachtung der unterschiedlichen Zielsetzungen und den damit verbundenen Implementierungsformen.

2.1.2 Einordnung von Mystery Shopping als Verfahren zur Erhebung von Dienstleistungsqualität

Ziel des vorliegenden Abschnitts ist es, Mystery Shopping gegen korrespondierende Verfahren zur Evaluierung des Dienstleistungsgeschehens abzugrenzen und das Verfahren einzuordnen. Dafür erfolgt zunächst eine Begriffsbestimmung von Dienstleistung und Dienstleistungsqualität. Dann wird ein Überblick über die unterschiedlichen Ansätze zur Erhebung der Dienstleistungsqualität gegeben. Dabei werden insbesondere die Defizite subjektiver Erhebungsverfahren adressiert und die Vorteilhaftigkeit der Methode Mystery Shopping als Verfahren der Datengewinnung objektiver Ansätze zur Erhebung der Dienstleistungsqualität herausgearbeitet, was die Relevanz der Methode untermauern soll.

2.1.2.1 Begriffsbestimmung Dienstleistung und Dienstleistungsqualität

Damit die Leistungsfähigkeit der verschiedenen Ansätze zur Erhebung der Dienstleistungsqualität beurteilt werden kann, ist zunächst eine Betrachtung der Begriffe „Dienstleistung“ und „Dienstleistungsqualität“ zweckmäßig.

Es existiert eine Vielzahl an Beiträgen, die sich mit der Bestimmung des Begriffes „Dienstleistung“ befassen. So lassen sich Dienstleistungen durch ihre typischen Eigenschaften wie Immaterialität, Intangibilität, direkter Kontakt zwischen Anbieter und Nachfrager bzw. Notwendigkeit der Integration des externen Faktors, Unteilbarkeit, Vergänglichkeit, Standortgebundenheit sowie Individualität bzw. Variabilität charakterisieren.³⁸ Zudem lassen sich Dienstleistungen auf einer übergeordneten Ebene hinsichtlich ihrer Orientierung auf die Dimensionen Tätigkeit, Potenzial, Prozess und Ergebnis unterscheiden.³⁹ Die sehr weit gefasste und daher nur bedingt überzeugende tätigkeitsorientierte Auffassung klassifiziert jede menschliche Tätigkeit als Dienstleistung. Die potenzialorientierte Interpretation beschreibt das durch Mitarbeiter und Maschinen vorhandene Potenzial zur Dienstleistungserbringung auf der Nachfrageseite. Bei der prozessorientierten Perspektive wird die Dienstleistung als Tätigkeit mit Gleichzeitigkeit von Produktion und Verbrauch betrachtet. Die ergebnisorientierte Auffassung hat schließlich das im Dienstleistungsprozess erstellte immaterielle Gut zum Inhalt. Die Dimensionen sind dabei nicht als sich gegenseitig ausschließend, sondern im Sinne einer integrierten Sichtweise zu betrachten.⁴⁰ Demnach kann

³⁸ Vgl. Bruhn (2006: 20ff.)

³⁹ Vgl. hierzu und im Folgenden Hentschel (1992: 19ff.) Corsten (2001: 21ff.) und Bruhn (2006: 22f.)

⁴⁰ Vgl. Hentschel (1992: 21), Siefke (1998: 7)

als Verknüpfung der skizzierten Eigenschaften und Dimensionen folgende Definition zu Grunde gelegt werden:

„Dienstleistungen sind selbständige, marktfähige Leistungen, die mit der Bereitstellung und/oder dem Einsatz von Leistungsfähigkeiten verbunden sind (Potenzialorientierung). Interne und externe Faktoren werden im Rahmen des Leistungserstellungsprozesses kombiniert (Prozessorientierung). Die Faktorkombination des Dienstleistungsanbieters wird mit dem Ziel eingesetzt, an den externen Faktoren – Menschen oder deren Objekten – nutzenstiftende Wirkungen zu erzielen (Ergebnisorientierung).“⁴¹

Ähnlich wie bei der Begriffsbestimmung der Dienstleistung sind die Beiträge zur Bestimmung des Begriffes „Dienstleistungsqualität“ vielfältig und zum Teil sehr unterschiedlich.⁴² So definiert die Deutsche Gesellschaft für Qualität e.V. Qualität als „die Gesamtheit von Eigenschaften und Merkmalen eines Produktes oder einer Tätigkeit, die sich auf deren Eignung zur Erfüllung gegebener Erfordernisse beziehen.“⁴³ In der vorliegenden Arbeit soll für den Begriff „Dienstleistungsqualität“ die von BRUHN auf Basis unterschiedlicher Qualitäts- und Dienstleistungsbegriffe hergeleitete Definition verwendet werden:

„Dienstleistungsqualität ist die Fähigkeit eines Anbieters, die Beschaffenheit einer primär intangiblen und der Kundenbeteiligung bedürftigen Leistung gemäß den Kundenerwartungen auf einem bestimmten Anforderungsniveau zu erstellen. Sie bestimmt sich aus der Summe der Eigenschaften bzw. Merkmale der Dienstleistung, bestimmten Anforderungen gerecht zu werden.“⁴⁴

Demnach ist das Urteil über die Dienstleistungsqualität das Ergebnis eines Vergleichs von erwarteter und tatsächlich erbrachter Dienstleistung.⁴⁵

Nach Herleitung des Dienstleistungsqualitätsbegriffs widmet sich der folgende Abschnitt der Betrachtung der unterschiedlichen Ansätze zur Erhebung der Dienstleistungsqualität.

2.1.2.2 Überblick der Verfahren zur Erhebung der Dienstleistungsqualität

Damit Dienstleistungsqualität im Sinne der Umsetzung eines Qualitätsmanagementkreislaufes geplant, gelenkt, geprüft und dargelegt werden kann, bedarf es zunächst ihrer Erhebung. Die dafür einsetzbaren Verfahren lassen sich in kunden- und unternehmensorientierte Ansätze

⁴¹ Meffert/Bruhn (2006: 30)

⁴² Vgl. Bruhn (2006: 33)

⁴³ Deutsche Gesellschaft für Qualität e.V. (1995: 30)

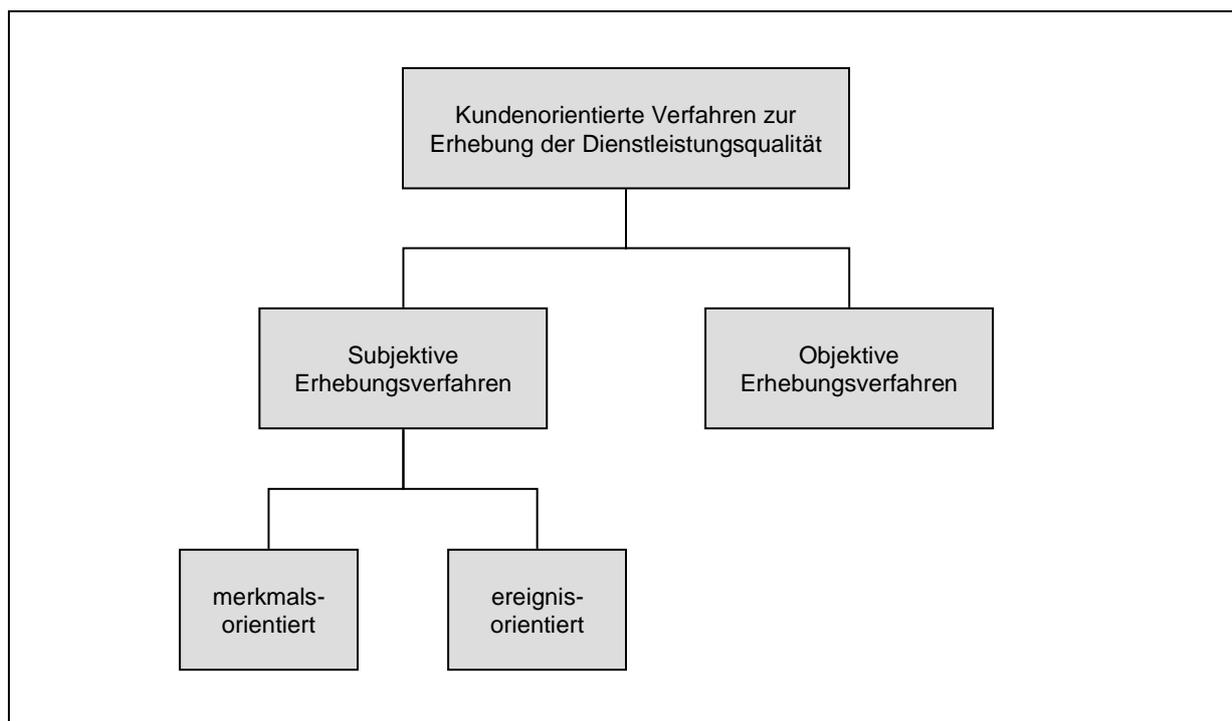
⁴⁴ Bruhn (2000: 29)

⁴⁵ Vgl. Corsten (2001: 299), Bruhn (2006: 41)

aufteilen.⁴⁶ In der nachfolgenden Betrachtung wird allein auf die kundenorientierten Ansätze eingegangen. Zum einen ist - wie noch zu zeigen ist – Mystery Shopping hier einzuordnen. Zum anderen bestimmt in den meisten Fällen allein das Kundenurteil die Kaufentscheidung, so dass die erbrachte Dienstleistungsqualität hauptsächlich aus Kundenperspektive zu erfolgen hat.⁴⁷

In Abbildung 2-1-1 wird die grundsätzliche Struktur der kundenorientierten Ansätze zur Erhebung der Dienstleistungsqualität veranschaulicht.

Abbildung 2-1-1: Systematisierung der kundenorientierten Verfahren zur Erhebung der Dienstleistungsqualität



Quelle: In Anlehnung an Bruhn (2006: 84)

Wie der Abbildung zu entnehmen ist, lassen sich die kundenorientierten Erhebungsverfahren in subjektive und objektive Ansätze unterteilen.⁴⁸ Bei den subjektiven Verfahren ist die Evaluation der Dienstleistungsqualität aus der Perspektive einzelner Kunden Kern der Betrachtung. Diese lassen sich weiterhin in merkmalsorientierte und ereignisorientierte Verfahren unterteilen.⁴⁹ Mit merkmalsorientierten Verfahren wird die globale Dienstleistungsqualität anhand einzelner Merkmale beurteilt; dies erfolgt häufig mittels strukturierter schriftlicher Kundenbefragungen. Die ereignisorientierten Ansätze

⁴⁶ Vgl. hierzu und im Folgenden Bruhn (2006: 83ff.)

⁴⁷ Vgl. Landgraf (1995: 18), Schmidt (2007: 75)

⁴⁸ Vgl. hierzu und im Folgenden Bruhn (2006: 83ff.)

⁴⁹ Daneben wird auch von problemorientierten Verfahren berichtet, bei denen qualitätsrelevante Problembereiche der Leistungserstellung aus Kundensicht Kern der Analyse sind. Vgl. hierzu Bruhn (2006: 126ff.)

konzentrieren sich auf die einzelnen Prozesse einer Kundenkontaktsituation. Dabei schildern die Kunden ihre Erlebnisse unstrukturiert in freier Rede. Bei den objektiven Verfahren erfolgt im Gegensatz zu den subjektiven Verfahren eine objektivierte Evaluation, d.h. dass diese zumindest zum Teil intersubjektiv nachprüfbar ist.⁵⁰ Nachfolgend werden die Verfahren näher erläutert und bezüglich ihrer Eignung zur Erhebung der Dienstleistungsqualität beurteilt.

2.1.2.3 Subjektive Verfahren zur Erhebung der Dienstleistungsqualität

In diesem Abschnitt werden zunächst kurz merkmals- und ereignisorientierte Verfahren erläutert. Dabei wird insbesondere auf multiattributive Verfahren eingegangen. Dann werden Methoden zur Datengewinnung, die mit subjektiven Verfahren korrespondieren, skizziert. Dabei werden Defizite subjektiver Messansätze herausgearbeitet.

Ausgangspunkt *merkmalsorientierter Verfahren* ist die Annahme, dass sich das globale Qualitätsurteil aus einzelnen Teilqualitätsurteilen zusammensetzt. Dabei sind differenzierte und undifferenzierte Herangehensweisen zu unterscheiden. Erstere fragen Teilqualitäten ab, die dann zu einem globalen Qualitätsurteil aggregiert werden (z.B. multiattributive Verfahren, Willingness-to-Pay-Ansatz, Penalty-Reward-Faktoren-Ansatz).⁵¹ Letztere fragen umgekehrt globale Qualitätsurteile ab und schließen dann auf Teilqualitäten (z.B. dekompositionelle Verfahren).⁵² Bei *ereignisorientierten Verfahren* werden Ereignisse aus der Kundenkontaktsituation zur Analyse herangezogen.⁵³ Dabei kann der Dienstleistungsprozess zur Analyse in seine einzelnen Teilschritte zerlegt werden, die in einer offenen Befragung sukzessive von der Auskunftsperson beschrieben werden (sequenzielle Ereignismethode). Ferner können kritische Ereignisse in strukturierten Interviews mit offenen Fragen ermittelt werden (Critical Incident- Technik).⁵⁴ Nachfolgend sollen im Rahmen der Betrachtung subjektiver Verfahren aus pragmatischen Überlegungen heraus mehrdimensionale multiattributive Verfahren herausgegriffen werden. Diese sind am weitesten verbreitet⁵⁵ und sind aufgrund ihrer standardisierbaren, strukturierten und schriftlichen Erhebungsform als Alternative zu einer Mystery Shopping-Erhebung am ehesten mit dieser vergleichbar.⁵⁶

⁵⁰ Vgl. ebenda (2006: 85)

⁵¹ Vgl. Hentschel (1992: 114) und für einen detaillierten Überblick Bruhn (2006: 90)

⁵² Vgl. Bruhn (2006: 99)

⁵³ Vgl. ebenda (2006: 114ff.)

⁵⁴ zu einer detaillierten Betrachtung und zu weiteren ereignisorientierten Verfahren siehe Bruhn (2006: 114ff.)

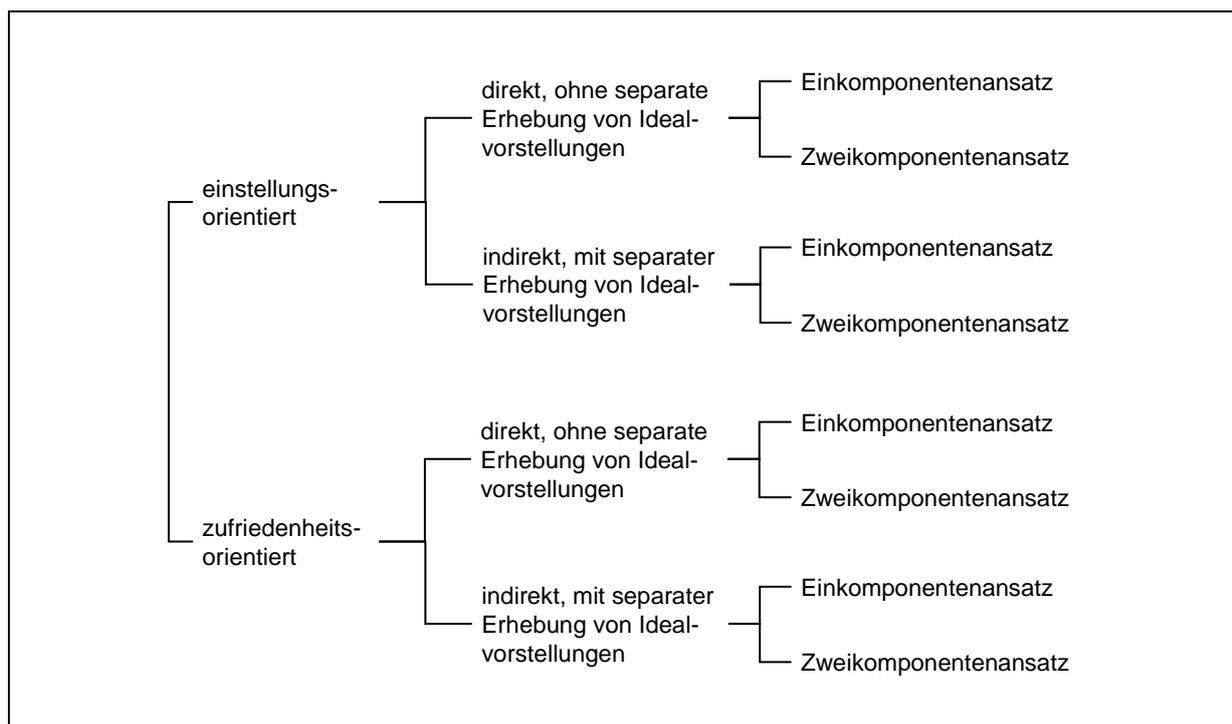
⁵⁵ Vgl. Beutin (2006: 129)

⁵⁶ So widmen sich praktische und theoretische Beiträge zum Teil der Diskussion um die Vorteilhaftigkeit von Kundenbefragung und Mystery Shopping, z.B. Deckers (1999: 242ff.), Finn/Kayandé (1999: 212), Haas (2003: 27ff.), Stücken (2003)

Folglich ist zur Beurteilung einer Vorteilhaftigkeit der Methode Mystery Shopping insbesondere die Kenntnis der Defizite dieser Verfahren relevant.

Multiattributive Verfahren lassen sich hinsichtlich der Perspektive ihrer Messung grundsätzlich in zwei Varianten aufteilen: Einstellungsorientierte und zufriedenheitsorientierte Ansätze.⁵⁷ Bei einstellungsorientierter multiattributiver Qualitätsmessung wird unterstellt, dass die Einschätzung des Kunden durch gelernte, dauerhafte, positive oder negative innere Haltung gegenüber einem Beurteilungsobjekt aufzufassen ist.⁵⁸ Während die Einschätzung dabei auch aus von anderen Kunden kommunizierten Erfahrungen resultieren kann, setzt die zufriedenheitsorientierte Variante hingegen eine konkrete, selbst erlebte Dienstleistungssituation voraus, da das Urteil auf der Diskrepanz zwischen erwarteter und wahrgenommener Dienstleistungsqualität beruht.⁵⁹ Über diese grundsätzliche Differenzierung hinaus, lassen sich die Ansätze weiter in direkte und indirekte sowie hinsichtlich ihrer Komponenten in Ein- und Zweikomponentenansätze unterteilen, wie in Abbildung 2-1-2 grafisch dargestellt ist.⁶⁰

Abbildung 2-1-2: Hauptvarianten multiattributiver Messung von Dienstleistungsqualität



Quelle: Hentschel (1992: 115)

⁵⁷ Vgl. hierzu und im Folgenden Hentschel (1992: 115ff.)

⁵⁸ Vgl. Trommersdorff (2004: 158ff.)

⁵⁹ Vgl. hierzu insbesondere das Expectancy-Disconfirmation-Paradigma, welches auf Oliver (1980) zurückgeht.

⁶⁰ Vgl. hierzu und im Folgenden Hentschel (1992: 115ff.)

Bei direkten Ansätzen wird lediglich der Eindruck abgefragt, während die Abfrage bei indirekten Ansätzen um die Idealvorstellung oder die erwartete Ausprägung ergänzt wird.⁶¹ Bei Zweikomponentenansätzen erfolgt im Gegensatz zu Einkomponentenansätzen zudem eine Abfrage der Wichtigkeit der Ausprägungen. Es wird deutlich, dass mit einer indirekten Messung und insbesondere mit den Zweikomponentenansätzen ein Informationsgewinn einhergeht. Allerdings gibt es auch Einschränkungen: Es besteht bei den Zweikomponentenansätzen die Gefahr der Anspruchsinflation, so dass alle Items als wichtig bewertet werden. Dieser Gefahr könnte entgegengewirkt werden, indem nur ein begrenztes Verteilungsbudget zur Bestimmung der Wichtigkeit zur Verfügung gestellt wird.⁶² Gleichwohl ist bereits ab einer geringen Anzahl von Items eine adäquate Zuordnung nur schwierig durchführbar. Für die indirekten Ansätze wurde festgestellt, dass die zweimalige Verwendung der gleichen Messskala zu konsistenten Antworten führt und somit die Diskrepanz zwischen Erwartungs- und Wahrnehmungswert als Maß für die Dienstleistungsqualität erheblich von der tatsächlich empfundenen Dienstleistungsqualität abweicht.⁶³ Zudem wird angeführt, dass die Diskrepanz zwischen Erwartungs- und Wahrnehmungswert nicht auf die Differenz im Sinne einer mathematischen Subtraktion, sondern auf die vom Kunden empfundene Diskrepanz zurückzuführen sein sollte, da alleinig die Betrachtung der Differenz irreführend sein kann.⁶⁴ Daneben besteht bei indirekten Ansätzen und Zweikomponentenansätzen die Gefahr, dass bei den Befragten aufgrund des Erhebungsaufwands Ermüdungserscheinungen auftreten können.⁶⁵ Ferner konnte für die Zweikomponentenansätze empirisch nachgewiesen werden, dass die direkte und ungewichtete Erhebung von Wahrnehmungswerten die höchste Korrelation mit dem Globalurteil innehat.⁶⁶ Schließlich ist der mit einer „doppelten“ Erhebung verbundene entstehende Aufwand anzuführen, welcher in höheren Kosten oder aufgrund des doppelten Umfanges in einem weniger umfangreichen Fragebogen resultieren kann.⁶⁷ Demnach erscheint die alleinige Messung von Wahrnehmungswerten vorteilhafter gegenüber Evaluationen, die Diskrepanzen aus Erwartungs- und Wahrnehmungswerten oder zusätzlichen entsprechenden Wichtigkeitswerten umfassen. In diesem Zusammenhang wurde festgestellt, dass bei einem solchen Ansatz eine implizite Gewichtung der einzelnen Qualitätsmerkmale erfolgt, die

⁶¹ Die Wahrnehmungswerte werden bei direkten und indirekten zufriedenheits- und einstellungsorientierten Ansätzen a posteriori abgefragt, während bei den indirekten zufriedenheitsorientierten Ansätzen zusätzlich eine Abfrage der Erwartungswerte a priori erfolgt.

⁶² Vgl. Schütze (1992: 176)

⁶³ Vgl. Homburg/Rudolph (1995: 44)

⁶⁴ Vgl. Hentschel (1992: 140)

⁶⁵ Vgl. Siefke (1998: 112)

⁶⁶ Vgl. Cronin/Taylor (1992: 66ff.)

⁶⁷ So geht Beutin von einem Anstieg der Gesamtkosten um 20-50% bei Kundenzufriedenheitsbefragungen aus, die statt eines Einkomponentenansatzes einen Zweikomponentenansatz verwenden. Vgl. Beutin (2006: 131)

einschließlich der Wirkung auf das globale Qualitätsurteil durch den Einsatz geeigneter statistischer Verfahren identifiziert werden kann.⁶⁸ Vor dem Hintergrund der dargelegten Argumente ist von einer Vorteilhaftigkeit der direkten Einkomponentenansätze auszugehen. So überrascht es nicht, dass in der Praxis schätzungsweise 95% der Befragungen derart konzipiert sind.⁶⁹

Es gibt unterschiedliche *Methoden zur Datengewinnung*, die mit subjektiven Ansätzen zur Evaluation der Dienstleistungsqualität korrespondieren. Grundsätzlich lassen sich die Methoden zur Datengewinnung in Sekundärforschung und Primärforschung unterteilen. Während im Rahmen ersterer auf verfügbare Informationen zurückgegriffen wird, werden bei letzterer gezielt Informationen erhoben.⁷⁰ Dafür kann bei den subjektiven Ansätzen eine mündliche oder schriftliche Form gewählt werden. Die mündliche Form, die Befragung oder das Interview, lässt sich in eine explorative und eine standardisierte Form unterteilen.⁷¹ Das explorative Interview ist weitgehend unstandardisiert, wobei der Interviewer den Verlauf mit selbst formulierten Fragen und freien Antwortmöglichkeiten steuern kann. Dabei kann zur Orientierung ein Interview-Leitfaden verwendet werden. Ziel kann es sein, schwierige Themenkomplexe zu besprechen oder vertiefende Einblicke in bestimmte Sachverhalte zu gewinnen, die dann für nachfolgende repräsentative, quantitative Erhebungen verwendet werden können.⁷² Während die Menge der gewonnenen Informationen als vorteilhaft anzusehen ist, birgt sie gleichzeitig höhere Anforderungen an Dokumentation, Kodierung, Auswertung und Interpretation.⁷³ Ferner besteht die Gefahr, dass der Befragte durch den Interviewer beeinflusst wird.⁷⁴ Bei der standardisierten Form des mündlichen Interviews wird für alle Befragten der gleiche Fragebogen verwendet, den der Interviewer im Gesprächsverlauf ausfüllt. So können die einzelnen Aussagen verglichen und reproduziert bzw. nachgeprüft werden.⁷⁵ Allerdings besteht der mögliche Nachteil, dass das Ergebnis durch die soziale Interaktion zwischen den Gesprächspartnern beeinflusst wird, indem Vorurteile zu Anpassungsmechanismen auf beiden Seiten führen.⁷⁶ Zudem können sich situative Faktoren, wie das Befragungsumfeld, der Befragungszeitpunkt oder die Anwesenheit

⁶⁸ Vgl. Homburg/Rudolph/Werner (1995: 322), die dafür den Einsatz kausalanalytischer Verfahren vorschlagen.

⁶⁹ Vgl. Beutin (2006: 131)

⁷⁰ Vgl. Bruhn (2001: 99)

⁷¹ Vgl. Homburg/Krohmer (2006: 263)

⁷² Vgl. Berekoven/Eckert/Ellenrieder (2004: 97)

⁷³ Vgl. Homburg/Krohmer (2006: 263). Gleichwohl kann den durch ein begrenztes Erinnerungsvermögen des Interviewers entstehenden Problemen durch gemischte Datenaufzeichnungstechniken wie Stichwörternotierung und Datenaufzeichnungen begegnet werden. Vgl. Berekoven/Eckert/Ellenrieder (2004: 98)

⁷⁴ Vgl. Beutin (2006: 135)

⁷⁵ Vgl. Berekoven/Eckert/Ellenrieder (2004: 101)

⁷⁶ Vgl. ebenda (2004: 106ff.)

Dritter, ergebnisverzerrend auswirken. Als ökonomisch bedeutsames Negativkriterium sind für mündliche Interviews die Kosten anzuführen, die rund 20-mal höher als bei schriftlichen Befragungen anzusetzen sind.⁷⁷ Schriftliche Befragungen erfolgen in der Regel durch postalische Zusendung eines Fragebogens, der vom Befragten ausgefüllt zurückgesendet wird. Alternativ kann die Befragung auch via Internet durchgeführt werden.⁷⁸ Die zeitliche und räumliche Ungebundenheit der Befragungssituation macht auch schwer erreichbare Gruppen zugänglich für eine Befragung. Dabei resultiert die Anonymität der Befragung in einer größeren Offenheit und in realitätsnäheren Antworten.⁷⁹ Trotz der erheblichen Vorteile der vergleichsweise niedrigen Kosten und der Erreichbarkeit, sind die Nachteile einer schriftlichen Befragung massiv: So ist in der Regel die Rücklaufquote als niedrig einzustufen, Missverständnisse können durch fehlende Interaktionsmöglichkeiten nicht geklärt werden, Erhebungen können durch kurzfristige Ereignisse bei ungleichen Erhebungsstichtagen verzerrt werden, und die Ergebnisse können durch Dritte verzerrt werden.⁸⁰ Damit eine Interpretierbarkeit und Generalisierbarkeit der Ergebnisse durch korrektes Ausfüllen und Rücksendung sichergestellt werden kann, ist die Konzeption durch geeignete Maßnahmen darauf auszurichten.⁸¹ Die genannten Vor- und Nachteile, die mit den aufgeführten Methoden der Datengewinnung einhergehen, sind in Tabelle 2-1-2 zusammengefasst.

⁷⁷ Vgl. Beutin (2006: 138)

⁷⁸ Vgl. ebenda (2006: 136f.)

⁷⁹ Vgl. Bruhn (2001: 100)

⁸⁰ Vgl. Beutin (2006: 136)

⁸¹ Vgl. Berekoven/Eckert/Ellenrieder (2004: 118ff.)

Tabelle 2-1-2: Vor- und Nachteile ausgewählter Methoden zur Datengewinnung

Methode	Vorteile	Nachteile
<i>Explorative Interview</i>	<ul style="list-style-type: none"> - Gewinnung großer Informationsmengen - Besprechung schwieriger Themenkomplexe - Tiefe Einblicke in Sachverhalte möglich 	<ul style="list-style-type: none"> - Schwierigkeiten bei der Dokumentation - Schwierigkeiten bei Kodierung, Auswertung und Interpretation - Gefahr der Beeinflussung durch Interviewer
<i>Standardisiertes Interview</i>	<ul style="list-style-type: none"> - Hoher Standardisierungsgrad ermöglicht vergleichbare und reproduzierbare Ergebnisse 	<ul style="list-style-type: none"> - Gefahr der Beeinflussung durch den Interviewer - Gefahr der Beeinflussung durch situative Faktoren - Hohe Erhebungskosten
<i>Schriftliche Befragung</i>	<ul style="list-style-type: none"> - Zugänglichkeit schwer erreichbarer Gruppen - Vergleichsweise niedrige Kosten 	<ul style="list-style-type: none"> - In der Summe niedrige Rücklaufquote - Nichtklärbarkeit von Missverständnissen - Ergebnisverzerrung durch ungleiche Erhebungszeitpunkte und/oder Dritte

Quelle: Eigene Darstellung

Neben den dargelegten allgemeinen *Defiziten der Datenerhebungstechniken* subjektiver Ansätze existieren auch solche, die auf Besonderheiten der Dienstleistung zurückzuführen sind. Zur Erläuterung sollen die drei Dienstleistungseigenschaften Intangibilität, direkte Interaktion zwischen Dienstleister und Kunde sowie prozessualer Charakter aufgegriffen werden.⁸² In Bezug auf die Intangibilität wird angenommen, dass sich die Dienstleistungsqualität vor, während oder am Ende des Prozesses nur bedingt erschließt.⁸³ Der direkte Kontakt zwischen Anbieter und Nachfrager führt dazu, dass Dienstleistungsprozesse aufgrund der Interaktion nicht vollständig detailliert planbar und standardisierbar sind und die Dienstleistungsqualität somit zu einem großen Teil aus der individuellen Interpretation einzelner Interaktionen resultiert. Der (zum Teil komplexe) prozessuale Charakter von Dienstleistungen führt dazu, dass ihre Qualität nicht anhand von spezifischen einzelnen Leistungsmerkmalen, sondern als Gesamtprozess wahrgenommen wird.⁸⁴ Die Anwendbarkeit der weiter oben aufgeführten standardisierten Verfahren zur Evaluation der Dienstleistungsqualität kann anhand der genannten Merkmale in Frage gestellt werden. Die Intangibilität erschwert im Vergleich zu Produkten die Transformation von

⁸² Vgl. hierzu und im Folgenden Matzler/Kittinger-Rosanelli (2000: 223ff.)

⁸³ Vgl. Bruhn (2006: 21)

⁸⁴ Vgl. Matzler/Kittinger-Rosanelli (2000: 223)

Erlebnissen in merkmalsorientierte Bewertungen.⁸⁵ Es wird ferner deutlich, dass die Beurteilung auf der Interpretation der beteiligten Individuen beruht, welche kaum von standardisierten Erhebungsmethoden erfasst werden kann.⁸⁶ Eine Dienstleistung ist als ein soziales Phänomen nicht objektiv zu beschreiben, sondern immer nur aus dem Kontext einer Situation zu interpretieren. So kann eine hilfsbereite Kassiererin, die einer Person beim Einpacken der Einkäufe hilft, von dieser mit einer hohen Qualität bewertet werden, während die wartenden Kunden in der Schlange möglicherweise zu einer gegenteiligen Bewertung kommen würden. Situationspezifisch widersprüchliche Qualitätseindrücke lassen sich von den oben beschriebenen standardisierten, multiattributiven Verfahren nicht ausreichend erfassen.⁸⁷ Vor dem Hintergrund des prozessualen Charakters einer Dienstleistung ist anzunehmen, dass quantitative, merkmalsorientierte Verfahren die wahrgenommene Dienstleistungsqualität nicht ausreichend erfassen, da die Dienstleistung nicht als eine Menge von einzelnen unabhängigen Merkmalen, sondern als eine Reihe von Interaktionen wahrgenommen wird.⁸⁸

Schließlich sind diejenigen Schwächen zu nennen, die den meisten Befragungstechniken zugrunde liegen. So wird zwischen tatsächlicher und beschriebener bzw. erlebter Situation häufig eine Abweichung festgestellt. Diese kann z.B. auf Vergessen, falsche Angaben wegen angenommener sozialer Erwünschtheit oder kognitive Dissonanzen zurückzuführen sein.⁸⁹ Mündliche Befragungen lassen häufig nur eingeschränkt eine Informationsgewinnung zu, da Auskunftspersonen in ihrem sprachlichen Ausdrucksvermögen verschieden sein können, was bei den intangiblen und somit schwierig zu beschreibenden Bestandteilen einer Dienstleistungen nicht verwunderlich ist.⁹⁰ Denkbar ist auch eine Einschränkung der Beurteilungsfähigkeit der Dienstleistungsqualität durch den Kunden selbst, da beispielsweise sein vorhandenes Unternehmensimage durch langfristig bestehende Kontakte zu Mitarbeitern und vergangene Erfahrungen mit dem Unternehmen geprägt ist. Ferner könnte sich während des Erhebungszeitraums das Verhalten der Mitarbeiter aufgrund der angekündigten Maßnahme ändern (Beobachtungseffekt), so dass es nicht repräsentativ ist.

2.1.2.4 Objektive Verfahren zur Erhebung der Dienstleistungsqualität

Damit die im vorangegangenen Abschnitt aufgezeigten Defizite subjektiver Ansätze zur Erhebung der Dienstleistungsqualität umgangen werden können, bieten sich unterschiedliche objektive Verfahren als Möglichkeit an. In diesem Abschnitt werden diese Verfahren sowie

⁸⁵ Vgl. Matzler/Pechlaner/Kohl (2000: 162)

⁸⁶ Vgl. mit der Kritik an quantitativen Verfahren von Girtler (1992: 26)

⁸⁷ Vgl. Matzler/Kittinger-Rosanelli (2000: 223)

⁸⁸ Vgl. Stauss/Hentschel (1992: 59ff.)

⁸⁹ Vgl. Matzler/Kittinger-Rosanelli (2000: 225), Finn (2001: 310), Wilson (2001: 722)

⁹⁰ Vgl. Matzler/Kittinger-Rosanelli (2000: 225), Wilson (2001: 722), Schmidt (2007: 2)

die mit ihnen korrespondierende Methode der Datengewinnung (die Beobachtung) kurz erläutert. Ferner werden Vorteile gegenüber dem subjektiven Verfahren Befragung genannt. Abschließend erfolgt die zusammenfassende Einordnung von Mystery Shopping.

Im Gegensatz zu den subjektiven Ansätzen, die auf subjektiven Einschätzungen einzelner Kunden basieren, fußen *objektive Ansätze* auf intersubjektiv nachprüfbar Beurteilungen von am Dienstleistungsprozess (un)beteiligten Dritten, deren Urteil als „objektiviert“ bezeichnet werden kann. Unterschieden werden die Verfahren Warentests, Expertenbeobachtungen und Mystery Shopping.⁹¹ Unter Warentests werden die Tätigkeiten der „Stiftung Warentest“ im Bereich des Vergleichs der Dienstleistungsqualität von unterschiedlichen Unternehmen jeweils einer Branche zusammengefasst. Dabei wird sich bei der Qualitätsmessung unterschiedlicher Verfahren der Datengewinnung bedient.⁹² Bei der Expertenbeobachtung wird der Prozess der Leistungserstellung durch geschulte Experten erfasst und analysiert. Ziel ist es, Erkenntnisse über offenkundige Mängel der Dienstleistungsqualität (z.B. Warteschlangen) und daraus resultierende Kundenreaktionen zu gewinnen.⁹³ Für Mystery Shopping sei auf die oben hergeleitete Definition verwiesen.⁹⁴

Sowohl die Expertenbeobachtung als auch Mystery Shopping bedienen sich spezieller Formen der *Beobachtung als Instrument der Datengewinnung*. Um einen theoretischen Überblick von Mystery Shopping zu erlangen, muss daher zunächst mit einer Betrachtung des relevanten sozialwissenschaftlichen Hintergrunds des Einsatzes von Beobachtungen (und somit auch von Mystery Shopping) als wissenschaftliche Methode begonnen werden. Dieser ist insbesondere in der Kulturanthropologie, in der Chicagoer Schule der qualitativen Sozialforschung und im Symbolischen Interaktionismus zu finden.⁹⁵ Der Einsatz quantitativer Methoden in der Sozialforschung wird dort kritisiert. Es wird darauf verwiesen, dass die soziale Realität zu komplex sei, als dass sie lediglich mit quantitativen Methoden, standardisierten Fragebögen usw. ausreichend erfasst werden könne. Vielmehr müssten die klassischen quantitativen Methoden durch die teilnehmende Beobachtung ergänzt werden, um eine größtmögliche Nähe zum Untersuchungsgegenstand zu erlangen und somit die Innenperspektive der zu beurteilenden Situation zu erschließen.⁹⁶ Unter Beobachtung versteht man in der empirischen Sozialforschung „die direkte Beobachtung menschlicher Handlungen, sprachlicher Äußerungen, nonverbaler Reaktionen (Mimik, Gestik, Körpersprache) und

⁹¹ Vgl. Bruhn (2006: 85)

⁹² Vgl. Rosenberger (2000: 341ff.); Da die Aktivitäten der Stiftung Warentest unter Warentests eine Vielzahl von Verfahren subsumieren, deren Vorgehensweisen nur bedingt Gegenstand dieser Arbeit sind, wird auf eine nähere Betrachtung verzichtet.

⁹³ Vgl. Bruhn (2006: 86)

⁹⁴ Vgl. Abschnitt 2.1.1.2 dieser Arbeit

⁹⁵ Vgl. z.B. Matzler/Pechlaner/Kohl (2000: 164)

⁹⁶ Vgl. Wilson (1998a: 149), Matzler/Kittinger-Rosanelli (2000: 225f.), Lingenfelder/Schmidt/Wieseke (2005: 195)

anderer sozialer Merkmale.“⁹⁷ Dabei können die unterschiedlichen Beobachtungsformen grundsätzlich anhand von unterschiedlichen Dimensionen klassifiziert werden: „teilnehmend vs. nicht-teilnehmend“, „strukturiert vs. unstrukturiert“, „offen vs. verdeckt“, „Feldbeobachtung vs. Laborbeobachtung“, „Selbstbeobachtung vs. Fremdbeobachtung“.⁹⁸ Während bei der nicht-teilnehmenden Beobachtung ablaufende Handlungen vom Beobachter lediglich protokolliert werden, ist der Beobachter bei der teilnehmenden Variante Interaktionspartner. Eine Strukturierung liegt vor, wenn die Beobachtung einem Schema unterliegt, ähnlich der Standardisierung bei den oben aufgeführten Interviews. Liegt eine Strukturierung vor, sind die gewonnenen Informationen leichter vergleichbar. Bei einer offenen Beobachtung ist der Beobachtete über die Beobachtung informiert, wohingegen der Interaktionspartner sich bei der verdeckten Beobachtung nicht zu erkennen gibt. Feldbeobachtungen finden im natürlichen Umfeld der beobachteten Person statt, während Laboruntersuchungen unter künstlichen Bedingungen stattfinden. Bei der Selbstbeobachtung wird das eigene Ich beobachtet, wohingegen bei der Fremdbeobachtung die Welt außerhalb des Ichs beobachtet wird. In Tabelle 2-1-3 werden Expertenbeobachtung und Mystery Shopping anhand der Dimensionen klassifiziert.

Tabelle 2-1-3: Klassifizierung von Expertenbeobachtung und Mystery Shopping

Dimension	Expertenbeobachtung	Mystery Shopping
<i>Teilnahme</i>	nicht-teilnehmend	teilnehmend
<i>Strukturierung</i>	je nach Forschungsfrage	strukturiert
<i>Offenheit</i>	je nach Forschungsfrage	verdeckt
<i>Beobachtungsort</i>	je nach Forschungsfrage	Feldbeobachtung
<i>Gegenstand</i>	Fremdbeobachtung	Selbst- und Fremdbeobachtung

Quelle: Eigene Darstellung

Es wird deutlich, dass sich Expertenbeobachtung und Mystery Shopping speziell durch die Teilnahme und den Gegenstand (Selbst- oder Fremdbeobachtung) unterscheiden. Die Leistungsfähigkeit der Expertenbeobachtung kann insbesondere durch die mangelnde Aussagekraft der erhobenen Kundenreaktionen in Frage gestellt werden, die nur bedingt Rückschlüsse auf die wirkliche Wahrnehmung zulassen. Daneben kann das Verhalten bei offener Beobachtung verzerrt werden.⁹⁹ Im Vergleich dazu bietet Mystery Shopping durch die Teilnahme an der Beobachtung, die zudem verdeckt erfolgt, sowie durch die Kombination von Selbst- und Fremdbeobachtung, je nach Grad der Ausgestaltung des Beobachtungskataloges, umfangreichere Möglichkeiten. Insbesondere die verdeckte

⁹⁷ Diekmann (2007: 456)

⁹⁸ Vgl. hierzu und im Folgenden Diekmann (2007: 469ff.)

⁹⁹ Vgl. Bruhn (2006: 86)

Teilnahme ermöglicht ein nicht-reaktives Verhalten der beobachteten Person, da diese ihr Verhalten nicht ändern, um beispielsweise in einem günstigeren Licht zu stehen.¹⁰⁰

Abschließend soll die *Vorteilhaftigkeit von Mystery Shopping gegenüber der Befragung* deutlich gemacht werden, die einerseits aus Charakteristika der Beobachtung und andererseits aus der Unterschiedlichkeit von Mystery Shopping und Kundenbefragungen resultiert.¹⁰¹ Es besteht eine deutliche Diskrepanz der Informationen, die mit einer Beobachtung und einer Befragung gewonnen werden. Die Beobachtungsergebnisse sind deutlich genauer und weniger verzerrt als die einer Befragung.¹⁰² Die Vorteile der Beobachtung ergeben sich überdies durch die Datengewinnung während des eigentlichen Dienstleistungsaktes und der gleichzeitigen Erhebung von Umwelteinflüssen. Dies lässt auch die Erfassung von Sachverhalten zu, die den beobachteten Personen selbst nicht bewusst sind.¹⁰³ Insbesondere bei Kundenbefragungen ist wie bei Expertenbeobachtungen mit einem Beobachtungseffekt zu rechnen, welcher sich in einer temporären Verhaltensänderung der Mitarbeiter während des Erhebungszeitraumes niederschlagen kann. Ferner ist es denkbar, dass bei Kunden eine Erwartungshaltung entsteht, nachdem Befragungen angekündigt wurden, wohingegen Mystery Shopping verdeckt durchgeführt werden kann. Als pragmatische Gründe sind Umfang und Flexibilität anzuführen: Kundenbefragungen sind im Umfang begrenzt, da die Kunden auf freiwilliger Basis antworten, während Mystery Shopper entgolten werden. Kundenbefragungen sind weniger flexibel, da eine Anpassung während des Erhebungszeitraumes nicht praktikabel und wenig sinnvoll ist. Mystery Shopping-Beobachtungen können hingegen bei Bedarf jederzeit angepasst werden. Zudem existieren unterschiedliche Studien, die Kundenbefragung und Mystery Shopping hinsichtlich Ergebnisqualität und Leistungsfähigkeit vergleichen. Mystery Shopping erwies sich dabei hinsichtlich der Ergebnisqualität im Vergleich zur Kundenbefragung als verlässlicher.¹⁰⁴ Zudem ist die Erhebung mit deutlich weniger Kosten verbunden als bei einer Kundenbefragung.¹⁰⁵ In Tabelle 2-1-4 sind verschiedene Aspekte von Kundenbefragung und Mystery Shopping gegenübergestellt, welche die Unterschiede hinsichtlich der Leistungsfähigkeit verdeutlichen.

¹⁰⁰ Vgl. Diekmann (2007: 470)

¹⁰¹ Für einen Überblick über das Thema Kundenbefragungen siehe Winter (2005)

¹⁰² So stellten Hürlimann/von Hebenstreit bei der Analyse von Verkehrsverhalten eine erhebliche Diskrepanz von befragtem und beobachtetem Verhalten fest. 98% der Befragten gaben etwa an, dass sie sich bei Autofahrern bedanken, wenn sie über die Straße gelassen werden, während dieses Verhalten nur bei 18% beobachtet werden konnte; vgl. Hürlimann/von Hebenstreit (1987)

¹⁰³ Vgl. Bruhn (2001: 105)

¹⁰⁴ z.B. Finn/Kayandé (1999), Finn (2001), Stücken (2003), Hesselink/van der Wiele (2003)

¹⁰⁵ Finn/Kayandé (1999: 215) beziehen die Kosten auf die Anzahl der jeweils notwendigen Tests, um eine angemessene und gleich hohe Reliabilität der Ergebnisse zu erreichen.

Tabelle 2-1-4: Kundenbefragung vs. Mystery Shopping

	Kundenbefragung	Mystery Shopping
<i>Beurteilung</i>	Subjektiv	Weitgehend objektiviert
<i>Beurteilungsfähigkeit</i>	Eingeschränkt, da Kundenwahrnehmung beeinflusst wird (z.B. durch langfristiges Unternehmensimage, persönliche Kontakte zu Mitarbeitern, bisherige Erfahrungen) und ggf. eine verzerrte Wahrnehmung durch eine zeitliche Versetzung zwischen Erleben und Erhebung besteht.	Hoch, da weitgehende Objektivierung erfolgt (z.B. durch Briefing, gezielte Selektion der Fragestellungen, geschärftes Bewusstsein, Erfahrungen mit anderen Tests, Realitätsnähe des Tests) und keine zeitliche Versetzung zwischen Test und Dokumentation erfolgt.
<i>Beobachtungseffekt</i>	Ergebnis ggf. nicht repräsentativ, da temporäre Verhaltensänderung der Mitarbeiter wahrscheinlich ist	Bei unangekündigter Vorgehensweise kann ein verzerrender Effekt der Verhaltensänderung ausgeschlossen werden
<i>Zielsetzung</i>	Abfrage der Zufriedenheit	Abfrage des Erfüllungsgrads
<i>Maßstab</i>	Subjektive Empfindungen	Objektive Vorgaben
<i>Wahrnehmung der Kunden</i>	Kunde nimmt Befragung möglicherweise als kundenorientierte Maßnahme wahr (Gefahr: Erwartungsdruck)	Keine Wahrnehmung der Maßnahme durch den Kunden
<i>Umfang</i>	Begrenzt, da Kunden auf freiwilliger Basis antworten	Umfangreicher, da Tester entgolten werden
<i>Kosten</i>	Höher (bezogen auf eine angemessene und gleich hohe Reliabilität)	Niedriger (bezogen auf eine angemessene und gleich hohe Reliabilität)
<i>Flexibilität</i>	Geringe Flexibilität, da z.B. eine Anpassung während der Erhebung nicht praktikabel und wenig sinnvoll ist	Hohe Flexibilität, z.B. durch kurzfristige Anpassbarkeit im Vorgehen

Quelle: Eigene Darstellung

Abschließend soll die Einordnung dieses Verfahrens erfolgen: Nachdem in Publikationen vor den 90er Jahren die Einordnung von Mystery Shopping zwischen Interview, Beobachtung und Experiment schwankte¹⁰⁶, ist sie in jüngeren Veröffentlichungen einheitlich und unstrittig:¹⁰⁷ Mystery Shopping ist innerhalb der Marktforschung in die Primärforschung einzuordnen. Abgegrenzt von den unternehmensorientierten Verfahren ist sie bei den kundenorientierten Verfahren und überdies bei den objektiven Messverfahren einzuordnen, da nicht die subjektive Wahrnehmung, sondern die möglichst objektive Beurteilung von Qualitätsaspekten zentraler Gegenstand des Verfahrens ist. Informationen werden in einer verdeckt teilnehmenden und strukturierten Feldbeobachtung erhoben. Dabei können sowohl quantitative wie qualitative Daten gewonnen werden.

¹⁰⁶ z.B. Geiger (1962: 101), Jauschowitz (1978: 35), Schumeckers (1980), Keller (1982: 116f.)

¹⁰⁷ Vgl. Deckers (1996: 29), Deckers (1999: 29), Matzler/Kittinger-Rosanelli (2000: 226), Eikenroth (2004: 21ff.), Bruhn (2006: 86f.) sowie Girtler (1992: 45) und Diekmann (2007: 469ff.) für die Dimensionen von Beobachtungsformen

Nach Erläuterung der Begriffe Dienstleistung und Dienstleistungsqualität, Darstellung subjektiver und objektiver Verfahren zur Evaluation der Dienstleistungsqualität sowie der mit ihnen korrespondierenden Verfahren zur Datengewinnung wurden die Defizite subjektiver Verfahren adressiert und die Vorteilhaftigkeit der Methode Mystery Shopping herausgearbeitet. Die Ausführungen bereiten den Boden für zu beachtende Aspekte, die Mystery Shopping in der Umsetzung zu einem leistungsfähigen Instrument machen.

2.1.3 Entwicklung von Mystery Shopping im Zeitverlauf

Bevor auf die gegenwärtige Struktur und Umsetzungsaspekte von Mystery Shopping eingegangen wird, ist es von Bedeutung, die Entwicklung der Methode näher zu betrachten. Dafür werden nachfolgend Ursprung und Wurzeln sowie die Entwicklung bis zum derzeitigen Status Quo dargelegt.

2.1.3.1 Ursprung und Wurzeln von Mystery Shopping

Die praktischen Wurzeln von Mystery Shopping sind bereits Ende des 19. Jahrhunderts zu finden, wenn auch in einer sehr speziellen Form: Die Gebrüder Edmond und Jules berichten in ihren Tagebuchaufzeichnungen über die Angestellte eines Hutladens, die im Laden eines Wettbewerbers die Produkte begutachtet, um Anhaltspunkte über Produktion und Material zu erhalten¹⁰⁸. Diese Vorgehensweise ist vergleichbar mit der Zerlegung von Wettbewerbsfahrzeugen in der Automobilindustrie und dem Warentest im Einzelhandel.¹⁰⁹

Die Beobachtung des Beratungs- und Kaufvorgangs begann erst in der zweiten Hälfte des 20. Jahrhunderts. Dabei sind insbesondere zwei Unternehmen als Wegbereiter zu nennen: die amerikanische Wilmark Service System Inc. (WSS) und die deutsche Gesellschaft für Konsumforschung (GfK). WSS schickte so genannte „Analysts“ in Geschäfte, die auf einem Erhebungsbogen das Geschick und die Erscheinung des Personals sowie Räume und Warenqualität festhielten und diese in einem Verkaufsleistungsindex („selling quotient“) zusammenfassten¹¹⁰. In den 40er Jahren verwendeten sie das erste Mal den Begriff „Mystery Shopping“¹¹¹. Die GfK hat in den 30er Jahren das „Befragungsgespräch“ eingeführt, bei dem der Interviewer sich als solcher nicht zu erkennen gibt. Die so genannte „Nürnberger Methode“ sollte den Vorteil einer ungezwungenen Atmosphäre aufweisen und dadurch zu erhöhter Validität führen.¹¹²

¹⁰⁸ Vgl. Goncourt/Goncourt (1983: 224), zitiert bei Deckers (1999)

¹⁰⁹ Vgl. Eichholz (1966: 36), Deckers (1999: 7)

¹¹⁰ Vgl. Schweisheimer (1962); Vgl. auch Eichholz (1966: 36)

¹¹¹ Vgl. Michelson (2005:6)

¹¹² Vgl. Deckers (1999: 8)

Diese Anfänge ignorierend, wurde die Vorgehensweise von Mystery Shopping von unterschiedlichen Autoren bis heute immer wieder als ein neuartiges und innovatives Instrument beschrieben, wenn auch mit unterschiedlichen Begriffen¹¹³. Gleichwohl merkt Berger bereits 1958 an: „Beim Kundentest handelt es sich um eine ganz bestimmte Befragungsmethodik. Sie ist in ihrer Art keineswegs neu.“¹¹⁴

2.1.3.2 Mystery Shopping von den 1950er bis zu den 2000er Jahren

Aus den 50er und 60er Jahren sind nur wenige Beiträge zu finden. Die Bezeichnung „Einkaufstest“ und die Veröffentlichungen im Bereich Absatz und Handel lassen darauf schließen, dass das Verfahren vor allem im Bereich des Einzelhandels zur Anwendung kommt.¹¹⁵ Dabei werden zwei entwickelte Anwendungsmöglichkeiten beschrieben: die herstellerorientierte und die handelsorientierte Variante.

Erstere dient dem Interesse der Industrie, die Informationen darüber gewinnt, wie sie im Handel mit ihren Produkten vertreten ist. Dabei lassen sich unterschiedliche Formen mit unterschiedlichen Zielsetzungen vergleichen. So kann der Einkaufstest beispielsweise Kontrollmittel sein, um die Einhaltung von Absatzwegen oder Preisbindung¹¹⁶ zu überprüfen. Er kann beim Identifizieren der Einstellung des Verkaufspersonals zu bestimmten Produkten helfen oder zur Überprüfung der Darbietung der Waren in den Geschäften dienen.¹¹⁷

Bei der handelsorientierten Variante wird der Schwerpunkt auf das Verkaufspersonal und das Ladenlokal gelegt. Erhoben werden beim Personal Erscheinungsbild und Umgangsformen, Fachkenntnisse und Verkaufsgeschick, etwa bei Einwänden oder beim Anbieten von Zusatzverkäufen.¹¹⁸ Bei den Verkaufsräumen werden Kriterien wie der Zustand der Räumlichkeiten oder die Schaufenstergestaltung bewertet.¹¹⁹

Bei Betrachtung der in der Praxis angesiedelten neueren Beiträge wird deutlich, dass insbesondere die handelsorientierte Variante in ihren Grundzügen auch heute noch eingesetzt wird¹²⁰, während über die herstellerorientierte Variante keine Beiträge über Erfahrungen zu finden sind.

¹¹³ Vgl. z.B. Geißler/Dingler (1969), Pergandé/Stücken (2004: 205), Böhm/Fischl/Rennhak (2005:1) sowie den Abschnitt 2.1.1.1. *Terminologie* dieser Arbeit

¹¹⁴ Vgl. Berger (1958: 122)

¹¹⁵ Vgl. Berger (1958), Geiger (1962), Eichholz (1966), Geißler/Dingler (1969)

¹¹⁶ Der Einsatz des Testkaufs als Mittel zur Überwachung der Preisbindung findet sich insbesondere in unterschiedlichen juristischen Beiträgen, z.B. Johannes (1961), Brücher (1962), Eckhardt (1968)

¹¹⁷ Vgl. Berger (1958: 124f.), Eichholz (1966: 35)

¹¹⁸ Vgl. Eichholz (1966: 36), Geißler/Dingler (1969)

¹¹⁹ Vgl. Eichholz (1966: 36), Geiger (1962: 101)

¹²⁰ z.B. Drees/Schiller (2001), Drees/Schiller (2003), Planung & Analyse 30 (5), Höhner/Schaper (2004)

In *den 1970er Jahren* entdecken Unternehmen aus der Finanzwirtschaft die Methode für sich und setzen sie als so genannten „Schaltertest“ ein.¹²¹ In dieser auf Kreditinstitute angepassten Form wird Mystery Shopping insbesondere angewendet, um das Verhalten im Neukundengeschäft zu professionalisieren und eine „sales culture“ zu entwickeln und diese vor allem nachweislich umzusetzen.¹²² Dabei wird bei einfachen Bankgeschäften der „beobachtende Schaltertest“ und bei komplexeren Beratungssituationen der „Kundenberatungstest“ eingeführt.¹²³ Die Einsatzmöglichkeiten der Methode erfahren (zumindest in den USA) einen Wandel: Vom reinen Beobachtungsinstrument wird die Methode zum Diagnoseinstrument und teilweise mit Ziel- und Anreizsystemen gekoppelt, um Leistungsanreize zu schaffen.¹²⁴

Die 1980er Jahre sind von einer zunehmenden Ausbreitung der Methode und eines strukturierteren Einsatzes gekennzeichnet. In Deutschland wird das Thema in diversen Beiträgen aufgegriffen und eine Anwendung in unterschiedlichen Bereichen des Dienstleistungsbereiches beschrieben.¹²⁵ Dabei wird das Instrument insbesondere für die Finanzbranche als „verbreitet“ und „routiniert eingesetzt“ beschrieben.

In den USA entwickelt sich das Verfahren in der Finanzbranche zu einem Standardinstrument.¹²⁶ Ferner wird dort eine Brücke zur Kundenzufriedenheit geschlagen: Im Rahmen der so genannten „dualen Methodologie“ werden zunächst die Einflussfaktoren der Kundenzufriedenheit bei den Kunden abgefragt. Dann wird mit Hilfe von Mystery Shopping überprüft, ob die aus den Einflussfaktoren abgeleiteten Servicestandards eingehalten werden. Die Verknüpfung der beiden Ansätze führt dazu, dass Mystery Shopping als Prognoseinstrument eingesetzt werden kann.¹²⁷

Für *die 1990er Jahre* stellt DECKERS die Frage, ob es das „Jahrzehnt des Testkundenverfahrens“ sei.¹²⁸ Im Vergleich zum vorangegangenen Jahrzehnt gewinnt das Verfahren zunehmend an Bedeutung. Der Einsatz der Methode ist in zahlreichen Branchen dokumentiert. DECKERS liefert eine akribisch erarbeitete Auflistung aller bis 1997 in (praktischen und wissenschaftlichen) Veröffentlichungen und Beiträgen dokumentierter Einsatzorte: Ämter und Behörden, Anwälte, Apotheken, Autohändler, Fast-Food-Restaurants,

¹²¹ Vgl. Achelis (1975), Küppers (1977), Jauschowitz (1978), Schumeckers (1980), Keller (1982), Leeds (1995)

¹²² Vgl. Leeds (1995: 18)

¹²³ Vgl. Jauschowitz (1978: 35)

¹²⁴ Vgl. Leeds (1995: 18), Lubin (2001: 29)

¹²⁵ Vgl. z.B. Achelis (1985), Landwehr (1984), Salcher/Esser (1987) und insbesondere Deckers (1999: 11) und die dort zitierte Literatur

¹²⁶ Vgl. Leeds (1995: 18), Lubin (2001: 29)

¹²⁷ Vgl. Leeds (1995: 18)

¹²⁸ Vgl. Deckers (1999: 12)

Fluggesellschaften, Freizeitparks, Gaststätten, Hotels, Krankenhäuser, Postdienstleister, Reformhäuser, Reisebüros, Spielhallen, Supermärkte, Tankstellen, Unternehmensberatungen, Verkehrsunternehmen, Versicherungen, Warenhäuser, Werkstätten.¹²⁹ In den 1990er Jahren gewinnt die Methode auch hinsichtlich der Einsatzhäufigkeit an Bedeutung. Aus den USA wird berichtet, dass jeder Mitarbeiter der Bank One in Denver einmal jährlich getestet wird und dass Filialen alle 90 Tage von Mystery Shoppern besucht werden.¹³⁰ Nachweise für eine vergleichbare Häufigkeit sind für den Einsatz in Deutschland allerdings nicht zu finden.

Hinsichtlich der Anzahl der Anbieter von Mystery Shopping ist im Ausland zum Teil ein erhebliches Wachstum zu verzeichnen.¹³¹ Allerdings ist dies für Deutschland nicht zu bestätigen, zumindest nicht anhand der Anbieter, die Mitglied beim Bundesverband Deutscher Markt- und Sozialforscher (BVM) sind.¹³² Die gebremste Entwicklung hierzulande ist nach DECKERS Überlegungen eine Folge der Diskussion über den Datenschutz sowie restriktiver ethischer Richtlinien der Marktforschungsverbände. Gleichwohl ist anzunehmen, dass sich aufgrund der Restriktionen eine Entwicklung außerhalb der Verbände vollzieht.¹³³ Diese Überlegung lässt sich anhand der Vielzahl von dokumentierten Mystery Shopping-Anwendungen nachvollziehen, wenngleich dies nur als Indiz gilt, da von der Quantität der Publikationen nicht auf Entwicklungen in der Realität geschlossen werden kann.

Die zunehmende Bedeutung ist nach Zanger/Sistenich (1996: 351) auf zwei Phänomene zurückzuführen:¹³⁴

Erstens hat sich der bereits in den 1980er Jahren diskutierte Ansatz der Servicequalität noch stärker verbreitet. Es wurde deutlich, dass viele Unternehmen in diesem Differenzierungskriterium einen Wettbewerbsvorteil sahen. Folglich hat auch die Marktforschungsseite diesen Aspekt aufgegriffen, so dass Mystery Shopping sich als Bestandteil der Qualitätsmessung etablieren konnte. Ferner wurde erneut das Einsatzgebiet erweitert. War es vor den 1990er Jahren die direkte Schnittstelle zum Kunden, die im Rahmen von Mystery Shopping bewertet wurde, wurde das Verfahren auf Mitarbeiter im „back office“ ausgedehnt.¹³⁵ Grund für die Ausdehnung ist der Gedanke, dass Erfolge oder Misstände an der „Verkaufsfront“ das Ergebnis des Funktionierens der gesamten Kunden-Lieferantenkette sind.

¹²⁹ Vgl. ebenda (1999: 12f.) und die dort zitierte Literatur. Auf eine Ergänzung um aktuellere Beiträge wird an dieser Stelle verzichtet, da der branchenübergreifende Einsatz durch Deckers Auflistung deutlich wird.

¹³⁰ Vgl. Morrall (1994: 16)

¹³¹ So hat sich in Großbritannien die Anzahl der Anbieter von 1994 auf 1995 von 28 auf 187 erhöht. Vgl. Dawson/Hillier (1995: 420)

¹³² Im Jahre 1997 waren es drei Unternehmen, Vgl. Deckers (1999: 13)

¹³³ Vgl. Deckers (1999: 13)

¹³⁴ Vgl. Zanger/Sistenich (1996: 351)

¹³⁵ Vgl. Morrall (1995: 66)

Zweitens wird von Zanger/Sistenich ein „Differenzierungsprozess auf Konsumentenseite“ angeführt, dem mit bisher eingesetzten Methoden der Marktforschung nicht begegnet werden kann. Mystery Shopping wird zugetraut, valide Einsichten in den neuartigen Konsumenten zu liefern. Zudem wird vermutet, dass der Einsatz der Methode dazu dient, dem gestiegenen Informationsbedarf von Konsumenten zu entsprechen und einem Wechsel der Kunden zu einem Wettbewerber entgegenzuwirken.¹³⁶ Dieser Informationsbedarf wird auch durch Zeitschriften aufgegriffen, die Produkte und Dienstleistungen testen, um ihren Lesern eine bessere Einschätzung der Qualität liefern zu können.¹³⁷

In *den 2000er Jahren* steht die Entwicklung von Mystery Shopping im Zeichen der Kommunikation sowie der EDV und des Internets. Unternehmen ermöglichen ihren Kunden über unterschiedliche Kommunikationskanäle mit ihnen in Kontakt zu treten (etwa Email, Homepage, Call Center, Newsletter). Mystery Shopping wird zunehmend eingesetzt, um das Gelingen dieser Kommunikation zu überprüfen (z.B. Mystery eMailing, Mystery Calling).¹³⁸ Daneben werden weitere Unternehmensbereiche in die Überprüfung durch Mystery Shopping eingeschlossen, z.B. im Vertriebscontrolling¹³⁹ und als Controlling-Instrument im Beschwerdemanagement¹⁴⁰.

Hinsichtlich des Branchenbezuges ist festzustellen, dass sich Mystery Shopping branchenübergreifend als eine Methode der Marktforschung etabliert hat.¹⁴¹ Hinzugekommen sind weitere Branchen mit hoher Wettbewerbsintensität (z.B. die Internet- und Telekommunikationsbranche). Ferner werden einst staatliche Branchen privatisiert (z.B. Energieversorgung, Verkehrsbetriebe). Daneben steigt die Wettbewerbsintensität in gesättigten Märkten, während diese von neuen Rahmenbedingungen getroffen werden (z.B. stagnierende Nachfrage und Novellierung der Gruppenfreistellungsverordnung in der Automobilbranche). Ferner steigt die Bedeutung der in den 80er Jahren geführten Diskussion um Servicequalität und Kundenzufriedenheit und kommt nun auch in marktfernen Bereichen an.¹⁴²

Es ist anzunehmen, dass elektronische Auswertungsmöglichkeiten sowie schnellere Informationsübertragung der Tester über das Internet an das ausführende Unternehmen zur

¹³⁶ Vgl. Leeds (1995: 20)

¹³⁷ Vgl. Deckers (1999: 16), Rosenberger (2000)

¹³⁸ Vgl. Pergandé/Stücken (2004: 209)

¹³⁹ Vgl. Keller/Stücken (2004), Ferchland/Paul (2006); Insbesondere im Vertriebscontrolling erfreut sich Mystery Shopping großer Beliebtheit: So ist Mystery Shopping im Einzelhandel mit 65% das präferierte Instrument vor der Kundenzufriedenheitsmessung (58%) und dem Beschwerdemanagement (47%), vgl. Infas TTR (2005: 9)

¹⁴⁰ Vgl. Höhner (2003)

¹⁴¹ Vgl. Krüger/Stücken (2003: 36)

¹⁴² Vgl. Deckers/Heinemann (2006: 17)

Verbesserung und weiteren Verbreitung des Methodeneinsatzes beitragen. Mit multivariaten Analysemethoden lassen sich z.B. auch komplexe Zusammenhänge von Daten ergründen, die mit Hilfe einer Software (etwa SPSS) vergleichsweise einfach auszuwerten sind. Über das Internet ist es Anbietern von Mystery Shopping zudem möglich, vergleichsweise schnell mit ihren Testern in Kontakt zu treten und auch größere Aufträge datenverarbeitungsgestützt effizient abzuwickeln (z.B. können freie Tests per Email angeboten und angenommen werden, ohne dass ein Mitarbeiter unter Zeitaufwand alle in Frage kommenden Tester identifiziert und abtelefoniert oder auf dem Postwege anschreibt).

Die vereinfachten Auswertungsmöglichkeiten bringen die Diskussion um Methoden auf den Plan. So beschäftigen sich unterschiedliche Beiträge mit der Einhaltung der Gütekriterien sozialwissenschaftlicher Methoden (Reliabilität, Validität, Repräsentativität) bei Mystery Shopping und erbringen den Nachweis, dass die Methode Dienstleistungsqualität reliabel, valide und repräsentativ messen kann.¹⁴³

Aus den in der Literatur beschriebenen Ansätzen und Einsatzgebieten lässt sich eine klare thematische Entwicklung abzeichnen: Während in den 50er und 60er Jahren das Verfahren im Einzelhandel eingesetzt wurde, wird es in den 70er Jahren von Kreditinstituten aufgegriffen und in den 80er Jahren weiter ausgebaut. Ab den 90er Jahren erfährt Mystery Shopping eine zunehmende Popularität sowie eine branchenübergreifende Verbreitung. Mit Beginn des neuen Jahrtausends steigt diese Verbreitung unterstützt durch die elektronischen Auswertungs- und Abwicklungsmöglichkeiten weiter an. Ferner werden erste erfolgreiche empirische Befunde zur Güteprüfung erbracht.

2.1.3.3 Zukünftige Entwicklung von Mystery Shopping

In der Zukunft ist weiterhin mit einer zunehmenden Bedeutung von Mystery Shopping zu rechnen. Dies resultiert aus folgenden Entwicklungstreibern:¹⁴⁴ Zunehmende Wettbewerbsintensität erfordert Stärken in Vertrieb und Service. Dabei sind insbesondere solche Branchen, die sich weniger über den Preis, sondern mehr über Kundenbeziehungen organisieren vom professionellen Umgang mit der knappen Ressource Kunde abhängig. Branchenübergreifend erfordert der steigende Kosten- und Effizienzdruck geeignete Maßnahmen, um Prozesse, Marketingaktionen und Mitarbeiterverhalten wirksam steuern zu können. Hierfür kann das Gelingen mit Mystery Shopping überprüft werden und die Grundlage für Verbesserungen bilden.

Zunehmende Erfahrung in der Abwicklung von Mystery Shopping-Projekten könnte sich zukünftig in einer Standardisierung von Abläufen und Inhalten niederschlagen. Diese

¹⁴³ Vgl. Finn/Kayandé (1999), Deckers (1999), Dawes/Sharp (2000), Haas (2003), Deckers (2003), Schmidt (2007)

¹⁴⁴ Vgl. hierzu und im Folgenden Deckers/Heinemann (2006: 18f.)

Vereinfachung erleichtert möglicherweise den Markteintritt weiterer Anbieter von Mystery Shopping. Diese können neben der Weiterentwicklung der Methode zur weiteren Verbreitung des Einsatzes beitragen. Es ist anzunehmen, dass sich solche Anbieter durchsetzen, die über besonders qualifizierte Tester und effiziente Abwicklungssysteme verfügen. Vor diesem Hintergrund ist eine weitergehende empirische Überprüfung von Gütekriterien der Methode sowie einer näheren Untersuchung von Aspekten, die sich auf die Qualifizierung der Tester beziehen, zwingend erforderlich.

2.1.4 Würdigung vergangener Forschungsbemühungen zur Mystery Shopping-Thematik

Wissenschaftlich orientierte Veröffentlichungen (etwa in renommierten Fachzeitschriften, in Diskussionspapieren von Universitäten oder als wissenschaftliche Beiträge in Fachbüchern) sind wenig zahlreich, was bei der Verbreitung der Methode am Markt von den entsprechenden Wissenschaftlern mit Erstaunen angemerkt wird.¹⁴⁵ Vielmehr sind als theorielos einzuordnende Beiträge zu finden, die von Praktikern für Praktiker verfasst wurden. So widmet die im Bereich der Praxis angesiedelte Zeitschrift „Planung & Analyse“ dem Thema 2003 sogar ein ganzes Heft mit zwölf Beiträgen¹⁴⁶.

Um einen Überblick über die als wissenschaftlich einzustufenden Beiträge zum Thema Mystery Shopping zu gewinnen, soll in diesem Abschnitt tabellarisch eine vollständige Übersicht geliefert werden, die Schwerpunkt, Branchenkontext sowie den relevanten Inhalt skizziert. Ferner soll der Forschungsstand knapp zusammengefasst werden.

2.1.4.1 Überblick der Beiträge

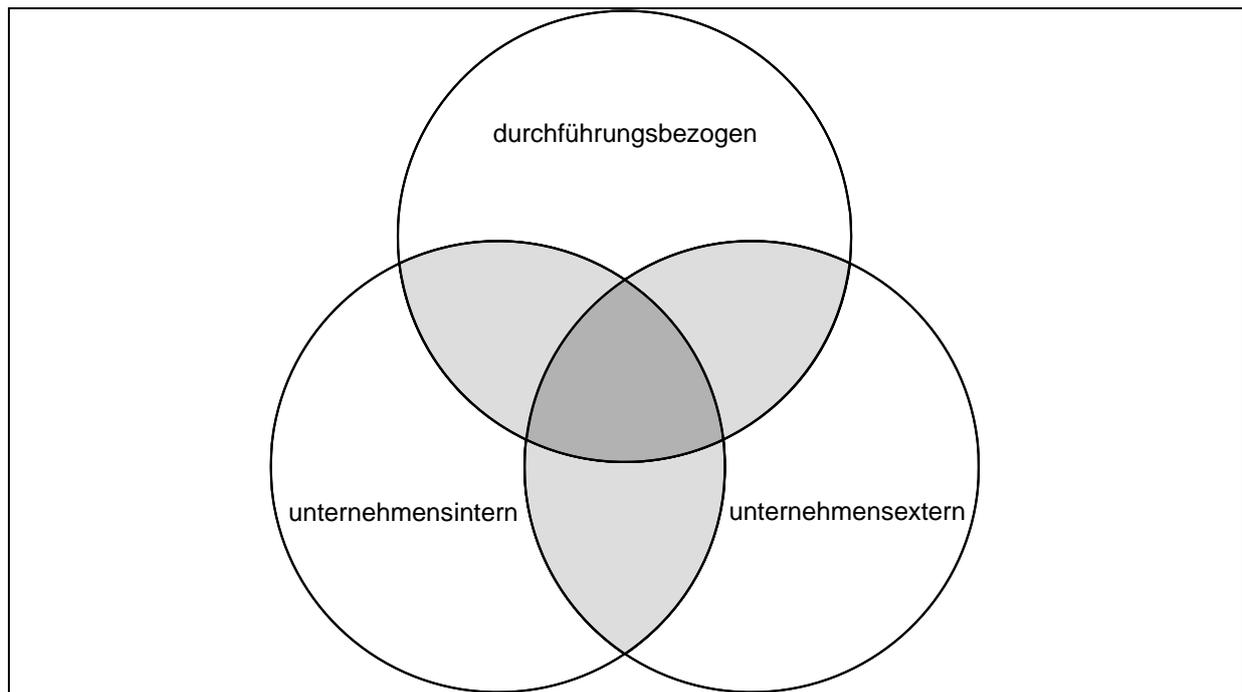
Beiträge vergangener Forschungsbemühungen lassen sich anhand folgender Teilaspekte klassifizieren:

- Durchführungsbezogene Aspekte (etwa methodische Betrachtung, ethische Richtlinien, Leitfäden)
- Unternehmensinterne Aspekte (etwa Fokus auf Personalbereich, Investitionsentscheidungen, Marketing, interne Kommunikation)
- Unternehmensexterne Aspekte (etwa Fokus auf Wettbewerbsvergleich, Markt Betrachtung)

Abbildung 2-1-2 soll verdeutlichen, dass nicht nur eine Zuordnung zu einem Teilaspekt möglich ist, sondern auch zu zwei oder allen drei Aspekten.

¹⁴⁵ Vgl. Wilson (1998: 148); Matzler/Kittinger-Rosanelli (2000: 221), Lingenfelder/Wieseke/Schmidt (2005: 183)

¹⁴⁶ siehe Planung & Analyse 30 (5), (2003)

Abbildung 2-1-2: Mystery Shopping – Klassifikationsschema vergangener Forschungsbemühungen

Quelle: Eigene Darstellung

Tabelle 2-1-5 verschafft einen Überblick vergangener Forschungsbemühungen. Die Beiträge werden nach Jahren aufsteigend dargestellt. Dabei wird der Schwerpunkt der Arbeit, der jeweilige Branchenkontext genannt sowie der relevante Inhalt wieder gegeben.

Tabelle 2-1-5: Wissenschaftliche Beiträge

Jahr	Autoren	Schwerpunkt	Branche	Relevanter Inhalt/ Erkenntnis	Einordnung
1958	Berger	Vorstellung der Methode „Kunden-Test“	Einzelhandel	Präsentation von Empfehlungen für Einsatzbereiche, Aufbau, Durchführung, Varianten und Kopplung mit andere Methoden, die auf eigenen Ideen und auf Erfahrungen der GfK basieren	-durchführungsbezogen
1982	Keller	Der „Schaltertest“ als Kontrollinstrument des persönlichen Verkaufs für Kreditinstitute	Banken	-Der Schaltertest wird als brauchbares Analyseinstrument für Verkaufsverhalten bei Kreditinstituten erkannt und beschrieben -Es wird empfohlen, die Ergebnisse als Basis für Trainings zu verwenden bzw. um deren Erfolg zu kontrollieren -Eine Kombination mit Mitarbeiter- und Kundenbefragungen wird empfohlen	-durchführungsbezogen -unternehmensintern
1992	Deges	Betrachtung des Einsatzes von Testkunden im Einzelhandel	Einzelhandel	Vorstellung als Instrument zur Beurteilung der Qualität der Kundenbetreuung, als Kassentest und als Diebstahltest	-durchführungsbezogen -unternehmensintern
1995	Dawson, Hillier	Verhaltenskodex bei Mystery Shopping im Rahmen von Wettbewerbsbeobachtung	keine	-Grundsätzliche Empfehlungen zum Verhalten bei Mystery Shopping im Rahmen von Wettbewerbsbeobachtungen -Hinweis auf branchenspezifische Unterschiede -Forderung einheitlicher Leitlinien auf Basis der Ergebnisse	-durchführungsbezogen -unternehmensextern
1997	Morrison, Colman, Preston	Reduzierung von Fehlern durch die Berücksichtigung von Erkenntnissen aus der	keine	-Ableitung von Maßnahmen und Gestaltungshinweisen aus der kognitiven Psychologie, um Reliabilität und Validität zu gewährleisten -Aufzeigen einer Forschungslücke: Effekte von soziodemografischen Faktoren auf die Güte der Ergebnisse	-durchführungsbezogen

		kognitiven Psychologie			
1998	Erstad	Mystery Shopping und Human Resource Management	keine	Mystery Shopping kann durch Einbinden der Mitarbeiter in den Prozess als ein nützliches Instrument des Personalwesens funktionieren	-durchführungsbezogen -unternehmensintern
1998 (und 2001)	Wilson	Explorativer Ansatz zur Bestimmung der Rolle von Mystery Shopping beim Messen der Servicequalität in England	keine	-Auf Basis von Experteninterviews wird die Rolle von Mystery Shopping in England skizziert -Mystery Shopping hat die Funktion eines Diagnoseinstruments, eines Wettbewerbsvergleichsinstruments und wird als Trainingsgrundlage verwendet -Ergebnisse von Mystery Shopping führen kurzfristig zu Erfolgen, langfristig werden die Ergebnisse nicht als Basis für Veränderungsprozesse verwendet, da keine Systematik vorhanden ist, der z.B. Mitarbeitern die Vorteile verdeutlicht -Mystery Shopping wird in England nicht mit anderen Messinstrumenten kombiniert eingesetzt	-durchführungsbezogen -unternehmensextern
1998	Wilson, Gutman	Die Rolle von Mystery Shopping bei Investitionsentscheidungen	Verkehr	-Einsatz von Mystery Shopping bei der Entwicklung von Business Cases im Rahmen von Investitionsentscheidungen der Londoner U-Bahn -Bestimmte Einflussfaktoren auf die Reisequalität wurden von Mystery Shoppem erhoben. Diese wurden in einem Modell mit Passagierzahlen und deren Ansprüchen zusammengeführt. Daraus wurden unterschiedliche „levels of improvement“ mit jeweils unterschiedlichen Kosten abgeleitet. -Mystery Shopping wird erfolgreich als Instrument eingesetzt, um vorab den Nutzen von Investitionsentscheidungen abzubilden	-durchführungsbezogen -unternehmensintern
1999	Deckers	Bestandsaufnahme	keine	-Kritische Prüfung und Weiterentwicklung des „Testkundenverfahrens“ als	-durchführungsbezogen

		unter methodischen Gesichtspunkten		Messinstrument -Der Fokus der Arbeit liegt auf eine Analyse der Methodenprobleme und einer Ableitung einer angemessenen Vorgehensweise. -Detaillierte theoretische Bestandsaufnahme bis zum Jahr 1997	
1999	Finn, Kayandé	Vergleich von Mystery Shopping und Kundenbefragung unter psychometrischen Gesichtspunkten	Einzelhandel	-Die mit Mystery Shopping erhobenen Daten erwiesen sich als reliabel -Die erhobenen Daten haben eine höhere Reliabilität als die von Kundenbefragungen -Die Reliabilität der Daten ist bei der Bewertung des Ladengeschäfts (etwa „hard facts“) höher als die bei der Bewertung von Serviceaspekten (etwa „soft facts“) -Reliable Beurteilungen, die mit Hilfe von Mystery Shopping generiert werden, sind kostengünstiger als die von Kundenbefragungen zu erheben, selbst wenn zehn mal höhere Kosten pro Datensatz angenommen werden -Zur Beurteilung des Ladengeschäfts sind zwölf Datensätze, für die Beurteilung von Servicequalität 40-60 Datensätze erforderlich	-durchführungsbezogen -unternehmensintern
2000	Matzler, Kittinger-Rosanelli	Mystery Shopping als Instrument zur Messung der wahrgenommenen Dienstleistungsqualität von Banken	Banken	-Mystery Shopping ist geeignet, das tatsächliche Verhalten von Mitarbeitern in einer Kundenkontaktsituation zu eruieren. -Es liefert ein umfassendes Bild der Dienstleistungsqualität -Es eignet sich, um die konkrete und subjektive Qualitätswahrnehmung des Kunden zu erfassen -Es ist in der Lage, die Qualitätswahrnehmung des Kunden reliabel und valide zu messen. -Ferner werden Anhaltspunkte zur praktischen Durchführung gegeben.	-durchführungsbezogen -unternehmensintern
2000	Matzler, Pechlaner,	Überprüfung von Servicestandards bei	Tourismus:	-Übertragung der Erkenntnisse von Matzler/Kittinger-Rosanelli auf	-durchführungsbezogen

	Kohl	touristischen Dienstleistungen durch „Mystery Guests“	Hotels	touristische Dienstleistungen -Formulierung von Anforderungen für den erfolgreichen Einsatz der Methode	-unternehmensintern
2000	Dawes, Sharp	Reliabilität und Validität von Mystery Shopping-Erhebungen	Einzelhandel	-Die mit Mystery Shopping erhobenen Daten erwiesen sich als reliabel -Kriteriums- und Konvergenzvalidität der Daten kann belegt werden -Es wird eine hohe Stabilität der Komponenten, die eine objektive Dienstleistungsqualität bestimmen, festgestellt -Es wird ein hoher Stichprobenfehler für Mystery Shopping festgestellt -Es wird empfohlen, Mystery Shopping nicht einmalig und auf einzelne Filialen zu beziehen, sondern aggregierte Daten für Regionen und/oder mehrere Zeitpunkte auszuwerten -Die Analyse der gewonnenen Informationen sollte dabei auf ein Gesamtmaß (wie gut sind wir?), Abweichungsgrad (z.B. zwischen den Filialen), Stärken und Schwächen sowie eine Veränderung der Werte über Zeit abzielen	-durchführungsbezogen -unternehmensintern
2001	Anderson, Groves, Lengfelder, Timothy	Einsatz von Mystery Shopping als Bestandteil von Personaltrainings	Tourismus	Tester erhöhen ihre persönliche Fähigkeit Entscheidungen zu treffen und Probleme (im Sinne von Problemlösungstechniken) zu lösen	-unternehmensintern
2001	Finn	Betrachtung der Verkaufsfähigkeit und des Erscheinungsbildes von Ladenlokalen bei Gebrauchsgüter-Filialisten und	Einzelhandel	-Wirtschaftlicher bei Einzelgeschäften als bei Filialisten umzusetzen -Testen von Verkaufsfähigkeit benötigt doppelt so viele Tests wie die des Erscheinungsbildes bei Einzelgeschäften und fünf mal so viele bei Filialisten -Im Vergleich zu Finn/Kayandé (1999) wird festgestellt, dass zur Beurteilung der Servicequalität für Konsumgütergeschäfte mehr Tests benötigt werden als zur Beurteilung der Verkaufsfähigkeit für	-durchführungsbezogen -unternehmensextern

		Einzelgeschäften im Leistungsvergleich mit dem Wettbewerb		Gebrauchsgütergeschäfte -Mystery Shopping wird als „brauchbare Alternative zu Kundenbefragungen im Leistungsvergleich“ angesehen	
2001	Haas	Ermittlung von Faktoren, die Beratungszufriedenheit beeinflussen	Einzelhandel	-Testkäufer werden hier eingesetzt, um Entstehung und Auswirkung von Beratungszufriedenheit zu ermitteln. -Die Methode wird nicht untersucht, sondern dient in der Studie lediglich als Vehikel, um Daten zu erheben.	-durchführungsbezogen
2002	Tribe	Betrachtung von Mystery Shopping zur Erhebung von Kundenzufriedenheit in der Tourismus- branche anhand von zwei rudimentären Fallstudien	Tourismus	-Als Zielsetzung von Mystery Shopping wird die Informationsgewinnung für Kunden und Management gesehen, die als Grundlage der Selbstbewertung von Tourismusunternehmen dient -Rudimentäre Diskussion unterschiedlicher Vorgehensweisen (strukturiert/unstrukturiert, fremde/eigene Tester) anhand von Stärken und Schwächen	-durchführungsbezogen -unternehmensintern
2002	Shing, Spence	Ethische Perspektive bei Wettbewerbs- beobachtungen durch Mystery Shopping	keine	Mystery Shopping im Rahmen von Wettbewerbsbeobachtungen wird anhand einer Argumentation als ethisch fragwürdig beschrieben	-durchführungsbezogen -unternehmensextern
2002	Haas	Diskussion über die Eignung von Mystery Shopping zur Messung und Analyse von Verkaufsgesprächen	Einzelhandel	-Mystery Shopping erweist sich als effektive Methode der Datengewinnung bei Verkaufssituationen -Formulierung von Anforderungen, um qualitativ hochwertige Datenbasis generieren zu können -Aufzeigen von Einsatzmöglichkeiten in der Verkaufsforschung -Es wird festgestellt, dass kein weiteres Verfahren existiert, das eine	-durchführungsbezogen -unternehmensintern

				vergleichbar umfassende, präzise und konsistente Datenbasis über Verkaufssituationen aus Kundensicht generieren kann	
2003	Haas	Untersuchung der Validität der Urteile von Mystery Shoppern anhand eines Vergleichs mit Kundenbefragungen	Einzelhandel	<ul style="list-style-type: none"> -Urteile von Mystery Shoppern lassen sich auch bei komplexeren Sachverhalten mit denen von echten Kunden vergleichen -Die auf die Beratungszufriedenheit von Kunden wirkenden Einflüsse wirken auch im Falle der Mystery Shopper -Es kann sinnvoll sein, bei Mystery Shopping die Operationalisierungen inhaltlicher Aspekte zu verändern -Eine Beschränkung auf die Messung objektiver Sachverhalte wird der Leistungsfähigkeit der Methode nicht gerecht 	-durchführungsbezogen
2003	Beck, Miao	Untersuchung der Einsatzgebiete von Mystery Shopping in der Hotelbranche in den USA	Tourismus: Hotels	<ul style="list-style-type: none"> -Ableitung von Aussagen über praktische Umsetzung von Mystery Shopping und Verwendung der Daten auf Basis einer Befragung von n=82 Geschäftsführern und Finanzleitern von Hotels (>300 Betten) -Mystery Shopping wird eher eingesetzt, um den Prozess der Dienstleistungserbringung zu beschreiben, als lediglich das Ergebnis einer Dienstleistungssituation abzubilden -Die Durchführung erfolgt bevorzugterweise durch externe Anbieter -Inhaltlich herrscht eine Balance zwischen der Erhebung von Dienstleistungsaspekten und objektiven Standards (z.B. Zustand der Räumlichkeiten) 	<ul style="list-style-type: none"> -durchführungsbezogen -unternehmensintern -unternehmensextern
2003	Lingenfelder, Wieseke, Schmidt	Modellierung und Messung von Dienstleistungsqualität in Reisebüros mittels Mystery Shopping	Tourismus: Reisebüros	<ul style="list-style-type: none"> -Erläuterung des Vorgehens bei der Erhebung von Daten mittels Mystery Shopping zur Messung der Servicequalität von Reisebürodienstleistungen -Keine nennenswerten Erkenntnisse über die Methode Mystery Shopping 	-durchführungsbezogen

2003	Hesselink, van der Wiele	Mystery Shopping als Instrument zur Messung von Kundenzufriedenheit	Zeitarbeit	<p>-Mystery Shopping wird als gute Ergänzungsmethode für Kundenbefragungen angesehen.</p> <p>-Sowohl quantitative als auch qualitative sowie objektive und subjektive Daten können mit der Methode generiert werden.</p> <p>-Es wird eine transparente Vorgehensweise in Verbindung mit einer internen Kommunikation vorgeschlagen, um Bewusstsein bei den Mitarbeitern zu schaffen. Es wird vorgeschlagen, dies gründlicher zu untersuchen.</p>	<p>-durchführungsbezogen</p> <p>-unternehmensintern</p>
2004	Hesselink, van Iwaarden, van der Wiele	Mystery Shopping als Instrument zur Einsicht in die Dienstleistungserbringung	Zeitarbeit	<p>-Kundenzufriedenheitsdaten aus Befragungen werden zur Einsicht in die Dienstleistungsqualität und zum Anstoßen von erforderlichen Veränderungsprozessen als nicht ausreichend angesehen</p> <p>-Mystery Shopping wird als sinnvolles Instrument zur effektiven und effizienten Begleitung von Veränderungsprozessen zur Verbesserung der Dienstleistungserbringung beschrieben</p>	-durchführungsbezogen
2005	Lingenfelder, Schmidt, Wieseke	Modellierung und Messung von Dienstleistungsqualität in Reisebüros mittels Mystery Shopping (analog 2003)	-Tourismus: Reisebüros	<p>-Weitgehend inhaltlich identisch mit Beitrag aus 2003</p> <p>-Zusatz: Diskussion der Analyse der Datengüte (insb. Reliabilität) der mittels Mystery Shopping erhobenen Daten</p> <p>-Feststellung einer Forschungslücke: Prüfung der Validität und Reliabilität der Methode Mystery Shopping</p>	-durchführungsbezogen
2005	Calvert	Untersuchung des Verfareneinsatzes in öffentlichen Büchereien	Büchereien	<p>-Identifizierung von drei Aspekten, die den Einsatz von Mystery Shopping in öffentlichen Bereichen begründen: Evaluierung des Kundenservice, Verbesserung von Mitarbeiterverhalten gegenüber dem Kunden und Wettbewerbsvergleich</p> <p>-Feststellung, dass Mitarbeiter den Einsatz von Mystery Shopping anfänglich ablehnen und nach Ergebnispräsentation befürworten</p>	<p>-unternehmensintern</p> <p>-unternehmensextern</p>

2006	Semel	Bewertung des Verfahrens hinsichtlich seines Potenzials im Rahmen des Qualitätsmanagements von Dienstleistungen	keine	<ul style="list-style-type: none"> -Feststellung der Eignung der Methode, um theoretische und wahrnehmungsrelevante Aspekte der Dienstleistungsqualität zu erfassen und zu analysieren -Bei der Analyse von Kundenschnittstellen eignet sich das Instrument als Ergänzung zu subjektiven Analyseinstrumenten -Annahme: Einsatz von ungeschulten Testern wird abgelehnt, da keine verlässlichen Ergebnisse zu erwarten sind 	-durchführungsbezogen
2006	Wieseke, Schmidt, Lingenfelder (Artikel) sowie	Leistungspotenzial von Mystery Shopping	-Tourismus: Reisebüro	<ul style="list-style-type: none"> -Aufsetzen auf der Studie von 2003 -Es besteht eine hohe Beobachtungsübereinstimmung zwischen Testkäufern hinsichtlich zahlreicher relevanter Faktoren der Dienstleistungsqualität -Diagnosen von Testkäufern sind dauerhaft stabil und verlässlich -Es besteht ein signifikanter Zusammenhang zwischen positiver Beurteilung und Verkaufserfolg von Mitarbeitern -Rudimentäre Ableitung von Richtlinien zur erfolgreichen Durchführung von Mystery Shopping 	-durchführungsbezogen -unternehmensintern
2007	Schmidt (Dissertation)				

Quelle: Eigene Darstellung

2.1.4.2 Erkenntnisstand vergangener Forschungsbemühungen

Nach näherer Betrachtung der zuvor aufgeführten wissenschaftlichen Beiträge lassen sich folgende Aussagen treffen:

- Als Begriff für Verfahren, die den Einsatz von Testkunden subsumieren, hat sich *Mystery Shopping* durchgesetzt.
- Die Beiträge widmen sich jeweils Einzelaspekten. Mystery Shopping wird nicht in einem Gesamtzusammenhang betrachtet.
- Es existiert keine umfassende Übersicht und kritische Auseinandersetzung der bisher publizierten Literatur. Eine Ausnahme bildet der Beitrag von Deckers.¹⁴⁷ Gleichwohl sind insbesondere nach seiner Veröffentlichung zahlreiche Beiträge entstanden, die dort natürlich nicht erfasst worden sind. Die vorliegende Arbeit versucht, diese Lücke zu schließen.
- Es gibt keinen Beitrag, der sich als Grundstein einstufen ließe. Ferner ist keine stringente aufeinander aufbauende Entwicklung nachzuvollziehen. Vielmehr scheinen die Beiträge weitgehend unabhängig voneinander entstanden zu sein, ohne auf den Erkenntnissen vorangegangener Forschungsbemühungen aufzubauen.
- Mit 21 Artikeln wird das Thema mehrheitlich im Kontext einer bestimmten Branche betrachtet, wobei Einzelhandel und Tourismus mit jeweils acht Publikationen am häufigsten vertreten sind. Während der Einzelhandel eine klassische Mystery Shopping-Branche darstellt, in der die Methode ihren Ursprung hatte, ist die Vielzahl an Publikationen aus dem touristischen Sektor verwunderlich. Zu erklären ist dies mit mehrfach publizierten Beiträgen aus der gleichen Forschungsgruppe.¹⁴⁸
- 26 Beiträge schildern das Vorgehen bei der Durchführung, wobei nur zum Teil Umsetzungsempfehlungen abgegeben werden. Es wird jedoch nicht deutlich, welche Vorgehensweise sich besonders bewährt hat und sich auf andere Unternehmen übertragen ließe. Möglicherweise hängt dies damit zusammen, dass Mystery Shopping überwiegend als Auftragsforschung durchgeführt wird. Durchführende Unternehmen wollen ihr Vorgehen bewahren, um ihr Know-how dem Wettbewerb nicht preiszugeben. Die Folge ist eine Anhäufung spezialisierten Wissens bei Praktikern, das ohne fehlende Diskussion und Prüfung nur schwer weiterentwickelt werden kann.
- Neun Artikel widmen sich der Methodik im Sinne einer Betrachtung von Messansätzen, Messgenauigkeit und der Vermeidung von Messfehlern. Es ist zu

¹⁴⁷ Vgl. Deckers (1999)

¹⁴⁸ So fußen die Publikationen von Lingenfelder/Wieseke/Schmidt (2003), Lingenfelder/Schmidt/Wieseke (2005), Wieseke/Schmidt/Lingenfelder (2006) und Schmidt (2007) auf derselben Studie. Für einen vollständigen Überblick des Studienverlaufs siehe Schmidt (2007)

konstatieren, dass insbesondere Reliabilität und Validität der Methode nachgewiesen worden sind. Gleichwohl erfolgte dieser Nachweis erst im Verlauf der letzten acht Jahre, rund ein halbes Jahrhundert nach bewährtem Einsatz der Methode.

- In unterschiedlichen Beiträgen wird die Eignung, theoretische und wahrnehmungsrelevante Aspekte der Dienstleistungsqualität zu erfassen und zu analysieren, nachgewiesen.
- Unterschiedliche Beiträge, die sich mit einem Vergleich von Mystery Shopping und Kundenbefragungen auseinandersetzen, konstatieren, dass sich Mystery Shopping hinsichtlich Ergebnisqualität, Leistungsfähigkeit und anfallender Kosten als vorteilhafter erwiesen hat. Ferner wird eine Kopplung von Mystery Shopping mit Kundenbefragungen empfohlen. Eine Bestätigung oder Widerlegung der Vorteilhaftigkeit des kombinierten Einsatzes erfolgt jedoch nicht.
- Hinsichtlich des Grads der Schulung der Testpersonen herrscht Uneinigkeit. Es ist nicht belegt, ob der Einsatz vollkommen ungeschulter Tester, teilweise geschulter Tester oder vollständig geschulter Tester die sinnvollste Vorgehensweise darstellt.
- In diesem Zusammenhang wird auch nicht deutlich, ob sich soziodemografische Faktoren auf die Güte der Ergebnisse auswirken und ob sich bestimmte soziodemografische Eigenschaften möglicherweise als vorteilhaft beim Einsatz als Mystery Shopper erweisen würden.
- Eine Verwendung der mittels Mystery Shoppings gewonnenen Ergebnisse in Folgeprozessen, die eine nachhaltige Verbesserung der Leistungsqualität eines Unternehmens ermöglichen würden, wird von verschiedenen Autoren propagiert. Allerdings sind keine Hinweise auf eine Systematik vorhanden, die es Unternehmen ermöglichen könnte, Mystery Shopping im Sinne eines Steuerungsinstruments effektiv und effizient einzusetzen.
- Eine Darstellung des Nutzens erfolgt lediglich qualitativ über die Eignung des Instruments für unterschiedliche Einsatzgebiete. Eine Quantifizierung im Sinne einer Kosten-Nutzen-Betrachtung erfolgt an keiner Stelle.
- Eine Diskussion der Durchführung in Form einer internen oder externen Leistungserstellung erfolgt an keiner Stelle.

Nachdem die theoretischen Grundlagen der Methode Mystery Shopping dargestellt wurden, wird im folgenden Abschnitt die praktische Umsetzung erläutert.

2.2 Mystery Shopping aus praktischer Perspektive

Im Folgenden wird der Einsatz des Instrumentes in der Praxis geschildert. Nach einer Betrachtung des Marktes wird die Umsetzung anhand der möglichen Zielsetzungen, Einsatzmöglichkeiten und anhand der Vorgehensweise in einem Projektverlauf von der Konzeption bis zur Verwendung der gewonnenen Informationen dargelegt.

2.2.1 Marktgröße und Marktumfeld

Nach Angaben des europäischen Verbandes, der Mystery Shopping Providers Association (MSPA), belief sich das Volumen, das 2005 europaweit von Unternehmen für Mystery Shopping-Dienstleistungen ausgegeben wurde, auf schätzungsweise 210 Mio. €. ¹⁴⁹ Der Verband geht von einem durchschnittlichen jährlichen Wachstum i.H.v. rund 11% auf 360 Mio. € im Jahre 2010 aus, wobei die durchschnittliche jährliche Wachstumsrate in Deutschland auf rund 8% geschätzt wird. Angaben für einzelne Länder werden nicht gemacht. Das Umsatzvolumen für Nordamerika wird auf rund 600 Mio. € für das Jahr 2005 geschätzt. Die Korrektheit dieser Angaben ist jedoch fraglich, denn europaweit wurden lediglich 41 Mitglieder des Verbandes befragt, unter ihnen nur eines aus Deutschland.

Schätzungen über das deutsche Marktvolumen unterscheiden sich bei einer Durchsicht von Interviews und Zeitungsartikeln aus der Vergangenheit. So wurde 2001 das Marktvolumen für 2005 in Deutschland auf 125 Mio. € geschätzt. ¹⁵⁰ Zwei Jahre später wird das Marktvolumen auf 20-30 Mio. € geschätzt, und es wird von einer Verdoppelung bis 2005 ausgegangen. ¹⁵¹ 2005 wird das deutsche Marktvolumen auf einer Bandbreite zwischen 30 und 50 Mio. € eingestuft. ¹⁵² Diese Relativierungen im Zeitverlauf machen deutlich, dass die Entwicklung den Prognosen nicht annähernd nachkommt. Vorsichtiger Schätzungen gehen von einer Entwicklung des Marktvolumens auf 75 Mio. € im Jahre 2010 aus. Demnach soll dann der Gipfel der Entwicklung erreicht sein. ¹⁵³ Bezogen auf das Marktvolumen von Marktforschungsausgaben 2005 in Deutschland ¹⁵⁴ sind dies 4,3%.

Die unterschiedlichen Angaben deuten zudem auf ein wenig transparentes Marktumfeld hin. Weder das Marktvolumen noch die Anzahl der Unternehmen, die in Deutschland (aktiv) Mystery Shopping-Dienstleistungen anbieten und letztlich auch durchführen, sind genau bekannt. Einen Anhaltspunkt über die Entwicklung liefern Daten des Bundesverbandes

¹⁴⁹ Vgl. hierzu und im Folgenden die Pressemitteilung der MSPA vom 19.05.2006

¹⁵⁰ Vgl. Lorentz (2001)

¹⁵¹ Vgl. Geißler (2003)

¹⁵² Vgl. Dostert (2005)

¹⁵³ Vgl. Deckers/Heinemann (2006: 19)

¹⁵⁴ Der Marktforschungsumsatz betrug 2005 in Deutschland 1.758 Mio. €, Vgl. <http://www.adm-ev.de/zahlen.html> (Stand: 24.10.2007)

Deutscher Markt- und Sozialforscher (BVM). Während 1997 lediglich drei Unternehmen als Anbieter von Mystery Shopping-Dienstleistungen beim BVM registriert waren,¹⁵⁵ belief sich die Anzahl 2007 auf 20 Unternehmen.¹⁵⁶ Allerdings ist anzunehmen, dass längst nicht alle Unternehmen dort Mitglied sind oder sich Anbieter registriert haben, denn eine 2005 von diesem Autor durchgeführte Internetrecherche ergab eine Anzahl von 44 Unternehmen, die Mystery Shopping-Dienstleistungen anbieten. Hinsichtlich des Marktumfeldes ist noch weniger bekannt, wie viele Unternehmen Mystery Shopping zur Überprüfung ihrer Dienstleistungsqualität über einen Anbieter oder in Eigenregie durchführen.

Der 2001 gegründete „Bundesverband Mystery Shopping“ ist im Mai 2007 aufgelöst worden.¹⁵⁷ Somit ist ein Markt entstanden, bei dem keinerlei qualitative Voraussetzungen für den Eintritt oder Richtlinien für die Durchführung existieren. Ein Gütesiegel, wie ihn Berufsverbände (z.B. der Bundesverband Deutscher Unternehmensberater) durch entsprechende Grundsätze festlegen, ist nicht existent. Grundsätzlich kann folglich jedes Unternehmen seine Dienstleistungen in dem Bereich anbieten, ohne dass potenzielle Auftraggeber die Güte des Unternehmens einschätzen können. Fraglich ist, ob die eingesetzten Methoden der Anbieter Nutzen bei den jeweiligen Auftraggebern stiften. Der mögliche Prozess einer mittel- oder langfristigen Bereinigung des Marktes aufgrund der Verdrängung von weniger professionellen Anbietern könnte auch zu einer Ablehnung der Methode durch den Markt führen. Schließlich erwartet ein Auftraggeber einen entsprechenden Nutzen, der möglicherweise von weniger professionellen Anbietern nicht oder nur teilweise erbracht werden kann.

In einer 2005 von dem Autor dieser Arbeit durchgeführten Stichprobe auf Basis einer Internetrecherche lassen sich die Anbieter von Mystery Shopping-Dienstleistungen in drei Gruppen aufteilen.¹⁵⁸

- Reine Mystery Shopping Unternehmen (63,6%)
- Unternehmensberatungen, die u.a. Mystery Shopping anbieten (20,6%)
- Marktforschungsunternehmen, die u.a. auch Mystery Shopping anbieten (15,9%)

Die Aktivitäten von Unternehmensberatungen auf diesem Gebiet lassen eine Ausdehnung ihrer Wertschöpfungskette vermuten. Zum einen können so „researchnahe“ Leistungen wie Mystery Shopping selbst angeboten werden, zum anderen könnten die erhobenen Daten

¹⁵⁵ Vgl. Deckers (1999: 13)

¹⁵⁶ Vgl. <http://www.bvm-net.de/bvm-net-suche.html> im Suchfeld Schlagworte nach „Mystery-Shopping“ (Stand 13.09.2007)

¹⁵⁷ Vgl. <http://www.bvms.de> (Stand: 20.05.2007); Nach mündlichen Angaben führten unterschiedliche Auffassungen der Mitgliedsunternehmen zur Trennung.

¹⁵⁸ Basis sind die Internetpräsenzen von 44 Unternehmen (Stand 20.08.2005)

zusätzliche Beratungsleistungen in Form von Folgeaufträgen einbringen. Erstaunlich ist der geringe Anteil von Marktforschungsunternehmen. Schließlich ist Mystery Shopping eine marktforschungsnahe Dienstleistung, die auf gegebenen Strukturen aufsetzen kann (etwa Testerstab aus Marktforscherstab rekrutieren, bestehenden Kundenstamm für Mystery Shopping gewinnen, Know-how bei Auswertungen von Befragungsdaten).

Eine weitere Unterteilung der Anbieter ist denkbar: Ein Unternehmen kann Mystery Shopping anbieten, ohne den Test selbst durchzuführen. Agenturen, die über einen großen Testerstab verfügen, generieren durch die Überlassung der Tester an andere Unternehmen zusätzliche Umsatzerlöse und lasten dadurch ihre Kapazitäten stärker aus. Dies könnte zudem Tester stärker an ihr Unternehmen binden, da sie so häufiger die Möglichkeit haben, als Tester eingesetzt zu werden. Die Anbieter ohne eigenen Testerstab sind wiederum nicht auf Ausbau und Pflege des Testerstabs angewiesen und haben ein geringes Markteintrittsrisiko.¹⁵⁹

2.2.2 Zielsetzungen, Einsatzmöglichkeiten und Vorgehen bei Mystery Shopping

Einsatz und Vorgehensweise bei Mystery Shopping hängen mit der angestrebten Zielsetzung zusammen. In der überwiegend praxisnahen Literatur sind unterschiedliche Zielsetzungen und Einsatzgebiete dokumentiert. Es herrscht uneingeschränkte Übereinstimmung darüber, dass Schwachstellen und somit Verbesserungspotenziale im Prozess der Erstellung von Dienstleistungen aufgedeckt werden und so handlungsrelevante Informationen gewonnen werden können.¹⁶⁰ Das Instrument ist insbesondere geeignet, um die Einhaltung von Vorgaben zu überprüfen (z.B. Servicestandards¹⁶¹, Erscheinungsbild, Platzierung von Produkten und Werbemitteln im Einzelhandel¹⁶², gesetzliche Vorgaben des Wertpapierhandelsgesetzes¹⁶³, Rassendiskriminierung bei Kreditvergaben¹⁶⁴, Verhalten bei Kassivorgängen¹⁶⁵). Die Daten werden an der Kundenschnittstelle erhoben (z.B. Point of Sale¹⁶⁶, Call Center¹⁶⁷, Schriftverkehr¹⁶⁸, Außendienst¹⁶⁹) und können die Basis für inner-

¹⁵⁹ Basierend auf diesen Überlegungen hat der Autor dieser Arbeit unter <http://www.mysterypanel.de> eine Datenbank eingerichtet, bei der sich interessierte Personen als Mystery Shopper registrieren lassen können und die von solchen Marktforschungsunternehmen genutzt werden kann, die über kein eigenes Testkundenfeld verfügen.

¹⁶⁰ Vgl. z.B. Platzek (1998: 44), Wilson (1998b: 419), Cook et al. (2002: 170), Zenz/Schweiger (2003: 3f.), Michel/Prykop (2006: 52), Schlich (2006: 274), Dobbelstein/Windbacher (2007: 107)

¹⁶¹ Vgl. Czopek (1998), Matzler/Kittinger-Rosanelli (2000), Almhofer (2007: 35)

¹⁶² Vgl. Holtz (2003: 76), Pergandé/Stücken (2004: 207), Turner (2007: 334)

¹⁶³ Vgl. Duttenhöfer/Lerch (2000: 31), Müller/Neuhaus/Löhnert (2006)

¹⁶⁴ Vgl. Leeds (1995: 20)

¹⁶⁵ Vgl. Deges (1992: 87)

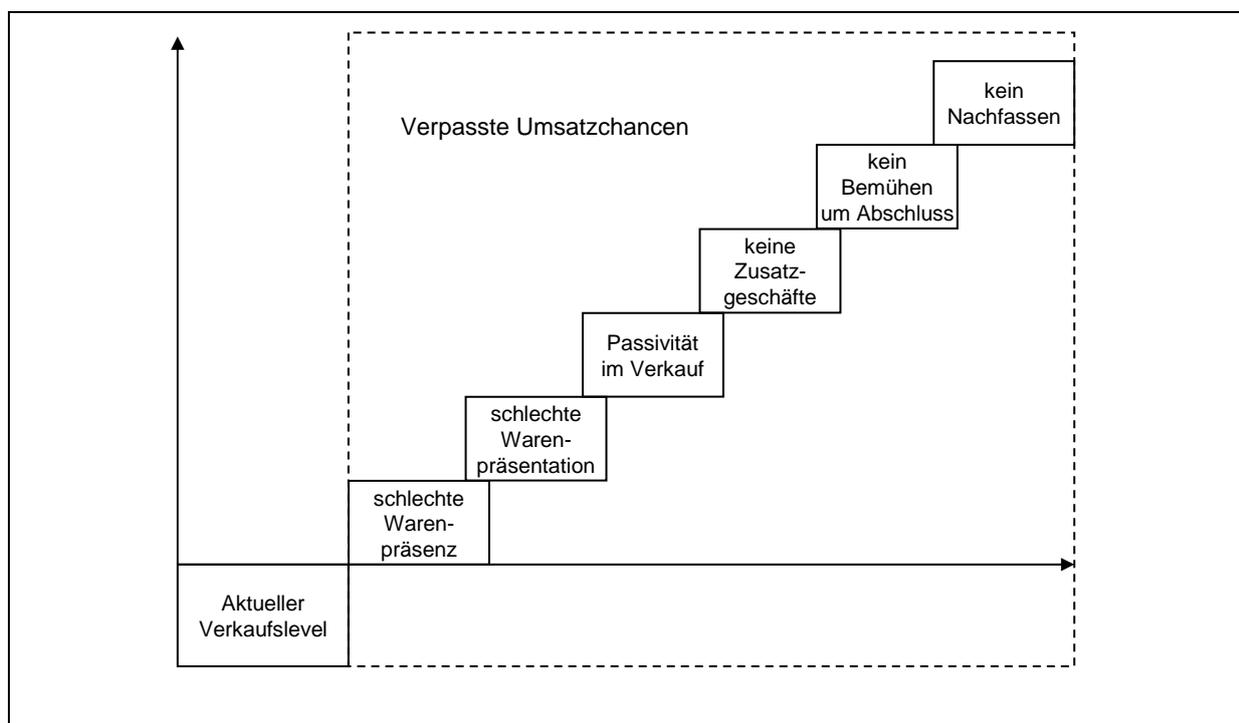
¹⁶⁶ Vgl. z.B. Höhner/Schaper (2004), Schrader (2006)

¹⁶⁷ Vgl. Beckmann (2003), Bauer/Schneider/Herrmann (2002), Gohr (2006), Turner (2007: 338)

¹⁶⁸ Vgl. Geißler/Thomas/Duscha (2005)

und außerbetriebliche Leistungsvergleiche bilden¹⁷⁰, mit Anreizsystemen gekoppelt werden¹⁷¹, als Basis für Schulungsmaßnahmen dienen¹⁷², zur Messung von Service-¹⁷³ oder Beratungsqualität¹⁷⁴ eingesetzt werden, als Ausgangspunkt zur Qualitätsentwicklung¹⁷⁵ verwendet werden oder, wie im Falle der Londoner U-Bahn, als Grundlage für die Entscheidungsfindung bei Investitionen dienen¹⁷⁶. Daneben können die Ergebnisse als Basis für Pressemitteilungen oder die PR-Arbeit dienen.¹⁷⁷ Ferner kann Mystery Shopping insbesondere eingesetzt werden, um konkrete Anhaltspunkte für zusätzliche Umsatzchancen im Verlauf eines Verkaufsprozesses zu gewinnen. Dies wird in der Abbildung 2-2-1 verdeutlicht.

Abbildung 2-2-1: Verpasste Umsatzchancen



Quelle: In Anlehnung an Deckers/Heinemann (2006: 24)

Zudem sind weitere Einsatzfelder denkbar, bei denen es nicht um den Kontakt mit Kunden, sondern um den mit anderen Stakeholdern eines Unternehmens geht, wie z.B. zukünftige

¹⁶⁹ Vgl. Guber/Kerper (2003)

¹⁷⁰ Vgl. z.B. Matzler/Kittinger-Rosanelli (2000: 222), Drees/Schiller (2001: 6), Douglas/Douglas (2006)

¹⁷¹ Vgl. Erstad (1998: 37)

¹⁷² Vgl. Keller (1982: 215)

¹⁷³ Vgl. Beck/Miao (2003), Calvert (2005: 24)

¹⁷⁴ Vgl. Schulte/Prins (2007)

¹⁷⁵ Vgl. Dreyer/Linne (2004)

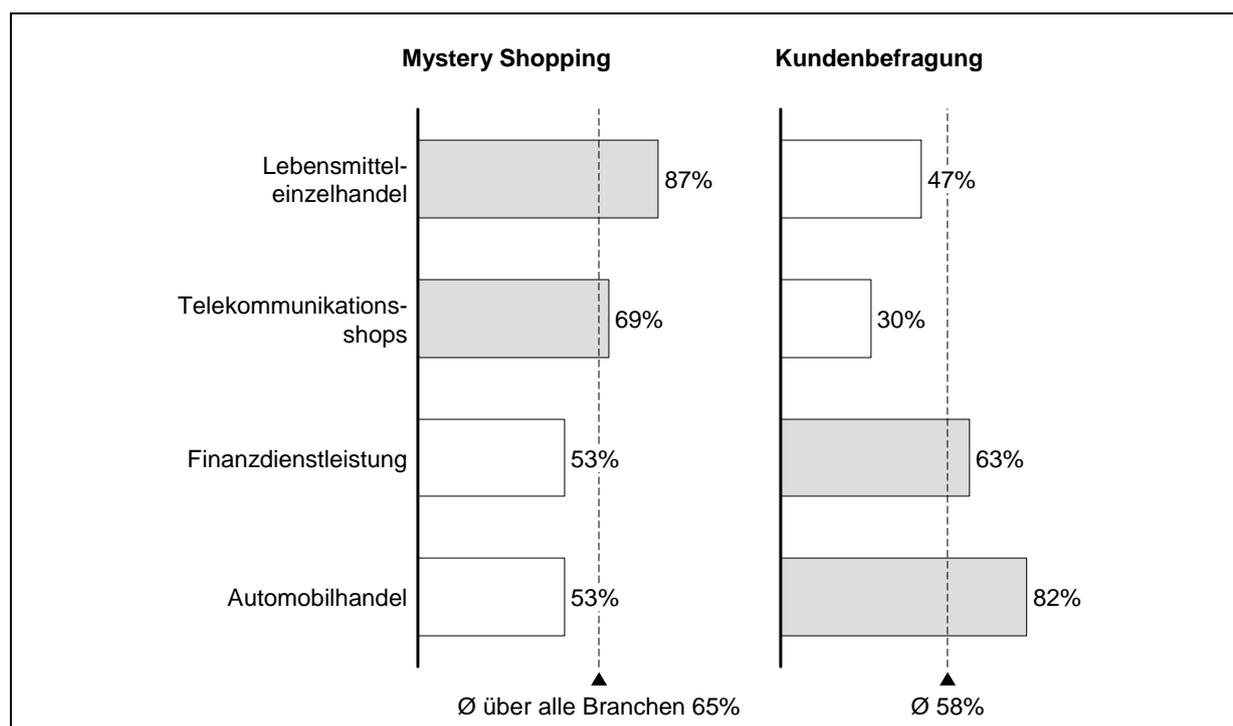
¹⁷⁶ Vgl. Wilson/Gutmann (1998)

¹⁷⁷ Vgl. Almhofer (2007: 37)

Mitarbeiter, die das Verhalten der Personalabteilung bei Einstellungsgesprächen beurteilen, Journalisten, die das Auftreten der Abteilung für Öffentlichkeitsarbeit bewerten oder Investoren, die die Abteilung für Investor Relations einschätzen. Darüber hinaus könnten in Franchise-Unternehmen die Umsetzung zentraler Vorgaben durch den Franchise-Nehmer überprüft werden. Nur schwer denkbar wäre ein Einsatz auf der Lieferantenseite. Zum einen wäre die Rekrutierung von entsprechenden Testern nur schwer möglich, zum anderen stellt sich die Frage nach dem Nutzen.¹⁷⁸

Hinsichtlich des branchenbezogenen Einsatzes scheint es kaum Einschränkungen zu geben, wie im Rahmen der Entwicklung bereits dargelegt wurde.¹⁷⁹ Es existieren nur wenige Anhaltspunkte, die auf einen branchenspezifischen Einsatzgrad von Mystery Shopping schließen lassen. Lediglich eine praxisnahe Studie hat sich mit dem Einsatz von Mystery Shopping im Vergleich zu anderen Marktforschungsinstrumenten in unterschiedlichen Branchen des Einzelhandels auseinandergesetzt.¹⁸⁰ Der Einsatzgrad der vier untersuchten Branchen ist in Abbildung 2-2-2 dargestellt.

Abbildung 2-2-2: Einsatz von Mystery Shopping und Kundenbefragung im Branchenvergleich



Quelle: Vgl. infas TTR (2005: 75), n=370

¹⁷⁸ Gleichwohl existiert eine Auftragsstudie, in der die Pre-Sales-Phase für eine Kundenzufriedenheitsbefragung durch professionelle Testkäufer bei entsprechenden Anbietern erhoben wurde; Vgl. Mayer (2007)

¹⁷⁹ Vgl. Abschnitt 2.1.3 der vorliegenden Arbeit

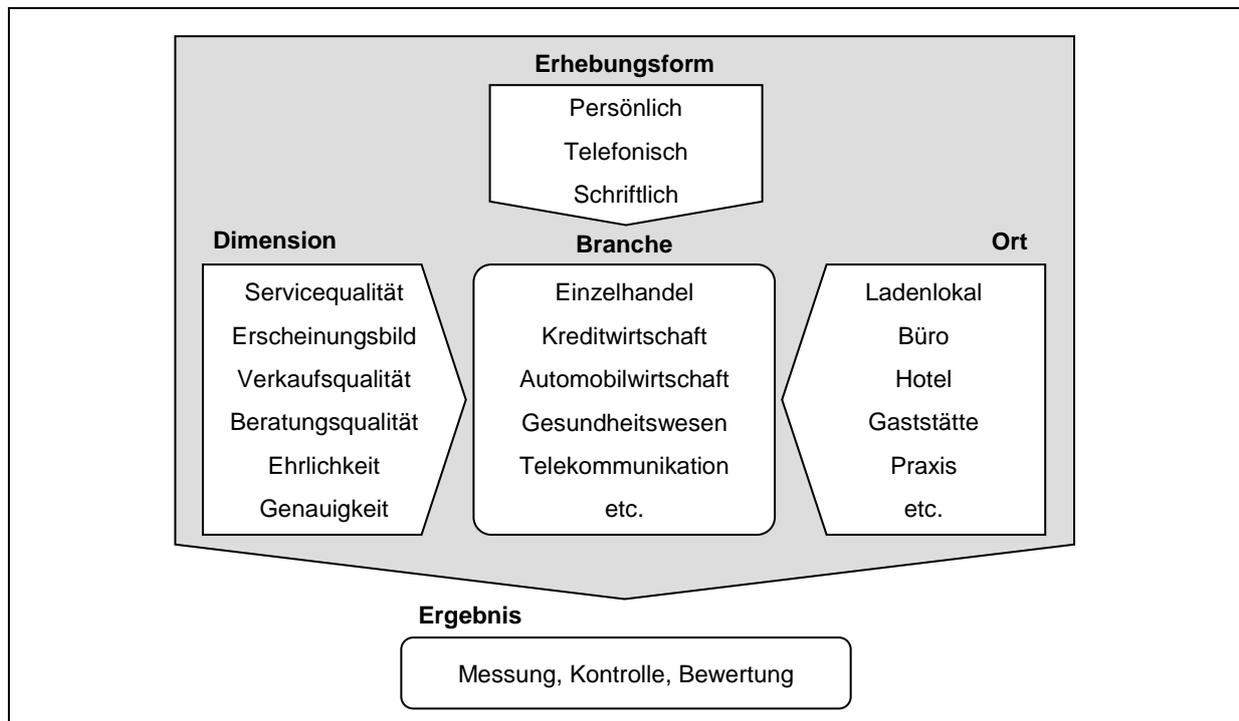
¹⁸⁰ Vgl. infas TTR (2005), die Automobilhändler, Telekommunikationssshops, Lebensmitteleinzelhändler und Finanzdienstleister hinsichtlich ihrer eingesetzten Marktforschungsinstrumente befragten (n=370).

Es wird deutlich, dass Mystery Shopping als Marktforschungsinstrument bei den untersuchten Branchen in der Summe vor der Kundenbefragung den höchsten Einsatzgrad aufweist. Während Lebensmitteleinzelhandel und Telekommunikationsshops mehr auf Mystery Shopping setzen, liegt der Schwerpunkt bei Finanzdienstleistern und im Automobilhandel auf Kundenbefragungen. Einschränkend ist anzumerken, dass lediglich vier Branchen des Einzelhandels untersucht wurden und daher kein Rückschluss auf andere Branchen möglich ist.

Damit der zielgerichtete Einsatz gewährleistet werden kann, ist ein reibungsloser Ablauf erforderlich. Zur Erläuterung sollen dafür nachfolgend die Gemeinsamkeiten unterschiedlicher Vorgehensweisen anhand eines Bezugsrahmens und einer übergeordneten Vorgehensweise im Sinne eines Projektverlaufs skizziert werden.

Gemeinsam haben alle Formen von Mystery Shopping das Auftreten von Testpersonen, die Daten anhand eines vorab definierten Beobachtungskataloges erheben und dokumentieren. Der Katalog bezieht sich auf die zu erhebende Dimension (z.B. Erscheinungsbild) und orientiert sich inhaltlich an den die jeweilige Branche betreffenden Eigenschaften, den Ort der Erhebung, die Erhebungsform und das angestrebte Ergebnis (siehe Abbildung 2-2-3).

Abbildung 2-2-3: Mystery Shopping – Grundsätzlicher Bezugsrahmen



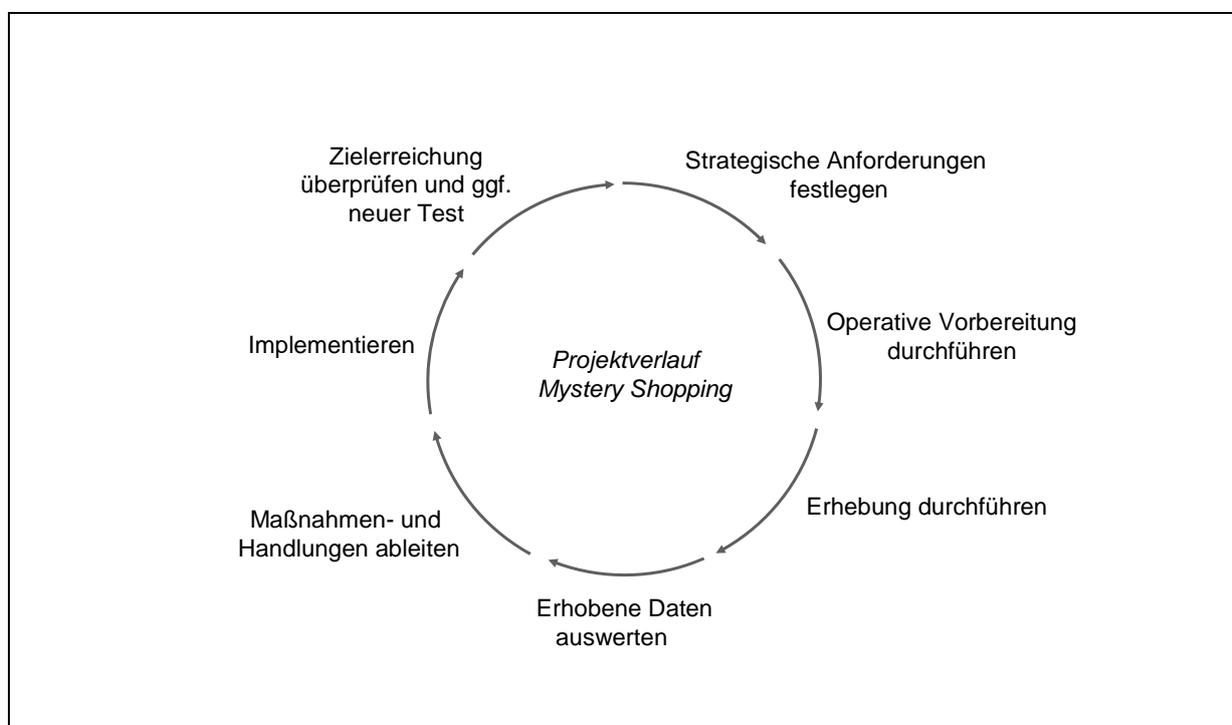
Quelle: Eigene Darstellung

Hinsichtlich der Vorgehensweise beim Mystery Shopping herrscht in der Literatur keine einheitliche Auffassung. Dies ist nicht verwunderlich, denn in der wissenschaftlichen Literatur wird das Instrument entweder nur als Hilfsmittel im Rahmen einer Untersuchung

herangezogen¹⁸¹ oder es werden Einzelaspekte beschrieben, die in der Regel mit der Auswertung der erhobenen Daten und einem Hinweis auf die Ableitung von Maßnahmen enden. Daher gibt es bisher auch keinen Aufschluss über eine entsprechende Systematik, die es Unternehmen ermöglicht, Mystery Shopping im Sinne eines Steuerungsinstruments effektiv und effizient einzusetzen, was in der wissenschaftlichen Literatur mit Erstaunen zur Kenntnis genommen wird.¹⁸² Auch bei den praktisch angesiedelten Beiträgen bleibt es bei einem Verweis auf die Notwendigkeit der weiteren Verwendung der Daten.

Das mögliche übergeordnete Vorgehen im Sinne eines Projektverlaufs einschließlich einer Rückkopplung des Erfolgs von ergriffenen Maßnahmen und Folgeprozessen ist in Abbildung 2-2-4 dargestellt.

Abbildung 2-2-4: Vorgehensmodell von Mystery Shopping als Projektverlauf



Quelle: Eigene Darstellung

Die Überlegungen des dargestellten Vorgehens sind theoretischer Natur, wobei auf Einzelerkenntnisse aus der Literatur zurückgegriffen wurde.¹⁸³

Im ersten Schritt sind die Anforderungen und Ziele festzulegen. Sollen beispielsweise Servicestandards überprüft und ggf. verbessert werden, so sind sie zunächst zu definieren. In

¹⁸¹ so z.B. bei Haas (2002), Haas (2006), Schrader (2006), Roberts/Campbell (2007) zur Analyse von Verkaufssituationen

¹⁸² Vgl. Matzler/Kittinger-Rosanelli (2000: 221), Wieseke/Schmidt/Lingenfelder (2006: 42)

¹⁸³ z.B. Salcher/Esser (1987: 19), Beister (2003), Drees/Schiller (2003: 166), Lammers/Schubert (2003: 60), Eikenroth (2004: 53), Böhm/Fischl/Rennhak (2006), Döbelstein/Windbacher (2007: 109), Turner (2007: 337ff.)

der operativen Vorbereitung ist eine Vielzahl an Aufgaben abzuarbeiten: Der Beobachtungskatalog ist zu formulieren, die Art der Durchführung ist festzulegen (eigene Durchführung vs. Beauftragung eines Fremdunternehmens, Einsatz von Profitestern vs. Laientestern), die Anzahl und die (mögliche) Art der Schulung der Tester ist zu bestimmen, der zeitliche Ablauf ist zu planen und die Rückspielung der Ergebnisse ist festzulegen. Im dritten Schritt wird die Erhebung durchgeführt. Anschließend werden die Daten ausgewertet und Handlungen abgeleitet. Diese werden umgesetzt, und schließlich wird die Zielerreichung überprüft.¹⁸⁴

2.2.3 Konzeption einer Mystery Shopping-Erhebung anhand der einzelnen Phasen eines Projektverlaufs

Für jede der im vorangegangenen Abschnitt genannten Aufgaben der einzelnen Phasen gibt es in der Literatur unterschiedliche Vorschläge zur Ausgestaltung in nahezu jede nur mögliche Richtung. In diesem Abschnitt sollen die Gestaltungsmöglichkeiten in den in Abbildung 2-2-4 dargestellten Phasen anhand der praktischen und wissenschaftlichen Beiträge ergänzt um eigene Überlegungen exemplarisch dargelegt werden, ohne jedoch einen Anspruch auf Allgemeingültigkeit zu erheben. Hiermit wird dem zweiten Teilziel der Arbeit Rechnung getragen, und der konzeptionelle Ansatz soll den Boden für die im Kapitel 3 darzustellende empirische Untersuchung bereiten.

2.2.3.1 Die Zielsetzungsphase einer Mystery Shopping-Erhebung

Im ersten Schritt sind die strategischen Anforderungen festzulegen. Unter strategischen Anforderungen sind Ziele, Themen und Inhalte einer Mystery Shopping-Erhebung zu verstehen. Dabei ist zunächst das Problem bzw. der Untersuchungsgegenstand zu identifizieren und zu konkretisieren. Daraus sind die spezifischen Untersuchungsziele, Themen und Inhalte abzuleiten.¹⁸⁵ So könnte das Problem ein Umsatzrückgang in den Filialen eines Einzelhändlers darstellen. Dieser Rückgang könnte z.B. die Folge niedriger Servicequalität, mangelnder Fachkompetenz oder verschmutzter Räumlichkeiten sein. Somit wären es mögliche Untersuchungsziele, die Servicequalität, die Fachkompetenz und den Zustand der Räumlichkeiten zu erheben. Es wird deutlich, dass die Untersuchungsziele in

¹⁸⁴ Neben diesem Vorgehen sei an dieser Stelle das 1999 eingereichte und 2005 genehmigte US-amerikanische Patent mit der Nummer US 6952679 angeführt, welches in der Klasse „Data Processing: Financial, Business Practice, Management, or Cost/Price Determination“ ein System zur Evaluierung der Servicequalität mittels anonymer Kunden beschreibt; vgl. United States Patent No. US 6952679. Der Erfolg dieser geschützten Methode kann allerdings in Frage gestellt werden, da lediglich dichotome Antwortmöglichkeiten („ja“ oder „nein“) gegeben sind, um unterschiedliche Aspekte des Dienstleistungsgeschehens auszuwerten. Es ist zu vermuten, dass der Patentschutz vielmehr als Gütesiegel einer vermeintlich höheren Qualität eingesetzt wird, indem mit einem „patentierten Ansatz“ geworben wird.

¹⁸⁵ Vgl. Drees/Schiller (2000: 70), Lammers/Schubert (2003: 60), Höhner/Schaper (2004: 33f.)

diesem Falle hypothesengesteuert festgelegt werden. Es kann auch Untersuchungsziel sein, die Einhaltung von zentral vorgegebenen Servicestandards zu überprüfen oder die eigenen mit denen von Wettbewerbern zu vergleichen.¹⁸⁶ Denkbar ist auch, die aus einer Kundenbefragung identifizierten Schwachstellen als Ziele festzulegen.¹⁸⁷

2.2.3.2 Operative Vorbereitung und Erhebungsphase einer Mystery Shopping-Erhebung

Nachdem die Themen identifiziert wurden, sind die Inhalte in einen Beobachtungskatalog zu übertragen. Dabei werden die Themen operationalisiert, so dass der jeweilige Sachverhalt vollständig abgebildet wird.¹⁸⁸ Ferner wird die Datenerhebungsphase vorbereitet.

Die Leistungsfähigkeit eines Mystery Shopping-Konzeptes ist vor allem von der Handlungsrelevanz der gewonnenen Informationen abhängig.¹⁸⁹ Für den erfolgreichen Einsatz der Methode sind im Rahmen dieser Phase vier Anforderungen zu erfüllen.¹⁹⁰

Anforderung 1: Mystery Shopping soll das tatsächliche Verhalten eines Mitarbeiters in einer Kundenkontaktsituation ermitteln.

Anforderung 2: Mystery Shopping soll ein umfassendes Bild der Dienstleistungsqualität liefern.

Anforderung 3: Mystery Shopping soll die konkrete, subjektive Qualitätswahrnehmung eines realen Kunden erfassen.

Anforderung 4: Mystery Shopping soll die Qualitätswahrnehmungen eines Kunden reliabel und valide messen.

Um diese Anforderungen erfüllen zu können, ist die Einhaltung von jeweils unterschiedlichen Voraussetzungen erforderlich. Nachfolgend werden die einzelnen Anforderungen und die zugehörigen Voraussetzungen erläutert.

¹⁸⁶ So behaupten Matzler/Kittinger-Rosanelli, dass die Überprüfung der Einhaltung von Servicestandards (und Konkurrenzanalysen) die häufigsten Anwendungsformen darstellen. Gleichwohl existieren über die Verteilung der Einsatzformen keine gesicherten Erkenntnisse. Siehe Matzler/Kittinger-Rosanelli (2000: 234)

¹⁸⁷ Siehe Abschnitt 2.2.2 für weitere Ziele, Themen und Inhalte

¹⁸⁸ Vgl. z.B. Eikenroth (2004: 54), Dobbstein/Windbacher (2007: 111f.)

¹⁸⁹ Vgl. Schmidt (2007: 93)

¹⁹⁰ Vgl. hierzu und nachfolgend Matzler/Kittinger-Rosanelli (2000: 226ff.), Matzler/Pechlaner/Kohl (2000: 164ff.) und Schmidt (2007: 101ff.). Die Anforderungen und Voraussetzungen für den erfolgreichen Einsatz von Mystery Shopping fußen insbesondere auf den Arbeiten von Matzler et al. Schmidt greift diese ebenfalls auf. In der vorliegenden Arbeit war eine Aufnahme unerlässlich. Dabei wurde die Wiedergabe dieser Grundlagen um zahlreiche Beispiele ergänzt, die vor allem bei den praktischen Beiträgen zu finden sind. Ferner wurden neuartige Überlegungen zu unterschiedlichen Aspekten angestellt.

Anforderung 1: Mystery Shopping soll das tatsächliche Verhalten eines Mitarbeiters in einer Kundenkontaktsituation ermitteln.

Bei Beobachtungen besteht die Gefahr, dass Mitarbeiter ihr Verhalten während einer Kundenkontaktsituation verändern, sofern sie Kenntnis von der Beobachtung haben. Um diesem als „Hawthorne-Effekt“ bezeichnetem Phänomen Rechnung zu tragen und das tatsächliche Verhalten der Mitarbeiter erheben zu können, ist es erforderlich, dass ein Mystery Shopping ohne die Kenntnis der Mitarbeiter im natürlichen, realen Umfeld durchgeführt wird.¹⁹¹

Allerdings ist eine vollständig reale Kundenkontaktsituation nur schwerlich herzustellen; dies hat zwei Gründe: Zum einen wird die Situation durch „Rosenthal-Effekte“ beeinflusst, wobei angenommen wird, dass der Mystery Shopper unbewusst non-verbale Signale über das gewünschte Verhalten des Mitarbeiters sendet. Zum anderen ist nur schwer auszuschließen, dass sich ein Mystery Shopper in einer bewusst unnatürlichen Situation anders verhält als in einer natürlichen Situation. Denn jede Interaktion ist als eine Kette von Reizen und darauf folgenden Reaktionen zu verstehen, deren Abfolge von der subjektiven Interpretation der interagierenden Personen gelenkt wird. Da die Deutung von Reizen wesentlich vom Ziel der Interaktion abhängt und da das Ziel aus Sichtweise des Testkäufers im Hervorrufen und Beobachten bestimmter Prozesse besteht, ist es nicht auszuschließen, dass sich die Interaktion in eine Richtung bewegt, die nicht einer „natürlichen Situation“ entspricht.¹⁹²

Um das tatsächliche Verhalten eines Mitarbeiters in einer Kundenkontaktsituation gemäß Anforderung 1 ermitteln zu können, sind vier Voraussetzungen zu erfüllen.

Voraussetzung 1: Die Kundenkontaktsituation soll möglichst realitätsnah konstruiert werden.

Der Mitarbeiter darf den Mystery Shopper nicht als solchen erkennen. Daher ist es von Bedeutung, ein realistisches Test-Szenario zu entwickeln. Die Anweisungen für den Mystery Shopper sollten also wie ein Drehbuch gestaltet sein, so dass eine tatsächliche Kundenkontaktsituation widergespiegelt wird und der Mystery Shopper sich in die Rolle eines typischen Kunden versetzen kann.¹⁹³

Voraussetzung 2: Die Mystery Shopper sollen die Charakteristika typischer Kunden aufweisen.

Sofern Mystery Shopper nicht die Charakteristika der Zielgruppe aufweisen, fallen Sie auf. Ist zudem bekannt, dass Mystery Shopper eingesetzt werden, besteht die Gefahr der Enttarnung.

¹⁹¹ Vgl. Matzler/Kittinger-Rosanelli (2000: 226), Matzler/Pechlaner/Kohl (2000: 164)

¹⁹² Vgl. Matzler/Kittinger-Rosanelli (2000: 226), Matzler/Pechlaner/Kohl (2000: 165)

¹⁹³ Vgl. Matzler/Pechlaner/Kohl (2000: 165), Haas (2002: 283), Schmidt (2007: 103)

Somit ist darauf zu achten, entsprechende soziodemografische Merkmale und Ausprägungen zu berücksichtigen, sofern dies erforderlich ist.¹⁹⁴

Voraussetzung 3: Die Mystery Shopper sollen sich wie normale Kunden verhalten.

Aufbauend auf der in Voraussetzung 1 postulierten realitätsnahen Kundenkontaktsituation ist es erforderlich, dass der Mystery Shopper diese auch glaubhaft umsetzt. In Verbindung mit den in Voraussetzung 2 genannten Charakteristika ist es wichtig, dass so wenige Fakten wie möglich erfunden werden. So kann ein „normales Kundenverhalten“ sichergestellt werden und es werden keine Situationen herbeigeführt, in denen der Mystery Shopper sich durch Widersprüche entlarvt.¹⁹⁵ Je nach Komplexität der zu testenden Situation kann dies eine Herausforderung für Drehbuch und Mystery Shopper darstellen. Gerade vor diesem Hintergrund gewinnt nachfolgende Voraussetzung an besonderer Bedeutung.

Voraussetzung 4: Mystery Shopper sollen hinsichtlich des Ablaufs der Testkaufsituation und des Verhaltens eines realen Kunden geschult werden.

Um die vorangegangenen Voraussetzungen zur Bewertung des tatsächlichen Verhaltens von Mitarbeitern in einer Kundenkontaktsituation zu erfüllen, ist eine Schulung der Mystery Shopper erforderlich. Der Mystery Shopper ist darüber aufzuklären, was „normales“ Verhalten in der jeweiligen Situation ist, damit er sich nicht zu aktiv oder zu passiv verhält. Ferner ist ihm eine Reihe von möglichen zu stellenden Fragen an die Hand zu geben.¹⁹⁶ In diesem Kontext ist es schwierig, einen allgemeingültigen Grad und eine Form der Schulung zu benennen, der für alle Testkaufsituationen Gültigkeit hat. Ein angemessener Schulungsumfang für die Vergabe eines Konsumentencredits unterscheidet sich beispielsweise von dem einer Erhebung in einem Buchhandel. So kann es zum Teil sinnvoll sein, die Testkaufsituation in entsprechenden Rollenspielen zu proben. In anderen Fällen ist die Schulung in Form eines Drehbuchs mit Verhaltensanweisungen möglicherweise ausreichend.¹⁹⁷ So wird aus der Praxis von unterschiedlichsten Schulungsformen berichtet. Dabei reichen die eingesetzten Maßnahmen von einfachen Verhaltensanweisungen über Bildmaterial und Videos bis hin zu Rollenspielen, in denen die Testkaufsituation einstudiert wird.¹⁹⁸

Zur Erfüllung der Anforderung und Einhaltung der weiteren vorstehenden Voraussetzungen sind in der Praxis unterschiedliche Beispiele zu finden. SALCHER/ESSER beschreiben den Einsatz eines „Rollenprotokolls“, in dem die Anweisungen für den Testkunden detailliert

¹⁹⁴ Vgl. Finn/Kayandé (1999: 195), Matzler/Pechlaner/Kohl (2000: 165), Haas (2002: 283)

¹⁹⁵ Vgl. Matzler/Pechlaner/Kohl (2000: 166)

¹⁹⁶ Vgl. Hackmann/Schäbe (1996: 1138), Wilson (1998: 155), Haas (2002: 283f.), Schmidt (2007: 104)

¹⁹⁷ Wie bereits bei der Übersicht vergangener Forschungsbemühungen angemerkt, ist die Auswirkung des Schulungsgrades bisher unerforscht.

¹⁹⁸ Vgl. Sander (2000: 32f.)

beschrieben sind.¹⁹⁹ Damit wird Voraussetzung 1 Rechnung getragen. Neben der Identifikation typischer Testgesprächsthemen und der Durchführung der Schulung beschreibt ACHELIS die Entwicklung einer so genannten „Life Story“ für den Testkunden, die so ausführlich wie möglich formuliert ist, so dass der Testkunde in allen denkbaren Situationen des Testverlaufs bestehen kann und überzeugend wirkt.²⁰⁰ Während ACHELIS noch den Einsatz eigener Mitarbeiter als Mystery Shopper propagiert, wird inzwischen der Einsatz von unternehmensfremden Testern vorgeschlagen, etwa über Testkaufagenturen.²⁰¹ Teilweise verfügen solche Anbieter von Mystery Shopping-Dienstleistungen über eine Anzahl von mehr als 20.000 Testkunden. Es ist denkbar, dass diese Anbieter Realkunden von beauftragenden Unternehmen in ihrem Testerbestand haben, so dass deren echte Kunden eingesetzt werden könnten. Dies kann zum Beispiel sinnvoll sein, sofern für die Testkaufsituation eine Legitimation in Form einer Kundenkarte oder einer Kundenhistorie erforderlich ist. Ferner könnte die durch Datenschutzrichtlinien erschwerte Rekrutierung eigener Kunden so erleichtert werden.²⁰² Zudem weisen solche Mystery Shopper die in Voraussetzung 2 geforderten Charakteristika typischer Kunden auf, und es ist anzunehmen, dass sie sich gemäß Voraussetzung 3 wie normale Kunden verhalten. Gleichwohl gibt es auch Argumente gegen die Rekrutierung eigener Kunden. Demnach könnten vergangene Erfahrungen mit dem Unternehmen oder persönliche Kontakte zu Mitarbeitern die Urteilskraft trüben.²⁰³ Es ist anzunehmen, dass sich die Erfüllung der Voraussetzungen 2 und 3 mit zunehmender Komplexität des Untersuchungsgegenstandes schwieriger gestaltet. So wird ein aus dem Arbeitermilieu rekrutierter Tester nur schwerlich überzeugend eine Beurteilung der Vermögensverwaltung bei einer Privatbank bestehen. Es ist anzunehmen, dass auch eine intensive Schulung nur bedingten Erfolg hat, sofern typische Kundencharakteristika nicht vorhanden sind.

Anforderung 2: Mystery Shopping soll ein umfassendes Bild der Dienstleistungsqualität liefern.

Neben der in Anforderung 1 dargelegten Notwendigkeit der Ermittlung des tatsächlichen Mitarbeiterverhaltens ist es darüber hinaus erforderlich, ein umfassendes Urteil über die Qualität der Dienstleistung zu erhalten.²⁰⁴ In der Praxis werden vor allem standardisierte Instrumente eingesetzt, die bei den quantitativen, multiattributiven Methoden eingeordnet

¹⁹⁹ Vgl. Salcher/Esser (1987: 19)

²⁰⁰ Vgl. Achelis (1985: 26)

²⁰¹ Vgl. z.B. Guber/Kerper (2003: 64)

²⁰² Vgl. Bauer/Urbahn/Markart (2003: 22)

²⁰³ Vgl. Krüger/Stücken (2003: 33)

²⁰⁴ Vgl. Matzler/Pechlaner/Kohl (2000: 166), Schmidt (2007: 104f.)

werden.²⁰⁵ Allerdings werden diese Verfahren zunehmend kritisiert. Insbesondere werden vier Punkte angeführt:²⁰⁶ Erstens reflektieren die mit den Verfahren gewonnenen Daten aufgrund der beschränkten Anzahl der einsetzbaren Items nicht das gesamte Bild der wahrgenommenen Dienstleistungsqualität. Zweitens sind die Probanden genötigt, die komplexen Verfahren in einem Fragebogen auf einer Beurteilungsskala zu verdichten. Drittens müssen Items so allgemein und abstrakt formuliert werden, dass konkrete Qualitätswahrnehmungen nicht ausreichend wiedergegeben werden können. Viertens wird dem prozessualen Charakter einer Dienstleistung durch die „merkmalsorientierte“ Abfrage nicht Rechnung getragen.

Da Mystery Shopping zu Beurteilung der Dienstleistungsqualität Kundenkontaktsituationen mit prozessualen Charakter zum Gegenstand hat, ist folgende Voraussetzung zu beachten:

Voraussetzung 5: Der Dienstleistungsprozess soll in seine einzelnen Phasen zerlegt werden, für die Beobachtungskriterien festzulegen sind.

Damit der prozessuale Charakter ausreichend berücksichtigt wird, ist der Dienstleistungsprozess bei der Erstellung des Beobachtungsbogens in seine einzelnen Phasen zu zerlegen. Dabei sind für die einzelnen Phasen relevante Beobachtungskriterien zu definieren.²⁰⁷ MATZLER et al. schlagen dafür das von SHOSTACK entwickelte „Service Blueprinting“ vor.²⁰⁸ Dies ermöglicht die Gewinnung von Informationen, die durch standardisierte Multiattributverfahren mit ihrer begrenzten Anzahl abstrakter Items nicht erfasst werden können.

Möglicherweise ist der Einsatz des Service Blueprinting in der Praxis gar nicht erforderlich, da Praktiker in der Regel ihre Prozesse kennen. Es ist anzunehmen, dass in der Praxis weniger standardisierte Multiattributverfahren eingesetzt werden, sondern dass vielmehr eine Zerlegung des Dienstleistungsprozess in seine einzelnen Phasen aufgrund eines pragmatischen Vorgehens erfolgt. So beschreiben beispielsweise SALCHER/ESSER ein ebensolches Vorgehen bei der Konzeption eines „idealen Beratungsgesprächs“ und eines Beobachtungsbogens, der sich an den einzelnen Schritten des Ablaufs orientiert.²⁰⁹

Anforderung 3: Mystery Shopping soll die konkrete, subjektive Qualitätswahrnehmung eines realen Kunden erfassen.

²⁰⁵ z.B. das Konzept SERVQUAL, welches weit verbreitet zur Messung der Dienstleistungsqualität eingesetzt wird. Vgl. Parassuraman/Zeithaml/Berry (1988)

²⁰⁶ Vgl. hierzu und nachfolgend Hentschel (1992: 155ff.), Stauss/Hentschel (1992: 59ff.), Stauss (1997: 34f.), Matzler/Pechlaner/Kohl (2000: 166)

²⁰⁷ Vgl. Matzler/Pechlaner/Kohl (2000: 166)

²⁰⁸ Die Methode des „Service Blueprinting“ ist eine Technik zur Strukturierung und Visualisierung eines Dienstleistungsprozesses aus Kundenperspektive, siehe Shostack (1984: 133ff.)

²⁰⁹ Vgl. Salcher/Esser (1987: 19)

In den meisten Fällen bestimmt einzig das Urteil des Kunden seine Kaufentscheidung. Folglich hat die Beurteilung der Dienstleistungsqualität hauptsächlich aus Kundenperspektive zu erfolgen.²¹⁰ Allerdings sind den quantitativen Kundenbefragungsmethoden enge Grenzen gesetzt, wenn Kundenkontaktsituationen abgefragt werden, die sehr stark ins Detail gehen.²¹¹ Dies hängt zum einen mit der zeitlichen Verschiebung zwischen erlebter Situation und Abfragezeitpunkt zusammen. Zum anderen kann dies mit falschen Angaben aufgrund sozialer Erwünschtheit sowie mit kognitiven Dissonanzen erklärt werden. Die Folge ist eine Abweichung zwischen tatsächlicher und beschriebener bzw. erlebter Situation.²¹²

Damit die konkrete, subjektive Wahrnehmung eines realen Kunden erfasst werden kann, sind folgende zwei Voraussetzungen zu erfüllen:

Voraussetzung 6: Der Beobachtungsbogen soll konkrete Erlebnisinformation und nicht abstrakte Dienstleistungsdimensionen beinhalten.

Mit dem Einsatz von Mystery Shoppern können konkrete Erlebnisinformationen gewonnen werden, da ihr Augenmerk durch den Beobachtungsbogen gezielt auf spezifische Sachverhalte gelenkt werden kann. Dabei können Aspekte abgefragt werden, die einem normalen Kunden möglicherweise nicht auffallen würden wie z.B. die Einhaltung von Servicestandards.²¹³ Mit multiattributiven Verfahren lassen sich hingegen lediglich weniger konkrete Informationen erheben, aus denen sich wiederum nur weniger spezifische Handlungsempfehlungen ableiten lassen.²¹⁴ Damit die konkrete Erlebnisinformation gewonnen werden kann, ist eine genaue Festlegung der Kriterien im Beurteilungsbogen erforderlich. Ferner ist vor dem Hintergrund des Umfangs einer Untersuchung die Erfüllung folgender Voraussetzung von Bedeutung:

Voraussetzung 7: Die Erlebnisse sind unmittelbar nach der Testkaufsituation aufzuzeichnen, um einen Informationsverlust zu verhindern.

Damit einem Informationsverlust oder einer Informationsverzerrung entgegengewirkt wird, sind die Erlebnisse unmittelbar im Anschluss an die Testkaufsituation aufzuzeichnen.²¹⁵ Dies ist vor allem deshalb erforderlich, weil Wissen, insbesondere bei schwachen Gedächtnisspuren, durch neue Informationen, Einstellungen, Erwartungen und Vorurteile

²¹⁰ Vgl. Landgraf (1995: 18)

²¹¹ Vgl. Abschnitt 2.1.2.3

²¹² Vgl. Matzler/Kittinger-Rosanelli (2000: 225), Wilson (2001: 722), Finn (2001: 310)

²¹³ Vgl. Matzler/Pechlaner/Kohl (2000: 167)

²¹⁴ Vgl. Stauss/Hentschel (1990: 232f.), Hentschel (1992: 153f.)

²¹⁵ Vgl. Matzler/Pechlaner/Kohl (2000: 168), Schmidt (2007: 107)

beeinflussbar ist. Verstärkt wird der Effekt, wenn eine Vielzahl an Beobachtungsdetails gespeichert werden sollen.²¹⁶

Aus der Praxis wird vor allem von der Zielsetzung handlungsrelevante Informationen zu gewinnen, berichtet.²¹⁷ Dafür werden nicht abstrakte Dienstleistungsdimensionen, sondern konkrete Erlebnisinformationen abgefragt, die sich am Prozess der Dienstleistungserbringung orientieren.²¹⁸ Somit wird Voraussetzung 6 Rechnung getragen. Dies ist vor allem deshalb nicht verwunderlich, da anzunehmen ist, dass Mystery Shopping nicht zum Selbstzweck durchgeführt wird. Die Voraussetzung 7, der unmittelbar nach der Testkaufsituation zu erfolgenden Aufzeichnung der Ergebnisse, ist in den praktischen Beiträgen überwiegend als erfüllt anzusehen.²¹⁹ Es gibt jedoch auch Indizien einer zeitverzögerten Aufzeichnung, wenn Mystery Shopper mit Laptops ausgestattet werden.²²⁰ Dies ist auch für diejenigen Mystery Shopping-Anbieter zu vermuten, die die Testdaten über das Internet von ihren Testkunden übermittelt bekommen. Anzunehmen ist, dass die Mystery Shopper ihre Beobachtungen erst zuhause in den Rechner übertragen.

Anforderung 4: Mystery Shopping soll die Qualitätswahrnehmungen eines Kunden reliabel und valide messen.

Neben der Ermittlung des tatsächlichen Mitarbeiterverhaltens in einem umfassenden Kontext und der Erfassung der konkreten, subjektiven Qualitätswahrnehmung aus Kundenperspektive ist es für einen erfolgreichen Einsatz von Mystery Shoppern schließlich erforderlich, dass reliabel und valide gemessen wird.²²¹

Die Reliabilität eines Beobachtungsverfahrens und der damit erhobenen Daten ist dann gegeben, wenn die wiederholte Anwendung unter gleichen Erhebungs- und Messbedingungen zu gleichen Ergebnissen führt. Dabei kann die Reliabilität des Beobachtungsverfahrens zum einen dadurch beurteilt werden, inwieweit verschiedene Beobachter einen gleichen Sachverhalt gleich einordnen und zum anderen, inwieweit ein Beobachter einen Sachverhalt zu unterschiedlichen Zeitpunkten gleich bewertet. Letzteres ist allerdings schwer umzusetzen, da ein Testkunde sich bei einem gleichen Szenario bei dem gleichen Mitarbeiter leicht enttarnen würde.²²² In der Marktforschungspraxis werden zur Sicherung bzw. Erhöhung der Reliabilität des Urteils Beobachtungskriterien eingesetzt, die ein möglichst objektives Urteil

²¹⁶ Vgl. Morrison/Colman/Preston (1997: 349ff.)

²¹⁷ Vgl. Drees/Schiller (2000: 69), Drees/Schiller (2003: 160), Dreyer/Linne (2004: 369f.)

²¹⁸ Vgl. Salcher/Esser (1987: 19)

²¹⁹ Vgl. z.B. Geißler/Dingler (1969: 60), Höhner/Schaper (2004: 34ff.)

²²⁰ Vgl. Hackmann/Schäbe (1996: 1136)

²²¹ Vgl. vergleiche hierzu und im Folgenden Matzler/Pechlaner/Kohl (2000: 168ff.)

²²² Vgl. Matzler/Kittinger-Rosanelli (2000: 230)

erlauben.²²³ Dabei werden solche Kriterien eingesetzt, die einfach zu beobachten und zu quantifizieren sind (z.B. „Trug der Mitarbeiter ein Namensschild?“). Allerdings sind nicht alle Kriterien auf Nominalskalenniveau reduzierbar. Ferner lassen Daten auf Ordinalskalenniveau ein dezidierteres Urteil über das Gewicht einzelner Kriterien zu. Die Ausprägung subjektiver Bewertungskriterien ist zudem bei der Einordnung der Handlungsrelevanz der ermittelten Daten hilfreich.²²⁴ Der Einsatz von ausschließlich nominalskalierten Beobachtungsbögen wird der Methode nicht gerecht. Will man dennoch Nominalskalenniveau einsetzen, so kann die Reliabilität durch eine klare Operationalisierung der jeweiligen Konstrukte erhöht werden, indem mehrere beobachtbare Sachverhalte bewertet werden.²²⁵ Bei ordinalskalierten Items ist die Reliabilität durch detaillierte Ratingskalen zu erhöhen, wobei die Konstrukte so zu operationalisieren sind, dass Interpretations- oder Wahrnehmungsspielräume ausgeräumt werden und dass weder Erinnerungslücken entstehen können oder künstliche Beobachtungskriterien das Differenzierungsvermögen der Tester trüben.²²⁶

Daneben ist die Validität bei Mystery Shopping gegeben, wenn ein Mystery Shopper in der Lage ist, tatsächlich das zu messen, was er messen soll, nämlich die vom Kunden wahrgenommene Dienstleistungsqualität.²²⁷ Semantische Validität liegt vor, wenn die Operationalisierung des zu messenden Konstruktes gelingt, etwa in der begrifflichen Erfassung in Items. Empirische Validität liegt vor, wenn das Messverfahren das Konstrukt unverfälscht abbildet.²²⁸ Semantische Validität ist nicht als spezifisches Problem von Mystery Shopping anzusehen, da es auch bei anderen Verfahren zu lösen ist. Die Sicherung der empirischen Validität könnte hingegen ein Problem darstellen. Allerdings gibt es unterschiedliche Argumente, die für eine valide Erhebung der Dienstleistungsqualität mittels Mystery Shopping sprechen.²²⁹ Erstens erfolgt die Erhebung im realen Umfeld des Mitarbeiters, so dass sein Verhalten nicht durch Laborbedingungen verfälscht wird. Zweitens kann der Einfluss des Mystery Shoppers durch entsprechende Anweisungen und ein realitätsnahes Testszenario reduziert werden. Drittens wird durch die Tatsache der verdeckt teilnehmenden Beobachtung eine Verhaltensänderung des Mitarbeiters ausgeschlossen.

²²³ Vgl. Grove/Fisk (1992: 222), Sander (2000: 31), Haas (2002: 284)

²²⁴ Vgl. Schmidt (2007: 108)

²²⁵ Das Konstrukt „Freundlichkeit“ lässt sich beispielsweise anhand der vier Kriterien „Begrüßung, Bedanken, Lächeln, Augenkontakt“ in einem als reliabel bewerteten Maß operationalisieren. Vgl. Sutton/Rafaelli (1988: 468)

²²⁶ Vgl. Wilson (1998a: 154), Wilson/Gutmann (1998: 289), Wilson (2001: 727) und insbesondere Schmidt (2007), der sich detailliert mit Reliabilität und Validität bei Mystery Shopping auseinandersetzt.

²²⁷ Vgl. Matzler/Kittinger-Rosanelli (2000: 230)

²²⁸ Vgl. Kromrey (2006: 189ff.)

²²⁹ Vgl. hierzu und im Folgenden Matzler/Kittinger-Rosanelli (2000: 231), Matzler/Pechlaner/Kohl (2000:169f.), Haas (2002: 283f.) und Schmidt (2007: 109)

Letztens kann durch entsprechend hohe Fallzahlen die externe Validität sichergestellt werden.²³⁰

Reliabilität und Validität von Mystery Shopping werden daneben auch durch den Prozess der Informationsaufnahme, ihrer Speicherung und ihres Abrufs beeinflusst. MORRISON et al. haben dies aus kognitionspsychologischer Ebene untersucht und haben zahlreiche Einflüsse identifiziert. Die Überlegungen sind in Tabelle 2-2-1 wiedergegeben. Zusammenfassend wird festgestellt, dass die Informationsaufnahme unvollständig oder ungenau sein kann, dass gespeicherte Informationen vergessen werden können und dass Informationen bei der Übertragung in den Beobachtungsbogen nicht korrekt aus dem Gedächtnis abgerufen werden.²³¹

²³⁰ Gleichwohl resultiert eine höhere Fallzahl in höheren Kosten. Finn/Kayandé konnten jedoch nachweisen, dass die Datenerhebung mit Mystery Shopping im Vergleich zu Kundenbefragungen deutlich weniger Kosten verursacht, wenn vergleichbare Ergebnisse erzielt werden sollen. Dies wird auf die gezieltere Beobachtung, die höhere Motivation und die verlässlicheren Angaben der Mystery Shopper zurückgeführt. Vgl. hierzu Finn/Kayandé (1999: 199) und Finn (2001: 311)

²³¹ Vgl. Morrison/Colman/Preston (1997)

Tabelle 2-2-1: Einfluss kognitiver Prozesse auf Reliabilität und Validität

Phase	Einflussfaktoren	Beispiele
Informations- aufnahme	Physische Faktoren	<ul style="list-style-type: none"> - Lichtverhältnisse können die Wahrnehmungsfähigkeit von Sauberkeit einschränken. - Der Tageszeitpunkt und die damit verbundene Verfassung des Mystery Shoppers hat Einfluss auf seine Wahrnehmungsfähigkeit.
	Aufmerksamkeit	<ul style="list-style-type: none"> - Menschen nehmen Situationen und Sachverhalte unterschiedlich wahr. - Wissenslücken werden oft unbewusst mit Annahmen rekonstruiert.
	Einstellungen und soziale Prägung	<ul style="list-style-type: none"> - Persönliche Präferenzen für z.B. Sauberkeit können trotz Schulung die Bewertung beeinflussen. - Die Sympathie für den Mitarbeiter kann die Objektivität der Bewertung gefährden.
	Fachwissen	<ul style="list-style-type: none"> - Mystery Shopper mit unterschiedlichem Kenntnisstand der zu bewertenden Dienstleistung bewerten diese unterschiedlich.
	Zeitraum	<ul style="list-style-type: none"> - Je kürzer der Besuch, desto weniger Information kann aufgenommen werden
Informations- speicherung	Verzerrung	<ul style="list-style-type: none"> - Sind viele Details zu speichern, werden die zu speichernden Informationen häufig durch Vorurteile, Erwartungen und Einstellungen verzerrt.
Informations- abruf	Beobachtungsbogen	<ul style="list-style-type: none"> - Eine zu detaillierte Gestaltung des Beobachtungsbogens wirkt sich negativ auf die Abrufgenauigkeit aus.
	Kontextbezug	<ul style="list-style-type: none"> - Bei Konzentration auf den Gefühlszustand der Phase der Informationsaufnahme steigt die Genauigkeit.
	Stimmung	<ul style="list-style-type: none"> - Unterschiedliche Stimmungen bei Informationsaufnahme und Abruf beeinflussen die Abrufleistung.
	Geschlecht	<ul style="list-style-type: none"> - Es wird angenommen, dass geschlechtsspezifische Unterschiede hinsichtlich Zuverlässigkeit und Genauigkeit beim Informationsabruf bestehen.
	Alter	<ul style="list-style-type: none"> - Es wird angenommen, dass altersspezifische Unterschiede hinsichtlich Zuverlässigkeit und Genauigkeit beim Informationsabruf bestehen.

Quelle: Basierend auf den Ausführungen von Morrison/Colman/Preston (1997: 352f.)

Es ist festzustellen, dass zahlreiche Einflussfaktoren das Urteil von Mystery Shoppers im Prozess Informationsaufnahme, -speicherung und -abruf beeinflussen können. Die Faktoren können zum Teil gar nicht (z.B. etwa Lichtverhältnisse) oder nur teilweise kontrolliert werden. Letzteren Faktoren sollte bei der Konzeption besondere Aufmerksamkeit zu Teil werden, so dass Reliabilität und Validität durch folgende Maßnahmen gesichert bzw. erhöht werden.²³²

Voraussetzung 8: Der Beobachtungsbogen ist klar zu strukturieren und das Szenario klar zu beschreiben.

Das Urteil eines Beobachters kann durch Erinnerungs-, Wahrnehmungs- und Interpretationsfehler verfälscht werden.²³³ Erinnerungsfehler sind von der Zahl der zu beobachtenden Merkmale und der Häufigkeit ihres Auftretens abhängig. So können selten vorkommende Merkmale leicht übersehen werden, während häufig auftretende Merkmale bei einer zu umfangreichen Gesamtanzahl an Beobachtungsmerkmalen nicht zuverlässig registriert werden. Der Erinnerungsfehler kann durch eine Fragebogenstruktur reduziert werden, die dem Verlauf der Beobachtungssituation entspricht. Wahrnehmungs- und Interpretationsfehler resultieren zumeist aus dem beim Beobachter gelassenen Beurteilungsspielraum. Daher ist dieser Spielraum durch entsprechende klare Szenariobeschreibungen zu eliminieren bzw. zu reduzieren.²³⁴ Ferner ist der Spielraum durch entsprechende Operationalisierungen einzuschränken, wie bereits oben ausgeführt.²³⁵

Voraussetzung 9: Individuelle und situationsspezifische Faktoren der Mystery Shopper sind bei der Untersuchungsplanung zu berücksichtigen.

Wie insbesondere in Tabelle 2-2-1 aufgeführt, existieren unterschiedliche individuelle und situationsspezifische Faktoren, die sich auf die Wahrnehmungsfähigkeit und somit auf das Urteil von Mystery Shoppers auswirken. So kann beispielsweise Müdigkeit, resultierend aus dem Tageszeitpunkt der Testdurchführung oder aufgrund eines lang andauernden Testszenarios, in Erinnerungs- und Wahrnehmungsfehlern resultieren.²³⁶ Folglich sind bei der Konzeption individuelle und situationsspezifische Faktoren zu berücksichtigen.

Voraussetzung 10: Beobachtungsspielräume sind durch ausreichende Operationalisierung der Konstrukte einzuschränken.

²³² Vgl. hierzu und im Folgenden Matzler/Kittinger-Rosanelli (2000: 232f.), Matzler/Pechlaner/Kohl (2000:171) und Schmidt (2007: 111ff.)

²³³ Vgl. Greve/Wentura (1991: 52f.)

²³⁴ Vgl. Schmidt (2007:111f.)

²³⁵ Vgl. die Ausführungen unter Anforderung 4

²³⁶ Vgl. die in Tabelle 2-2-1 zusammengefassten Ergebnisse von Morrison/Preston/Colman (1997) und vgl. Schmidt (2007: 112)

Die Reduzierung bzw. Eliminierung von Beobachtungsspielräumen und die damit verknüpfte Notwendigkeit einer Operationalisierung wurde bereits unter den Ausführungen unter Anforderung 4 behandelt. Daher wird an dieser Stelle auf eine Wiederholung verzichtet.

Voraussetzung 11: Es ist auf eine angemessene Anzahl an Bewertungskategorien zu achten, um die Erinnerung der Mystery Shopper nicht zu überfordern.

Wie bereits in den Ausführungen zu Voraussetzung 8 diskutiert, hängt die Zuverlässigkeit der Beobachtung der Mystery Shopper von der Anzahl der zu beobachtenden Merkmale und der Häufigkeit ihres Auftretens zusammen. Sind sehr viele Details einer Kundenkontaktsituation zu speichern, steigt das Risiko, dass bei der Aufzeichnung nicht alles erinnert werden kann. Das Urteil kann dann durch Annahmen, persönliche Einstellungen und Erwartungen beeinflusst werden.²³⁷ Folglich ist auf einen angemessenen Umfang des Beobachtungsbogens zu achten.

Voraussetzung 12: Es sind ausreichend hohe Fallzahlen zu erbringen.

Damit eine zuverlässige Informationsbasis geschaffen wird, die Fehlentscheidungen aufgrund nicht repräsentativer Einzelurteile vermeidet, sind angemessen hohe Fallzahlen anzustreben.²³⁸ FINN/KAYANDÉ zeigen, dass ca. 40 Testfälle pro Filiale notwendig sind, um vergleichbare reliable Ergebnisse wie bei einer Kundenbefragung zu erhalten, wenn die Servicequalität erhoben wird. Sofern rein tangible, objektiv beurteilbare Sachverhalte zu bewerten sind, kann diese Zahl auf elf Testfälle reduziert werden.²³⁹ Demnach ist die Höhe der anzusetzenden Testfälle vom Grad der Objektivierung bzw. der Subjektivität der herangezogenen Bewertungskriterien abhängig.

Zur Erfüllung der Anforderung 4 und Einhaltung der entsprechenden Voraussetzungen sind in der Praxis unterschiedliche Beispiele zu finden. Damit die in Voraussetzung 8 geforderte klare Szenariobeschreibung sichergestellt ist, arbeiten Marktforschungsexperten zum Teil mit den jeweiligen Prozessexperten aus dem beauftragenden Unternehmen zusammen und entwerfen gemeinsam Szenario, Verhaltensanweisungen und Beobachtungsbogen.²⁴⁰ Dies erscheint sinnvoll, um einerseits handlungsrelevante Informationen zu gewinnen und andererseits die Validität zu sichern bzw. zu erhöhen. Hinsichtlich der Strukturierung könnten internetbasierte Anwendungen zur Dateneingabe vorteilhaft sein, da die beobachteten Aspekte iterativ abgefragt werden können, ohne dass der Mystery Shopper durch die Gesamtansicht des Beobachtungsbogens überfordert wird. Bezüglich der Einhaltung der Voraussetzung 9 sind lediglich wenige Hinweise für die Berücksichtigung situationspezifischer Hinweise zu

²³⁷ Vgl. Morrison/Colman/Preston (1997: 352), Matzler/Pechlaner/Kohl (2000:171)

²³⁸ Vgl. Matzler/Kittinger-Rosanelli (2000: 232)

²³⁹ Vgl. Finn/Kayandé (1999: 212)

²⁴⁰ Vgl. Leeds (1995: 21), Höhner/Schaper (2004: 33)

finden, die sich dann insbesondere dem Testumfang widmen.²⁴¹ Daneben werden auch sehr umfangreiche Beobachtungskataloge beschrieben, bei denen die Einhaltung der Voraussetzung fraglich ist.²⁴² Zur Berücksichtigung individueller Faktoren bei der Untersuchungsplanung sind bei den praktischen Beiträgen keine konkreten Hinweise zu finden. Die nur oberflächliche Einhaltung der Voraussetzung 9 in der Praxis ist verwunderlich, denn Daten, die nicht über eine entsprechende Güte verfügen, können möglicherweise zu fehlgeleiteten Maßnahmen führen. Es ist anzunehmen, dass die Gründe für die ausbleibende Berücksichtigung individueller Faktoren dem Tester überlassen bleibt. Bezüglich der situationsspezifischen Faktoren ist zu mutmaßen, dass gern eine Vielzahl an Aspekten abgefragt wird, statt sich auf die wirklich handlungsrelevanten Informationen zu beschränken. Agenturen fördern diese Aufstockung des Beobachtungsbogens, da anzunehmen ist, dass ihre Vergütung sich auch am Testumfang orientiert. Voraussetzung 10, die Einschränkung des Interpretationsspielraums durch ausreichende Operationalisierung der Konstrukte, wird in der Praxis Rechnung getragen und wird in unterschiedlichen Beiträgen thematisiert.²⁴³ In Bezug auf die in Voraussetzung 11 geforderte angemessene Anzahl der Bewertungskategorien sind die Feststellungen und Überlegungen zu Voraussetzung 9 anzuführen. Demnach ist zu vermuten, dass Beobachtungsbögen in der Praxis zum Teil mit Items überfrachtet werden, so dass eine Überforderung der Mystery Shopper nicht auszuschließen ist. Abschließend ist die praktische Umsetzung der in Voraussetzung 12 geforderten ausreichend hohen Zahl an Testfällen zu betrachten. Hier zeichnet sich bei den praktischen Beiträgen ein undurchsichtiges Bild. Zum Teil wird trotz Vorgehensbeschreibung in der Datenerhebungsphase der Hinweis auf die Anzahl der Testfälle ausgespart.²⁴⁴ Unterschiedliche Autoren führen an, dass die Fallzahl bei Mystery Shopping im Vergleich zu Kundenbefragungen deutlich niedriger angesetzt werden kann, da nicht die statistische Signifikanz, sondern die grundsätzliche Aufdeckung vorhandener Missstände bedeutsam sei.²⁴⁵ BAUER/URBAHN/MARKART argumentieren im Zusammenhang der niedrigeren Fallzahl, dass Tester auf ein einheitliches Bewertungsmaß geschult sind, was in einer niedrigen Varianz resultiert, während die Bewertungsunterschiede zwischen realen Kunden die Fehlervarianz erhöhen. Anhaltspunkte auf eine sinnvolle Größe der Datenbasis werden aber nicht gegeben.²⁴⁶ DECKERS greift die von FINN/KAYANDÉ genannte Anzahl von mindestens 40 Testfällen pro Filiale auf und weist darauf hin, dass eine solche Zahl, insbesondere bei

²⁴¹ Vgl. Hessler (1999: 62f.)

²⁴² So beschreiben Drees/Schiller einen Beobachtungskatalog mit 100 Oberfragen, denen weitere Unterfragen zugeordnet werden. Vgl. Drees/Schiller (2003: 168)

²⁴³ Vgl. Sander (2000: 31), Drees/Schiller (2000: 70), Deckers (2003: 35), Lammers/Schubert (2003: 58),

²⁴⁴ z.B. bei Drees/Schiller (2003: 169)

²⁴⁵ Vgl. Bössow/Zinnegger (2003: 67), Krüger/Stücken (2003: 35)

²⁴⁶ Vgl. Bauer/Urbahn/Markart (2003: 18)

Unternehmen mit ausgedehntem Filialnetz, nicht praktikabel sei.²⁴⁷ Entscheidend bei der Festsetzung der Fallzahl ist offenbar die Zielsetzung. GUBER/KERPER beschreiben ein Mystery Shopping in der Versicherungswirtschaft. Dabei wurden Versicherungsvermittler nach Produktionsleistung in vier Gruppen aufgeteilt. Mit Testberatungen sollte ermittelt werden, ob ein Zusammenhang zwischen Produktionsleistung und Beratungsqualität der Vermittler besteht.²⁴⁸ Unabhängig vom Ergebnis dieser Untersuchung wird deutlich, dass eine Aggregation der Ergebnisse je nach Zielsetzung sinnvoll sein kann. So können die Zielsetzungen darin liegen, bestimmte Gruppen (z.B. Regionen oder Filialtypen) miteinander zu vergleichen, Abweichungsgrade innerhalb der Gesamtheit aufzudecken, Stärken des Gesamtunternehmens zu identifizieren und die Veränderung der Daten über bestimmte Zeiträume hinweg zu analysieren. Es ist anzunehmen, dass die Erhebung von aggregierten Daten ebenfalls zu handlungsrelevanten Informationen führt, wenn auch auf einer übergeordneten Ebene, gleichwohl aber aufgrund der reduzierten Testanzahl zu deutlich niedrigeren Kosten als auf Filialebene.

Es ist zu konstatieren, dass für den leistungsfähigen Einsatz von Mystery Shopping und die damit verknüpfte Generierung handlungsrelevanter Informationen unterschiedliche Anforderungen erfüllt sein müssen, die die Einhaltung einer Reihe von Voraussetzungen erfordern. In der Praxis wird diesem Anforderungsgerüst weitgehend Rechnung getragen. Allerdings sind Schwächen hinsichtlich des inhaltlichen Umfangs der Beobachtungsbögen anzunehmen und Unsicherheiten bezüglich der erforderlichen Anzahl der Testfälle nicht ganz ausgeräumt. In nachfolgender Tabelle 2-2-2 sind die Anforderungen und Voraussetzungen für den erfolgreichen Einsatz von Mystery Shoppern zusammengefasst.

²⁴⁷ Demnach kämen Automobilhersteller wie Audi oder Volkswagen auf die exorbitant hohe Zahl von 50.000 notwendigen Testfällen für ein Mystery Shopping ihres Vertriebsnetzes. Vgl. Deckers (2003: 38)

²⁴⁸ Vgl. Guber/Kerper (2003: 62f.); Anmerkung: Es wurde lediglich über das Vorgehen, nicht aber über das Resultat berichtet.

Tabelle 2-2-2: Anforderungen an und Voraussetzungen für den erfolgreichen Einsatz von Mystery Shoppern

Anforderung	Voraussetzung
1. Mystery Shopping soll das tatsächliche Verhalten eines Mitarbeiters in einer Kundenkontaktsituation ermitteln.	1. Die Kundenkontaktsituation soll möglichst realitätsnah konstruiert werden. 2. Die Mystery Shopper sollen die Charakteristika typischer Kunden aufweisen. 3. Mystery Shopper sollen sich wie normale Kunden verhalten. 4. Mystery Shopper sollen hinsichtlich des Ablaufs der Testkaufsituation und des Verhaltens eines realen Kunden geschult werden.
2. Mystery Shopping soll ein umfassendes Bild der Dienstleistungsqualität liefern.	5. Der Dienstleistungsprozess soll in seine einzelnen Phasen zerlegt werden, für die Beobachtungskriterien festzulegen sind.
3. Mystery Shopping soll die konkrete, subjektive Qualitätswahrnehmung eines realen Kunden erfassen.	6. Der Beobachtungsbogen soll konkrete Erlebnisinformation und nicht abstrakte Dienstleistungsdimensionen beinhalten. 7. Die Erlebnisse sind unmittelbar nach der Testkaufsituation aufzuzeichnen, um einen Informationsverlust zu verhindern.
4. Mystery Shopping soll die Qualitätswahrnehmungen eines Kunden reliabel und valide messen.	8. Der Beobachtungsbogen ist klar zu strukturieren und das Szenario klar zu beschreiben. 9. Individuelle und situationsspezifische Faktoren der Mystery Shopper sind bei der Untersuchungsplanung zu berücksichtigen. 10. Beobachtungsspielräume sind durch ausreichende Operationalisierung der Konstrukte einzuschränken. 11. Es ist auf eine angemessene Anzahl an Bewertungskategorien zu achten, um die Erinnerung der Mystery Shopper nicht zu überfordern. 12. Es sind ausreichend hohe Fallzahlen zu erbringen.

Quelle: In Anlehnung an Matzler/Kittinger-Rosanelli (2000: 226ff.), Matzler/Pechlaner/Kohl (2000: 172) und Schmidt (2007: 114)

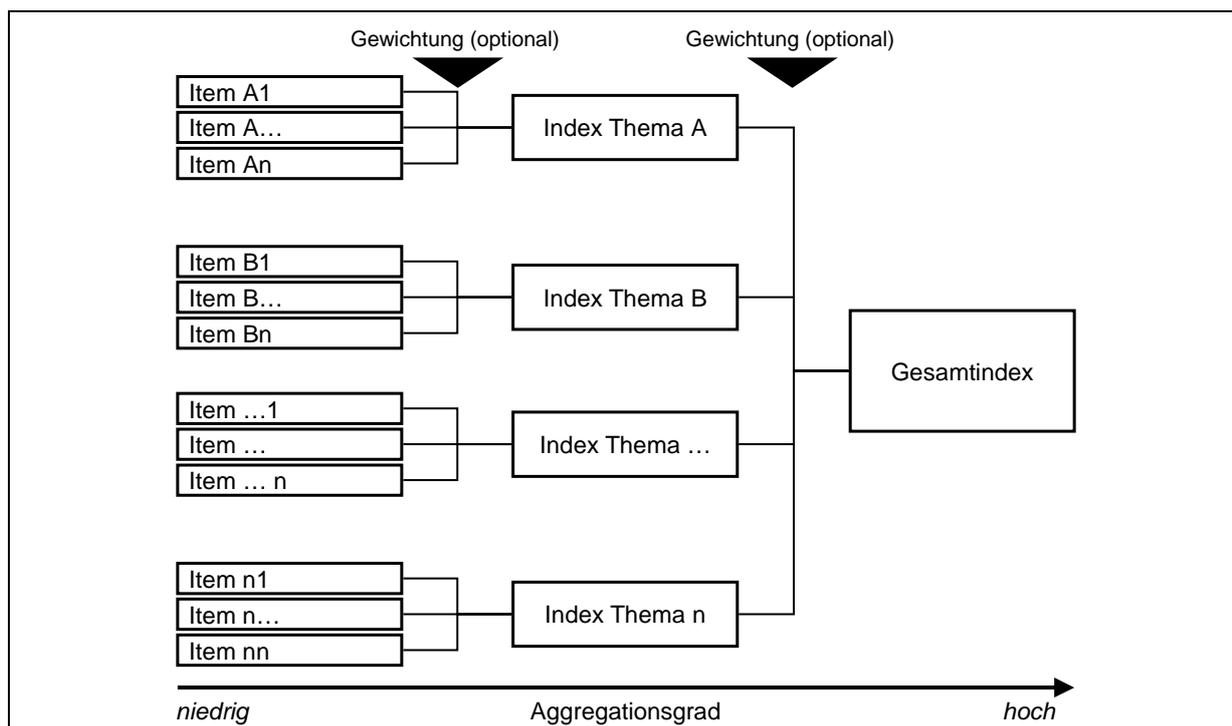
2.2.3.3 Auswertungsphase und Verwendung der Daten einer Mystery Shopping-Erhebung

Ausgehend von der übergeordneten Zielsetzung werden bei einer Mystery Shopping-Untersuchung handlungsrelevante Informationen gewonnen. Damit diese Informationen in konkrete Handlungen umgesetzt werden können, bieten sich unterschiedliche Werkzeuge zur Analyse und Darstellung an, deren Einsatz in diesem Abschnitt kurz erläutert werden soll.

Die Ergebnisse einer Untersuchung können unterschiedliche Adressaten haben. Während Mitarbeiter auf Filialebene besonderes Interesse haben, die in ihrer Filiale erhobenen Daten in Form von Einzelfallreports einzusehen, ist es auf Managementebene interessant, über aggregierte Werte informiert zu werden.²⁴⁹ Die Aggregation von gewonnenen Informationen kann beispielsweise durch Indizes erfolgen. Für jedes Item, jedes Untersuchungsthema bis hin zu einem Gesamtindex können Werte ermittelt werden, die optional mit einer Gewichtung für besonders herauszustellende Items oder Themen versehen werden können.²⁵⁰ Themenspezifische Indizes liefern im Rahmen einer Stärken-Schwächen-Analyse rasch Informationen über handlungsrelevante Bereiche. Dabei kann die zusätzliche Verwendung von Schwellenwerten, die nicht unterschritten werden dürfen, die Handlungsrelevanz besonders verdeutlichen. Ferner sind die ermittelten Werte von Filialen oder Regionen durch die Aggregation leichter miteinander und an einem gemeinsamen Mittelwert vergleichbar. Es kann beispielsweise ein Ranking erstellt werden, welches sich an einem oder verschiedenen aggregierten Werten orientiert. Abbildung 2-2-5 soll die Aggregation von gewonnenen Informationen beispielhaft verdeutlichen.

²⁴⁹ Vgl. Pergandé/Stücken (2004: 219)

²⁵⁰ Wie aber bereits in Abschnitt 2.1.2.3 gezeigt wurde, ist eine Gewichtung auch mit Hilfe eines Verteilungsbudgets bereits bei lediglich wenigen Items nur schwierig realisierbar und daher nicht praktikabel.

Abbildung 2-2-5: Aggregationsgrad gewonnener Informationen

Quelle: Eigene Darstellung

Zur Analyse von Stärken und Schwächen scheinen aggregierte Werte besonders geeignet. Der Blick wird dabei auf besonders extreme Werte gerichtet. Mit dem Einsatz von Schulungs- und Weiterbildungsmaßnahmen kann eine Verbesserung der Werte angestrebt und deren Erfolg letztlich auch überprüft werden.²⁵¹ Stehen Vergleichswerte anderer Unternehmen zur Verfügung, lässt sich überprüfen, wo der Wettbewerb besser oder schlechter ist.²⁵² Dies gewinnt dann an Bedeutung, wenn ein Unternehmen beispielsweise überprüfen möchte, ob das gesteckte Ziel der Qualitätsführerschaft im Service erreicht wurde. Die ermittelten Werte können aber nicht nur zu Vergleichszwecken ermittelt werden. Auch eine Kopplung mit leistungsabhängigen Gehaltsbestandteilen ist möglich, so dass die Erhebung eine Konsequenz für die Mitarbeiter hat und einen Leistungs- und Motivationsanreiz stiftet.²⁵³ Die unternehmensweite Veröffentlichung der ermittelten Mystery Shopping-Werte kann sich positiv auf den Teamgeist und die Zusammenarbeit innerhalb von Filialen auswirken. Dies kann durch materielle Anreize für Filialen mit besonders guten Ergebnissen gefördert werden.²⁵⁴

Die mit Mystery Shopping ermittelten objektivierten Daten können auch mit den subjektiven Einschätzungen einer Kundenbefragung verglichen werden. Dabei ist eine Darstellung in

²⁵¹ Vgl. Bössow/Zinnegger (2003: 67)

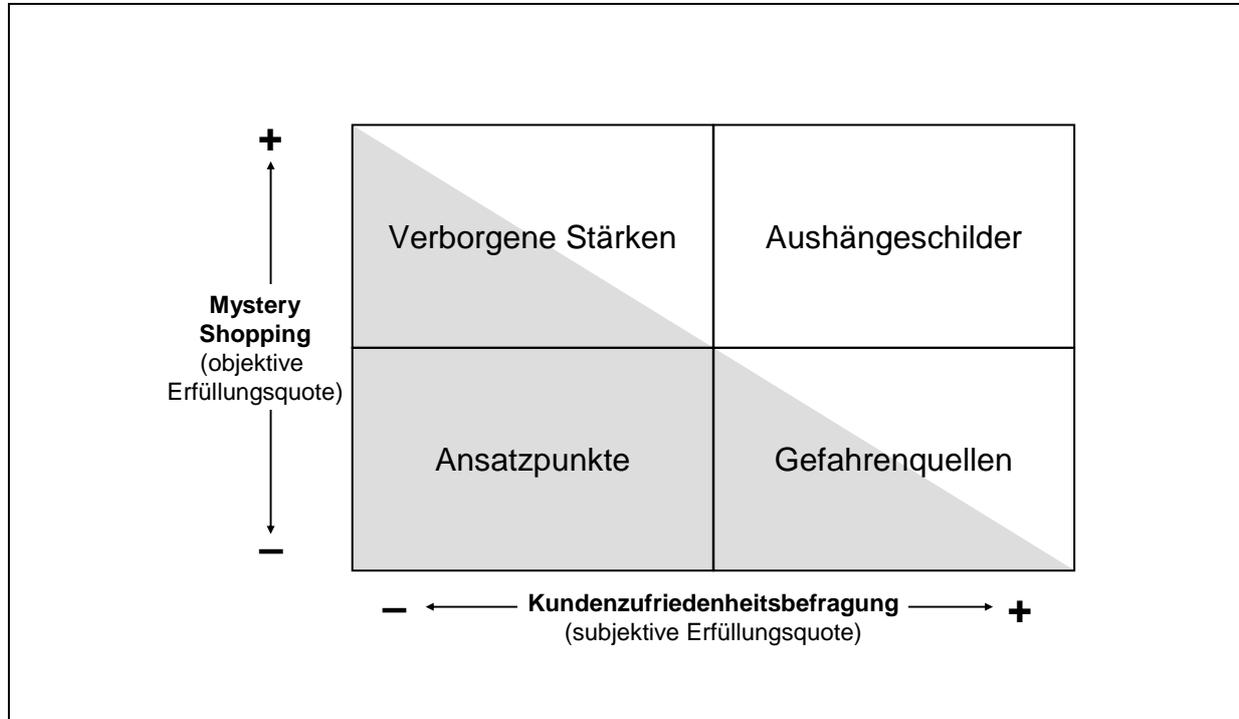
²⁵² Vgl. Pergandé/Stücken (2004: 220)

²⁵³ Vgl. Kohlbacher/Egloff (2003: 74)

²⁵⁴ Vgl. Hesselink/van der Wiele (2003: 6)

Form einer Handlungsrelevanzmatrix sinnvoll, die die Wichtigkeit der Handlungsbereiche abbildet.²⁵⁵ Dabei wird die Handlungsrelevanz durch die subjektive und objektive Erfüllungsquote verdeutlicht, wie in Abbildung 2-2-6 veranschaulicht ist.

Abbildung 2-2-6: Handlungsrelevanzmatrix auf Basis subjektiver und objektiver Erfüllung



Quelle: in Anlehnung an Stücken (2003: 46) und Semel (2006: 96)

Wird die Erfüllungsquote von bestimmten Themen objektiv und subjektiv als hoch wahrgenommen, so sollte die Leistung in diesen Bereichen konsequent fortgeführt und positiv kommuniziert werden. Sind Themen in der Kundenzufriedenheitsbefragung als positiv wahrgenommen worden, die sich in den Ergebnissen einer Mystery Shopping-Untersuchung aber nicht widerspiegeln, so sind sie zu untersuchen, da sie zukünftig zu subjektiven Negativwahrnehmungen führen können. Aspekte, die subjektiv als schlecht und objektiv als gut erfüllt angesehen werden, bedürfen einer verstärkten Kommunikation, da sie möglicherweise in der Vergangenheit besonders schlecht erfüllt waren und die positive Veränderung sich in der Wahrnehmung der Kunden noch nicht niedergeschlagen hat. Denkbar ist auch, dass diese Standards zwar objektiv gut erfüllt sind, aber nicht den Kundenwünschen entsprechen. Kriterien, die durch Mystery Shopping und Kundenzufriedenheitsbefragung als mangelhaft erfüllt angesehen werden, sind dringend handlungsrelevante Themen. Neben der Ableitung der Handlungsrelevanz durch objektive und subjektive Erfüllung ist auch eine Ableitung aus den Werten der subjektiven Wichtigkeit und objektiver Erfüllung möglich. Dies liefert Aufschluss, ob entwickelte Standards mit den Kundenerwartungen übereinstimmen.

²⁵⁵ Vgl. hierzu und im Folgenden Stücken (2003: 46)

Es wird deutlich, dass die Auswertungs- und Verwendungsmöglichkeiten von Mystery Shopping selbst beim Einsatz einfacher arithmetischer Verfahren wie der Mittelwertberechnung bereits vielfältig und nützlich sind. Gleichwohl können durch den Einsatz multivariater statistischer Analyseverfahren deutlich mehr Informationen gewonnen werden.²⁵⁶ So kann beispielsweise mit ihnen die Beobachterübereinstimmung bewertet werden.²⁵⁷ Ferner lassen sie vor allem die Trennung von wichtigen und unwichtigen Ergebnissen zu, denn eine Priorisierung der Handlungsrelevanz kann nur erfolgen, wenn gleichzeitig mehrere Kriterien in einen Zusammenhang mit den relevanten Erfolgsfaktoren gebracht werden.²⁵⁸ Bei den praktischen Beiträgen sind allerdings keine Hinweise auf den Einsatz derartiger Methoden zu finden. Es ist daher anzunehmen, dass in der Praxis aus Kosten- oder Komplexitätsgründen darauf verzichtet wird.

Nachdem im ersten Abschnitt die theoretischen Grundlagen von Mystery Shopping erläutert wurden, widmete sich dieser zweite Abschnitt der Umsetzung in der Praxis. Dabei wurde neben einer Erläuterung der Zielsetzungen und Einsatzmöglichkeiten ein Marktüberblick gegeben und insbesondere auf die Konzeption einer Mystery Shopping-Untersuchung eingegangen, die den Boden für die empirische Untersuchung bereitet, die im anschließenden Kapitel dargestellt wird. Davor widmet sich der folgende dritte Abschnitt dieses Kapitels kritischen Aspekten.

²⁵⁶ Für einen anwendungsorientierten Überblick multivariater Analysemethoden siehe Backhaus/Erichson/Plinke/Weiber (2006)

²⁵⁷ Vgl. Schmidt (2007: 196)

²⁵⁸ Vgl. Rust/Zahorik/Keinigham (1995: 60f.), Anderson/Mittal (2000: 108ff.)

2.3 Kritische Würdigung von Mystery Shopping

Im Rahmen einer kritischen Würdigung sollen vier unterschiedliche Aspekte aufgegriffen werden, die teilweise bereits im Zuge der Darlegung der theoretischen Grundlagen genannt wurden. Erstens ist dies der ökonomische Nutzen von Mystery Shopping, der bisher weder in praktischen noch in theoretischen Beiträgen Beachtung gefunden hat. Zweitens werden Überlegungen zur Frage nach eigener oder fremder Leistungserstellung einer Mystery Shopping-Untersuchung angestellt. Drittens werden ethische Aspekte, die mit der Durchführung von Mystery Shopping einhergehen, betrachtet. Schließlich wird der Fragestellung nach Qualifikation und möglichen vorteilhaften (soziodemografischen) Eigenschaften von Testern nachgegangen. Dabei werden Forschungshypothesen formuliert, die im anschließenden Kapitel untersucht werden sollen.

2.3.1 Nutzen von Mystery Shopping

Der Nutzen von Mystery Shopping wird in der Literatur lediglich hinsichtlich qualitativer Aspekte der Eignung der Methode dargestellt. So eignet sich Mystery Shopping beispielsweise als Kontrollinstrument²⁵⁹, als Methode zur Analyse von Verkaufssituationen²⁶⁰, um theoretische und wahrnehmungsrelevante Aspekte der Dienstleistungsqualität zu erfassen und zu analysieren²⁶¹ sowie zur Messung der wahrgenommenen Dienstleistungsqualität.²⁶²

Hinsichtlich quantitativer Aussagen über den Nutzen, findet sich nur die Feststellung, dass Mystery Shopping im Vergleich zur Kundenbefragung bezogen auf die Kosten effizienter einsetzbar ist.²⁶³ Eine Quantifizierung des Kosten-Nutzen-Zusammenhangs von Mystery Shopping erfolgte bisher nicht. Wie bereits dargelegt, eignet sich Mystery Shopping als Instrument zur Messung der Dienstleistungsqualität. Erfolgen auf Basis der mittels Mystery Shopping gewonnenen Erkenntnisse entsprechende Maßnahmen, die der Verbesserung der erbrachten und wahrgenommenen Qualität dienen, so ist eine Steigerung der Kundenzufriedenheit anzunehmen. Die Kundenzufriedenheit wirkt sich über die medierenden Konstrukte Kundenloyalität und Markterfolg positiv auf den wirtschaftlichen Erfolg eines Unternehmens aus.²⁶⁴ Allerdings herrscht Unklarheit über die Form der Zusammenhänge. So ist es bisher nicht gelungen, ein Modell zu entwickeln, welches den Bogen von der Ermittlung

²⁵⁹ Vgl. Deges (1992)

²⁶⁰ Vgl. Haas (2002)

²⁶¹ Vgl. Semel (2006)

²⁶² Vgl. Matzler/Kittinger-Rosanelli (2000)

²⁶³ Vgl. Finn/Kayandé (1999)

²⁶⁴ Vgl. Homburg/Bucerius (2006) für einen Überblick vergangener Forschungsbemühungen

der Kundenzufriedenheit bis hin zu konkreten finanzwirtschaftlichen Größen spannt.²⁶⁵ Dies lässt sich auch für Mystery Shopping behaupten. Aus ökonomischer Sicht sowie vor dem Hintergrund von Budgetrestriktionen und Einsparungsdruck ist es daher verwunderlich, dass eine genaue Untersuchung bisher nicht erfolgte²⁶⁶ (oder möglicherweise aus Komplexitätsgründen nicht gelungen ist). Zusammenfassend ist festzustellen, dass eine Aussage über den quantitativen Zusammenhang im Sinne einer detaillierten Kosten-Nutzen-Betrachtung bisher nicht getroffen werden konnte.

Gleichwohl gibt es Ansätze, die Maßnahmen zur Steigerung von Dienstleistungsqualität oder Kundenzufriedenheit hinsichtlich ihrer Rentabilität untersuchen. Der Return on Quality-Ansatz liefert beispielsweise ein Modell, welches eine monetäre Bewertung von Qualitätsverbesserungsmaßnahmen ermöglicht.²⁶⁷ Ein konzeptionell ähnliches Modell ist der Return on Customer Satisfaction-Ansatz, der basierend auf dem Kundenwert einen Zusammenhang zwischen Kundenzufriedenheit, Kundenbindung und Rentabilität herstellt.²⁶⁸ Die Modelle gehen jedoch nicht auf die Entstehung und die Beeinflussbarkeit der Kundenzufriedenheit durch das Unternehmen ein. Denkbar ist, die Ansätze mit Mystery Shopping zu verknüpfen und einen Ansatz zu entwickeln, der sich nicht nur mit Kosten-Nutzen-Aspekten der aus einer Mystery Shopping-Untersuchung abgeleiteten Maßnahmen auseinandersetzt, sondern der vorher ansetzt. Dabei wäre auch zu untersuchen, ob Unternehmen, die Mystery Shopping einsetzen, einen höheren Grad an Kundenzufriedenheit im Vergleich zu Unternehmen vorweisen können, die kein Mystery Shopping einsetzen.²⁶⁹ Dies ließe sich auch auf den Bereich Mitarbeiterzufriedenheit ausdehnen. Dabei wäre zu untersuchen, ob der Einsatz von Mystery Shopping beispielsweise Auswirkungen auf die Produktivität der Arbeitskräfte hat. Die Kenntnis der Auswirkungen wäre für Unternehmen insbesondere sinnvoll, um neben „weichen“ qualitativen Aussagen, „harte“ quantitative Aussagen über den Nutzen des Einsatzes von Mystery Shopping treffen zu können. Konkret bedeutet ein solcher *Return on Mystery Shopping* eine Hilfestellung für Unternehmen bei der Entscheidung hinsichtlich Durchführung und Umfang einer Mystery Shopping-Untersuchung. Hier besteht eine konkrete Forschungslücke, die nicht nur Mystery Shopping, sondern auch andere Verfahren der Erhebung der Dienstleistungsqualität und Kundenzufriedenheit betrifft. Ein möglicher konzeptioneller Rahmen für einen Return of Mystery Shopping wird in Abbildung 2-3-1 vorgeschlagen.

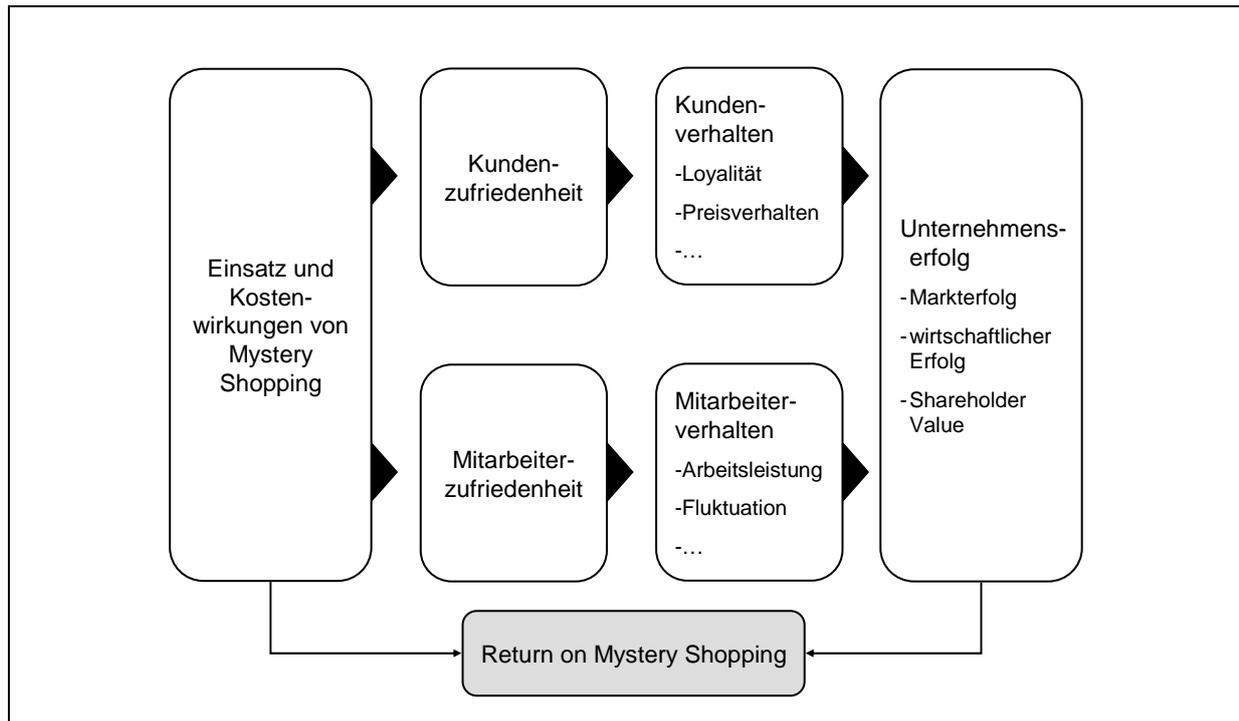
²⁶⁵ Vgl. Buhl/Mellwig/Renz (2003: 3)

²⁶⁶ Vgl. hierzu insbesondere Krafft (1999), der u.a. den fehlenden Bezug zu ökonomischen Final-Zielen wie Unternehmensrentabilität bei der Messung von Kundenzufriedenheit kritisiert.

²⁶⁷ Vgl. Rust/Zahorik/Keiningham (1995)

²⁶⁸ Vgl. Fischer/Herrmann/Huber (2001)

²⁶⁹ Vgl. Cook et al. (2002: 172)

Abbildung 2-3-1: Möglicher konzeptioneller Rahmen eines Return on Mystery Shopping

Quelle: Eigene Darstellung

Allerdings ist eine Quantifizierung für bestimmte Formen des Mystery Shoppings denkbar, sofern das Ziel nicht die Erhebung der eigenen Dienstleistungsqualität ist. Als Beispiel soll ein Industrieunternehmen angeführt werden, das die Umsetzungsqualität von konsumentengerichteten Promotion-Aktivitäten überprüfen möchte, mit denen eine entsprechende Agentur beauftragt wurde. Mit Hilfe von Mystery Shoppern kann dezentral effizient überprüft werden, ob die vereinbarte Umsetzungsqualität eingehalten wird, so dass bei Nichterfüllung eine Preisreduzierung erfolgen kann. In einem zweiten Beispiel soll ein Industrieunternehmen angeführt werden, das mit dem Handel gegen eine entsprechende Gebühr die bevorzugte Beratung hin zu ihrem Produkt vereinbart hat. Mit Mystery Shoppern kann zeitgleich im gesamten Filialnetz überprüft werden, ob die Beratung vereinbarungsgemäß erfolgt oder ob aufgrund von Nichterfüllung eine Preisreduzierung verhandelt werden kann. Beide Beispiele machen deutlich, dass die Ergebnisse von Mystery Shopping als Verhandlungsposition eingesetzt werden können, um eine Preisanpassung zu erreichen oder die Umsetzungsqualität sicherzustellen.

2.3.2 Exkurs: Make or Buy bei Mystery Shopping?

Hat sich ein Unternehmen entschlossen, Mystery Shopping durchzuführen, stellt sich die grundsätzliche Frage, ob die Durchführung selbständig (make) oder mit Hilfe eines externen Anbieters (buy) erfolgen soll. In der Literatur wird diese Entscheidungsfindung lediglich am Rande und insbesondere hinsichtlich qualitativer Attribute diskutiert. Beispielsweise wird die Möglichkeit des Zugriffs auf einen großen Testerstab als positiv beurteilt und der mangelnde

Einfluss auf den Testerstab als negativ angesehen.²⁷⁰ Daneben wird als quantitative Größe der Agenturaufschlag angeführt, der bei eigener Durchführung eingespart werden könnte.²⁷¹ Ferner erfolgt eine Empfehlung für die Erbringung durch externe Unternehmen, da die Zustimmung des Betriebsrates gem. § 94 BetrVG dann nicht erforderlich ist.²⁷² Es ist zu konstatieren, dass in der Literatur bisher keine fundierte Diskussion des Für und Wider einer internen oder externen Durchführung erfolgt ist. Dessen ungeachtet wird die Beauftragung eines externen Anbieters empfohlen oder als gesetzt angenommen. Dies ist nicht verwunderlich, da die Autoren solcher Beiträge häufig Anbieter von Mystery Shopping-Dienstleistungen sind.²⁷³

Zur Beurteilung einer Make-or-Buy-Entscheidung bieten sich unterschiedliche Ansätze an. Bei kostenrechnerischen Ansätzen werden die Kosten der Make-Alternative mit denen der Buy-Alternative (in der Regel in Form des Marktpreises) verglichen.²⁷⁴ Kritisch angemerkt wird die Monokausalität, da bei diesem Ansatz zusätzliche Entscheidungskriterien innerhalb des Zielsystems des Entscheidungsträgers nicht berücksichtigt werden.²⁷⁵ Eine Weiterentwicklung ist die Verwendung eines Zielkataloges oder von Checklisten, die Vor- und Nachteile beider Alternativen gegenüberstellen.²⁷⁶ Problematisch ist die Beliebigkeit der Bewertung einzelner Kriterien, da eine fehlende inhaltliche Systematik keine Konzentration aller Kriterien auf ein einzelnes Bewertungskriterium erlaubt.²⁷⁷ Einen weiteren Ansatz stellt die Transaktionskostentheorie dar.²⁷⁸ Diese wird zunächst abstrakt dargestellt und schließlich in ihrer Anwendung anhand eines Beispiels verdeutlicht.

2.3.2.1 Grundlagen der Transaktionskostentheorie

Grundeinheiten der Analyse sind *Transaktionen*, nämlich die Übertragung von so genannten Verfügungsrechten an Gütern und Dienstleistungen. Da die Übertragung der Verfügungsrechte dem materiellen Austausch vorausgeht, sind Transaktionen die impliziten und expliziten (Vertrags-)Verhandlungen.

²⁷⁰ Vgl. z.B. Landwehr (1984: 26), Deges (1992: 94)

²⁷¹ Vg. Calvert (2005: 25)

²⁷² Vgl. Deckers/Deckers (2004)

²⁷³ z.B. Collins (2003: 30), Gohr (2006: 32), Almhofer (2007: 34)

²⁷⁴ Vgl. z.B. Picot (1991: 340)

²⁷⁵ Vgl. ebenda (1991: 341f.)

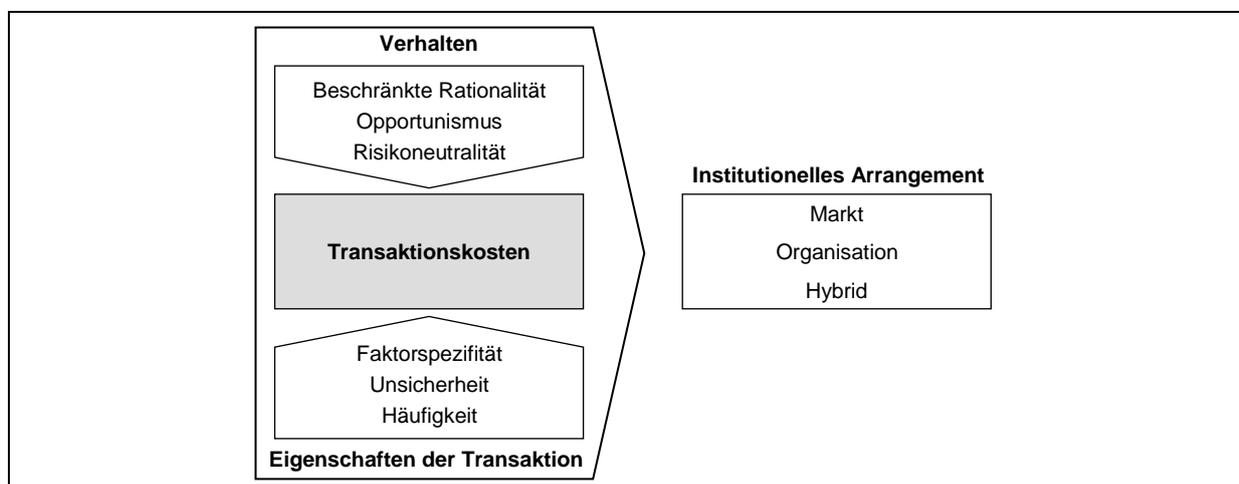
²⁷⁶ Vgl. ebenda (1991: 342f.)

²⁷⁷ Vgl. Haase (2002: 116)

²⁷⁸ Vgl. insbesondere Williamson (1975 und 1985), Nienhüser/Jans (2004), Ebers/Gotsch (2006: 277ff.) und die dort zitierte Literatur für die folgenden Ausführungen. Zur Eignung des Ansatzes vgl. Williamson (1985: 41)

Die vor (*ex ante*) und nach (*ex post*) dem Vertragsabschluss entstehenden Kosten sind die *Transaktionskosten* (Anbahnungskosten, Vereinbarungskosten, Abwicklungskosten, Kontrollkosten und Anpassungskosten). Die Produktionskosten, die für die eigentliche Erstellung der Güter und Dienstleistungen anfallen, werden bei der Transaktionskostentheorie konstant gehalten (implizite *ceteris-paribus*-Bedingung). Es wird daher im Prinzip ausschließlich untersucht, wie sich zwei Alternativen hinsichtlich der Transaktionskosten unterscheiden, gleichwohl am Ende die Summe der Alternativen verglichen wird. Die Transaktionskosten sind das zentrale Element der Transaktionskostentheorie. Sie werden durch das Verhalten der Vertragspartner sowie unterschiedliche Eigenschaften beeinflusst und resultieren in der Wahl eines bestimmten institutionellen Arrangements (vgl. Abbildung 2-3-2).

Abbildung 2-3-2: Schematische Darstellung der Transaktionskostentheorie



Quelle: Eigene Darstellung

Die Transaktionskostentheorie geht von drei unterschiedlichen *Verhaltensannahmen* aus. Den Akteuren wird erstens eine *beschränkte Rationalität* unterstellt, was sich trotz Anstrebens eines rationalen Verhaltens in einem nur unvollkommen rationalen Handeln äußert. Hieraus resultieren Probleme in der Vertragsgestaltung, da insbesondere in einer unsicheren und komplexen Umwelt zukünftige Einflussfaktoren nicht antizipiert werden können. Zweitens wird den Akteuren *Opportunismus* unterstellt, der sich in einem strategischen Verhalten gegenüber dem Vertragspartner äußert. So können unvollständige Verträge, etwa aus Informationsasymmetrien, zu einer nachvertraglichen Durchsetzung des Eigeninteresses eines Akteurs führen. Drittens wird den Akteuren (zur analytischen Vereinfachung aber dennoch kontrafaktisch) eine *Risikoneutralität* unterstellt. Die Akteure stehen den Gestaltungsalternativen demnach neutral gegenüber und unterscheiden sich nicht in ihrer Neigung, ein Risiko einzugehen. Bei der Gestaltung von Transaktionen sollen Unsicherheitsprobleme, die aus beschränkter Rationalität resultieren, weitestgehend minimiert werden, und es soll einem möglicherweise eintretenden opportunistischen Verhalten durch entsprechende Vorkehrungen entgegengewirkt werden.

Vor dem Hintergrund der Verhaltensannahmen werden die Transaktionskosten insbesondere durch drei *Faktoren* beeinflusst. Der bedeutendste ist die *Faktorspezifität*. Eine Transaktion kann die Investition in eine produktspezifische Fertigungsanlage oder eine unternehmensspezifische Qualifikation erforderlich machen. Die damit gewonnenen Spezialisierungsvorteile können in komparativen Kostenvorteilen und somit in sinkenden Produktionskosten resultieren. Im Gegensatz dazu können die Transaktionskosten durch eine solche transaktionsspezifische Investition steigen. Eine Zunahme der Spezifität der eingesetzten Faktoren ändert die Beziehung der Akteure, da ein Wechsel zu einem anderen Aktionspartner schwieriger wird. Entweder gehen bei einem Wechsel die Erlösdifferenz zur nächsten Verwendungsmöglichkeit der Faktoren (Quasi-Rente) und/oder die induzierten Kostenvorteile verloren. Vor dem Hintergrund der genannten Verhaltensannahmen bestehen ferner Anreize, Abhängigkeiten ex-post auszunutzen, indem sich die Quasi-Rente des Transaktionspartners, z.B. durch erhöhte Preise, angeeignet wird. Dies resultiert in erneuten Vertragsverhandlungen und somit in erhöhten Transaktionskosten, oder entsprechende Vertragsverhandlungen wurden bereits ex-ante mit ebenfalls verbundenen Transaktionskosten geführt. Eine weitere die Kosten beeinflussende Größe ist *Unsicherheit*. Parametrische Unsicherheit bezieht sich auf zukünftige Umweltzustände und ihren Einfluss auf die Transaktion, deren Unkenntnis die Akteure aufgrund der unterstellten beschränkten Rationalität nicht in die Lage versetzt, diese entsprechend in die Vertragsgestaltung bzw. die Institution einzubinden. Verhaltensunsicherheit besteht hinsichtlich des opportunistischen Verhaltens der Akteure, welches insbesondere durch Informationsasymmetrien begünstigt wird. Während kein Einfluss der Unsicherheit auf die Produktionskosten angenommen wird, steigen die Transaktionskosten bei beiden Formen der Unsicherheit ex-ante und ex-post. Zum einen müssen unterschiedliche Umweltzustände ex-ante antizipiert werden, und zum anderen ist eine Anpassung und Kontrolle der Einhaltung ex-post erforderlich. Spezifische Investitionen fördern diesen Effekt durch eine erhöhte Motivation der Akteure, Situationen zu spezifizieren oder nachzuverhandeln. Ein dritter Einflussfaktor der Kosten ist die *Transaktionshäufigkeit*. Je häufiger eine identische Transaktion vorgenommen wird, desto mehr besteht die Möglichkeit (Fix-)Kostendegressions-, Skalen- und Synergieeffekte zu realisieren. Dies resultiert bei Nutzung in sinkenden Produktions- und Transaktionskosten. Alle anderen Einflussfaktoren auf die Kosten, wie z.B. politische und rechtliche Rahmenbedingungen, werden in der Gruppe der *Transaktionskostenatmosphäre* zusammengefasst und zur Vereinfachung aus der Analyse ausgeklammert. Die Kostenwirkungen der Transaktionsmerkmale sind in Abbildung 2-3-3 dargestellt.

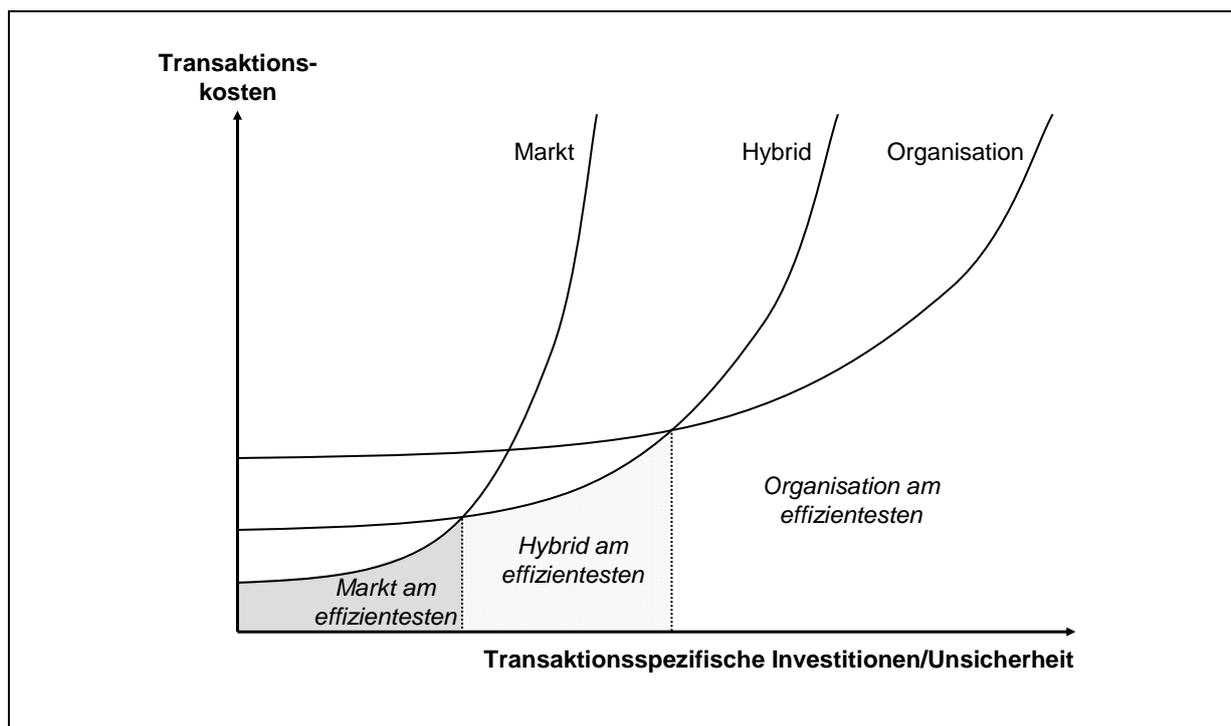
Abbildung 2-3-3: Kostenwirkungen c.p. zunehmender Ausprägungen der Transaktionsmerkmale

	Faktorspezifität	Unsicherheit	Häufigkeit
Produktionskosten			
Transaktionskosten			

Quelle: Eigene Darstellung in Anlehnung an EBERS/GOTSCH (2006: 284)

Letztlich empfiehlt die Transaktionskostentheorie je nach Ausprägung der Kosteneinflussfaktoren für unterschiedliche Situationen drei *institutionelle Arrangements*. Ist eine Situation dadurch gekennzeichnet, dass keine spezifischen Investitionen erforderlich und vereinbarte Leistungen leicht zu messen und kontrollieren sind, wird der *Markt* als institutionelles Arrangement empfohlen. Die nicht-spezifische Leistungserstellung ermöglicht durch Konkurrenzdruck einen jederzeitigen Wechsel des Transaktionspartners und eine autonome Anpassung der Transaktion bzw. deren Koordinationsform. Ferner sind keine zusätzlichen formalen Steuerungs- und Kontrollmechanismen zum Schutz vor Opportunismus und den Folgen beschränkter Rationalität erforderlich. Dies hält die Transaktionskosten im Vergleich zum institutionellen Arrangement *Organisation* (auch Hierarchie genannt) gering. Die organisationsinterne Leistungserstellung ist empfehlenswert, wenn Transaktionen mit spezifischen Investitionen und/oder die Leistungen mit hoher Unsicherheit verbunden und dadurch schwer zu messen und kontrollieren sind. Ex-ante können Transaktionskosten, die für Informationsbeschaffung, Verhandlungen und Verträge anfallen, eingespart werden. Ex-post sind Anpassungen intern mit geringeren Kosten verbunden. Ferner kann durch ein organisationseigenes Steuerungs- und Kontrollsystem opportunistischem Verhalten weitestgehend entgegengewirkt werden. Es wird deutlich, dass mit zunehmender Spezifität und Unsicherheit eine hierarchische Koordinationsform als effizienter anzusehen ist. Neben den Reinformen Markt und Organisation existiert eine *Hybridform*, die mittlere oder jeweils in die eine oder andere Richtung tendierende Ausprägungen aufweist. Sie ist beispielsweise mit langfristigen Vertragsbeziehungen zu beschreiben. Der Zusammenhang zwischen den entstehenden Transaktionskosten und den transaktionsspezifischen Investitionen sowie der Unsicherheit in den unterschiedlichen institutionellen Arrangements wird in Abbildung 2-3-4 verdeutlicht.

Abbildung 2-3-4: Kostenwirkungen c.p. zunehmender Ausprägungen der Transaktionsmerkmale



Quelle: Eigene Darstellung in Anlehnung an EBERS/GOTSCH (2006: 294)

Nachdem die Transaktionskostentheorie abstrakt skizziert wurde, wird sie nachfolgend anhand ihrer grundlegenden Prämissen am Beispiel einer Make-Or-Buy-Entscheidung, hinsichtlich der Durchführung von Mystery Shopping demonstriert.²⁷⁹

2.3.2.2 Mystery Shopping aus transaktionskostentheoretischer Perspektive

Ausgangspunkt ist die eingangs formulierte Frage, ob Mystery Shopping von einem externen Anbieter (buy-Alternative) oder selbständig (make-Alternative) durchgeführt werden soll. Im Folgenden wird eine vollständige Ausgliederung der Phasen „Operative Vorbereitung“ und „Datenerhebung“ betrachtet: Ein Unternehmen beauftragt eine entsprechende Agentur damit, auf Basis des formulierten Fragebogens und demografischer Testervorgaben in den vorgegebenen Filialen Daten zu erheben und diese zurückzuspielen. Durch eine Ausgliederung ließen sich Betriebskosten senken, indem die für diese Phasen erforderlichen Kapital- und Humanressourcen freigesetzt würden. Allerdings stehen dem Transaktions- und Delegationskosten aufgrund der eingegangenen Kooperationsbeziehung gegenüber. Die Frage nach der Vorteilhaftigkeit der Ausgliederung muss daher auf Basis der Transaktionskosten beantwortet werden.

²⁷⁹ Diese Demonstration erfolgt aufgrund des Rahmens als Exkurs lediglich rudimentär. So werden Weiterentwicklungen der Theorie nicht berücksichtigt. Es besteht daher kein Anspruch auf Allgemeingültigkeit.

Die *Spezifität* einer Transaktion lässt sich aus der Wertdifferenz zwischen beabsichtigter und zweitbesten Verwendungsmöglichkeit bestimmen. Sofern Mystery Shopping stets nahezu identisch durchführbar ist, wäre von einer niedrigen Spezifität auszugehen. Eine generelle Bestimmung der Spezifität von Mystery Shopping ist jedoch nicht möglich. Vielmehr hängt diese von der Art der Zielsetzung und somit von der Durchführung ab. Sofern es sich um eine Implementierungsform handelt, die sehr individuell auf das beauftragende Unternehmen zugeschnitten ist, ist von einer hohen Spezifität auszugehen. Die Individualität kann sich an einer sehr speziellen Demografie (z.B. nur Testkunden, die über eine Kundenkarte verfügen), an zu verwendenden Hilfsmitteln (z.B. Gewichts- oder Temperaturmessgeräte in der Gastronomie), am Umfang von Schulungsmaßnahmen (z.B. Vermittlung von Kenntnissen über komplexe Finanzdienstleistungsprodukte) und letztlich an der Anzahl einzusetzender Tester festmachen lassen. Je individueller ein Mystery Shopping gestaltet wird, desto mehr steigt die Spezifität, da eine problemlose wiederholte Durchführung bei einem anderen Unternehmen nicht mehr ohne weiteres gegeben ist. Die erhöhte Spezifität resultiert in einem gegenseitigen Abhängigkeitsverhältnis, was wiederum einen Anstieg der Transaktionskosten zur Folge hat: Eine aufwendige Vertragsgestaltung ist erforderlich, um opportunistischem Verhalten vorzubeugen und um antizipierten Unsicherheiten entgegenzuwirken.

Die *Unsicherheit* einer Transaktion basiert auf der Schwierigkeit der Informationsgewinnung und der Informationsverarbeitung. Durch den relativ jungen Markt für Mystery Shopping-Dienstleistungen haben Anbieter bisher kaum Renommee aufbauen können, es sei denn die Reputation würde sich aus bereits langfristig erbrachter anderer Leistungen, wie etwa Marktforschung, übertragen lassen. Daher besteht insbesondere bei relativ jungen Marktteilnehmern z.B. hinsichtlich der Qualität der erbrachten Leistungen oder der Qualität der eingesetzten Tester ein hohes Maß an Unsicherheit. Folglich fallen Transaktionskosten bei der Gewinnung von Informationen über den Anbieter an. Des Weiteren entstehen bei der Nutzung Transaktionskosten für Beurteilung und Kontrolle, z.B. bei der Durchführung von Stichproben. Zum gegenwärtigen Zeitpunkt muss demnach noch mit hohen Kontrollkosten gerechnet werden. Nach Etablierung und einem damit verbundenen Aufbau eines Renommées und einer Glaubwürdigkeit der Anbieter können diese gesenkt werden und bestehen nur noch für neu in den Markt eintretende Anbieter. Eine weitere Unsicherheit besteht hinsichtlich zukünftiger Modifikationen der Tests, die sich beispielsweise aufgrund einer Verlagerung der zu erhebenden Schwerpunkte ergeben könnten. Sofern es erforderlich wird, die Durchführung zu verändern, muss im Rahmen einer Anpassung nachverhandelt werden und es entstehen Transaktionskosten.

Die *Häufigkeit* ist eine weitere Kosteneinflussgröße, die bei zunehmender Anzahl identischer Transaktionen die Realisierung von (Fix)Kostendegressionspotenzial sowie Potenzial für Skalen- und Synergieeffekte ermöglicht. Kosten bei Mystery Shopping resultieren beispielsweise aus der Erstellung des Beobachtungsbogens, der Identifikation von geeigneten Testern, der Durchführung von Schulungsmaßnahmen und durch das Set-up der

Untersuchung im EDV-System des Anbieters. Bei wiederholter Durchführung kann von einer Reduzierung dieser Positionen und zudem der Steuerungs- und Kontrollkosten im Sinne eines Degressionseffektes ausgegangen werden.

Geht man davon aus, dass die drei Kosteneinflussfaktoren Spezifität, Unsicherheit und Häufigkeit entweder hoch oder niedrig sein können, so gibt es insgesamt acht mögliche Transaktionskonstellationen. Geht man ferner davon aus, dass Mystery Shopping nicht nur einmalig sondern mehrfach durchgeführt werden soll, um beispielsweise einen langfristigen Leistungsmonitor zu etablieren, verbleiben vier denkbare Konstellationen (vgl. Tabelle 2-3-1; Transaktionen mit mehrfacher Durchführung sind schattiert dargestellt).

Tabelle 2-3-1: Konstellationen der Kosteneinflussfaktoren

Spezifität	hoch	hoch	hoch	hoch	niedrig	niedrig	niedrig	niedrig
Unsicherheit	hoch	hoch	niedrig	niedrig	hoch	hoch	niedrig	niedrig
Häufigkeit	hoch	niedrig	hoch	niedrig	hoch	niedrig	hoch	niedrig

Quelle: Eigene Darstellung

Für die beiden Konstellationen „hohe Spezifität/hohe Unsicherheit“ sowie „niedrige Spezifität/niedrige Unsicherheit“ kann aus transaktionskostentheoretischer Perspektive ohne weiteres eine relative Vorteilhaftigkeit der Make-or-Buy-Entscheidung abgeleitet werden, da die Ausprägung „hoch“ jeweils mit hohen und die Ausprägung „niedrig“ jeweils mit niedrigen Transaktionskosten verbunden ist. Die relative Vorteilhaftigkeit der Make-Alternative oder der Buy-Alternative bei den beiden Konstellationen hinsichtlich der Transaktionskosten ist in Tabelle 2-3-2 abgebildet.

Tabelle 2-3-2: Relative Vorteilhaftigkeit bei unterschiedlichen Konstellationen

	Make	Buy
Spezifität hoch	Interne, leichter handhabbare Nachverhandlungen zu erwarten	Nachverhandlungen sind zu erwarten, da eine Abhängigkeit vom Anbieter besteht
Unsicherheit hoch	<ul style="list-style-type: none"> -Kenntnisse über Leistungsfähigkeit und -qualität des eigenen Unternehmens sind vorhanden. -Anpassungen der Maßnahme sind mit geringem Aufwand verbunden Kosten: -Informationskosten geringer -Kontroll- und Überwachungskosten geringer -Vertragsgestaltung einfacher -Keine Gefahr der Preiserhöhung -Geringere Gefahr der Qualitätsverschlechterung 	<ul style="list-style-type: none"> -Keine Kenntnisse über den Anbieter -Anpassungen sind mit Nachverhandlungen verbunden Kosten: -Informationskosten über den Anbieter sind hoch -Kontroll- und Überwachungskosten hoch -Komplizierte Vertragsgestaltung -Gefahr der Preiserhöhung -Gefahr der Qualitätsverschlechterung -Hohe Nachverhandlungskosten bei Anpassungen
	<i>Transaktionskosten insgesamt geringer (=vorteilhaftere Option)</i>	<i>Transaktionskosten insgesamt höher</i>
Spezifität niedrig	Keine Nachverhandlungen zu erwarten	Keine Nachverhandlungen zu erwarten, da keine Abhängigkeit besteht (Anbieterwechsel ist ohne weiteres möglich)
Unsicherheit niedrig	<ul style="list-style-type: none"> -Kenntnisse über Leistungsfähigkeit und -qualität des eigenen Unternehmens sind vorhanden. -Anpassungen der Maßnahme sind mit geringem Aufwand verbunden Kosten: -Informationskosten gering -Kontroll- und Überwachungskosten gering -Vertragsgestaltung einfacher -Keine Gefahr der Preiserhöhung -Geringe Gefahr der Qualitätsverschlechterung 	<ul style="list-style-type: none"> -Spezielle Kenntnisse über den Anbieter sind nicht erforderlich -Anpassungen sind mit geringem Aufwand verbunden Kosten: -Informationskosten kaum relevant -Kontroll- und Überwachungskosten gering -Einfache Vertragsgestaltung -Gefahr der Preiserhöhung und Qualitätsverschlechterung bestehen, aber Anbieter kann leicht gewechselt werden -Kaum Nachverhandlungskosten, da Anbieter gewechselt werden kann
	<i>Transaktionskosten insgesamt höher</i>	<i>Transaktionskosten insgesamt geringer (=vorteilhaftere Option)</i>

Quelle: Eigene Darstellung

Die Vorteilhaftigkeit der verbleibenden zwei Konstellationen wäre erst eindeutig zu bestimmen, wenn sie mit entsprechenden quantitativen Größen unterlegt würden. Schließlich könnte eine niedrige Spezifität sich in den Transaktionskosten stärker niederschlagen als eine hohe Unsicherheit. Bei gleicher Höhe wäre eine Vorteilhaftigkeit zugunsten der Make- oder der Buy-Alternative hinsichtlich der Transaktionskosten nicht nachzuweisen.

Festzuhalten bleibt, dass ein generelles Outsourcing angesichts der entstehenden Transaktionskosten nicht sinnvoll erscheint. Grundsätzlich ist Mystery Shopping bei niedriger

Spezifität und Unsicherheit aus Sicht der Transaktionskostentheorie extern vorteilhafter und bei hohen Ausprägungen intern vorteilhafter ausführbar. Allenfalls ist bei hohen Ausprägungen eine sehr enge Kooperation mit einem einzigen externen Anbieter denkbar. Dadurch ließen sich Effizienzvorteile auf Anbieterseite nutzen und gleichzeitig Anpassungskosten minimieren. Gleichwohl sind die Kontrollkosten durch die Unsicherheit hoch. Solange junge Anbieter nicht ausreichend Renommee aufgebaut haben, erscheint eine Ausgliederung aus Sicht der Transaktionskostentheorie nicht sinnvoll. Vor dem Hintergrund der ausgeführten qualitativen Betrachtung ist es zudem erforderlich, die Einflussgrößen mit quantitativen Größen zu unterlegen. Erst dann kann eine mögliche Vorteilhaftigkeit der einen oder anderen Variante nachgewiesen werden.

Eine Make-or-Buy-Entscheidung auf alleiniger Grundlage der Transaktionskosten greift allerdings zu kurz. Weitere Einflussfaktoren sind bei der Entscheidungsfindung zu berücksichtigen. Zum einen sind die Herstellungskosten eines externen Anbieters denen eines internen Anbieters gegenüberzustellen. Der Anbieter könnte die Leistung möglicherweise wesentlich rentabler erbringen, wenn man beispielsweise die Vorgehensweise bei der Gewinnung und Pflege der Tester näher betrachtet. Schließlich hat der Auftraggeber möglicherweise einen Vorteil, wenn er nur ein einziges Vertragsverhältnis mit dem Anbieter einzugehen hat, statt viele mit den jeweiligen Testern. Des Weiteren ist ein Zielkonflikt zwischen Anbieter und Auftraggeber anzunehmen. Der Anbieter verfolgt möglicherweise das Ziel, seinen Umsatz zu maximieren, indem er bei einer Abrechnung pro Testfall die Anzahl der durchgeführten Tests erhöht. Dies hätte gleichzeitig einen Einfluss auf den Nutzen des Auftraggebers. Ferner ist ein Anbieter an einer langfristigen Bindung interessiert und ist dadurch möglicherweise zu Preiszugeständnissen bereit. Dieses könnte wiederum durch opportunistisches Verhalten des Auftraggebers genutzt werden, um seinen Nutzen zu erhöhen.

Zur weiteren Spezifizierung der Fragestellung – und um somit eine fundiertere Make-or-Buy-Entscheidung als hier dargelegt fällen zu können – sollten daher zum einen weitere Perspektiven hinzugezogen werden, die sich auf die Make-or-Buy-Fragestellung im Allgemeinen und Kooperationsbeziehung im Besonderen, anwenden lassen.²⁸⁰ Dabei sollte idealerweise Bezug auf die Art der Durchführung genommen werden. Zum anderen sollte die Transaktionskostentheorie in einer detaillierteren Form, insbesondere im Hinblick auf Weiterentwicklungen, Berücksichtigung finden. Insofern ist der Einblick, der in diesem Abschnitt gegeben wurde, lediglich eine Annäherung an die Make-or-Buy-Fragestellung und daher als Exkurs zu betrachten.

²⁸⁰ z.B. die Prinzipal-Agent-Theorie

2.3.3 Ethische Aspekte von Mystery Shopping

Mit der Durchführung einer Mystery Shopping-Untersuchung gehen unterschiedliche bedeutsame ethische Aspekte einher. Diese werden bei der Durchführung im eigenen Unternehmen insbesondere durch die möglichen Konsequenzen für die Mitarbeiter deutlich. Im Rahmen der Durchführung einer Wettbewerbsbeobachtung beziehen sie sich auf mögliche ökonomische Auswirkungen.

Mystery Shopping richtet sich als Verfahren zur Überprüfung der Dienstleistungsqualität in der praktischen Anwendung insbesondere auf Schwächen in Form von Verhaltensfehlern einzelner Mitarbeiter.²⁸¹ Dabei wird gemäß theoretischer Beiträge die übergeordnete Zielsetzung verfolgt, positives Verhalten zu verstärken und nicht negatives zu bestrafen.²⁸² Allerdings kann diese Zielsetzung der Motivation in der praktischen Umsetzung nicht bestätigt werden, da es hier vorwiegend als Diagnose- denn als Motivationsinstrument eingesetzt wird.²⁸³ Es überrascht daher nicht, dass Mitarbeiter den Einsatz des Instruments als Kontrolle und nicht als Motivation auffassen und sich bereits Gerüchte über ein anstehendes Mystery Shopping negativ auf das Betriebsklima auswirken.²⁸⁴ Diese Auffassung wird durch die Wahrnehmung des Verfahrens in der Öffentlichkeit noch geschürt, in der das Ziel „Service-Sünder“²⁸⁵ zu entlarven propagiert wird oder Tester als „Undercover-Agenten“²⁸⁶ oder „Mysteriöse Kunden“²⁸⁷ bezeichnet werden. Eine entsprechende an die Mitarbeiter gerichtete offene Kommunikation über Zielsetzung, Verwendung der Daten und damit verbundene Folgen für die Mitarbeiter ist daher bedeutsam.²⁸⁸ Sie sollen sich auf die Erbringung ihrer Leistung in der Kundenkontaktsituation konzentrieren und sich nicht ablenken lassen, indem sie sich gedanklich auf ein schlechtes Abschneiden und eine mögliche Kündigung einstellen.²⁸⁹ Gleichwohl ist der Einsatz von Mystery Shoppern bedenklich, denn ihr Urteil kann zu disziplinarischen Maßnahmen oder gar zur Kündigung führen²⁹⁰ oder den leistungsbezogenen Gehaltsbestandteil beeinflussen.²⁹¹ Marktforschungsverbände haben die mögliche Willkür, die Mitarbeiter durch die Beurteilung ausgesetzt sein können, erkannt und

²⁸¹ Vgl. Semel (2006: 99)

²⁸² Vgl. z.B. Erstad (1998: 34)

²⁸³ Vgl. Beck/Miao (2003: 34), die in einer Befragung der US-amerikanischen Hotelbranche insbesondere die Verwendung als Diagnoseinstrument festgestellt haben.

²⁸⁴ Vgl. Nerdinger (1994: 212)

²⁸⁵ Timpe (2005)

²⁸⁶ Paersch (2006)

²⁸⁷ Schott (2007: 25)

²⁸⁸ Vgl. Guber/Kerper (2003:62)

²⁸⁹ Vgl. Dorman (1994: 18)

²⁹⁰ Vgl. Deges (1992: 95)

²⁹¹ Vgl. Dawson/Hillier (1995: 423)

Richtlinien zur Durchführung von Mystery Shopping formuliert.²⁹² Dabei stimmen sie insbesondere in zwei Punkten überein. Erstens sollen keine Rückschlüsse auf einzelne Mitarbeiter möglich sein, um etwa disziplinarische Abmahnungen zu verhindern. Zweitens soll die Zeit eines Mitarbeiters nicht über Gebühr in Anspruch genommen werden, um etwa leistungsabhängige Gehaltsbestandteile nicht zu gefährden. Letztendlich könnte der Mitarbeiter die Zeit verwenden, um einen echten Kaufabschluss herbeizuführen. Allerdings gibt es Indizien, dass sich in der Praxis nur selten an die Richtlinien gehalten wird.²⁹³ Da Mystery Shopping zum Teil mit dem eindeutigen Ziel der individuellen Mitarbeiterüberprüfung in Auftrag gegeben wird, werden die Richtlinien gar als „Augenauswischerei“ bezeichnet.²⁹⁴ Ferner wird der Einsatz eines direkten Mitarbeitergesprächs im Anschluss an die Kundenkontaktsituation gefordert.²⁹⁵ Die Umsetzung der Kodizes kann auch in Frage gestellt werden, da nur wenige durchführende Unternehmen in Deutschland Mitglied bei den Verbänden sind und sich daher möglicherweise nicht an diese gebunden fühlen.²⁹⁶ Ferner ist bei entsprechendem Wettbewerbsdruck die Annahme auch solcher Aufträge zu vermuten, die von den Richtlinien abgelehnt werden. Allerdings gibt es auch Hinweise, dass der Akzeptanz der Mitarbeiter ein kritischer Stellenwert zugeschrieben wird und daher praktiziert wird.²⁹⁷ Die Einbeziehung von Mitarbeitern in die Konzeptionsphase einer Untersuchung ist hilfreich, um eine „Torpedierung“ der gewonnenen Informationen zu verhindern.²⁹⁸ Ferner können sie als Experten ihrer Prozesse zur Entwicklung des Beobachtungsbogens beitragen.²⁹⁹ Letztlich ist noch auf eine deutsche Besonderheit hinzuweisen, die sich aus dem Mitbestimmungsrecht des Betriebsrates gemäß Betriebsverfassungsgesetz (BetrVG) ergibt. Demnach hat der Betriebsrat gemäß § 94 BetrVG nur in zwei besonderen Fällen die Möglichkeit, der Durchführung zuzustimmen oder an ihr mitzuwirken:³⁰⁰ Sofern das Unternehmen Mitarbeiter als Tester für die Untersuchung einstellt oder wenn die Auswertung auf einzelne Mitarbeiter zurückzuführen ist. Die erste Bedingung kann durch die Beauftragung einer Agentur umgangen werden, die zweite allerdings nicht. Gleichwohl gibt es Hinweise, dass Unternehmen, den Betriebsrat bereits in der Konzeptionsphase einbinden.³⁰¹

²⁹² Vgl. z.B. die Richtlinien folgender Verbände: Arbeitskreis Deutscher Markt- und Sozialforscher (2006), ESOMAR (2005) und Market Research Society (2003)

²⁹³ Vgl. Semel (2006: 99)

²⁹⁴ Vgl. Warmuth/Weinhold (2005: 83)

²⁹⁵ Vgl. Villwock (2004: 55)

²⁹⁶ Vgl. Abschnitt 2.2.1 dieser Arbeit

²⁹⁷ Vgl. Collins (2003: 31)

²⁹⁸ Vgl. Krüger/Stücken (2003: 36)

²⁹⁹ Vgl. Erstad (1998: 35)

³⁰⁰ Vgl. hierzu und im Folgenden Deckers/Deckers (2004)

³⁰¹ Vgl. Pergandé/Stücken (2004: 216)

Im Rahmen eines Mystery Shoppings bei einem Wettbewerber kommen insbesondere ökonomische Auswirkungen zum Tragen. Diese resultieren einerseits aus Einbußen, die entstehen, weil Mitarbeiter des besuchten Unternehmens sich einem Mystery Shopper widmen, statt einem Kunden mit echtem Kaufinteresse.³⁰² Daher gibt es Empfehlungen über Häufigkeit und Dauer eines Mystery Shoppings bei Wettbewerbern.³⁰³ Andererseits resultieren die ökonomischen Auswirkungen aus im guten Glauben an einen Wettbewerber herausgegebenen Informationen, die freiwillig nicht herausgegeben werden würden.³⁰⁴ Dies könnte zum Beispiel der Fall sein, wenn Angebote eingeholt werden, um Kenntnis über die Preispolitik des Wettbewerbers zu erlangen. Gerade bei komplexen Dienstleistungen in einem Markt mit einem Angebotsoligopol kann die Kenntnis schädigend sein, da Wettbewerber in einem Bieterverfahren leicht ausgehebelt werden können. In diesem Zusammenhang wird deutlich, dass Kundenzufriedenheit resultierend aus einer hohen Dienstleistungsqualität als Zielgröße große Bedeutung hat, denn während preispolitische Maßnahmen relativ einfach kopiert werden können, ist dies bei Serviceaspekten ungleich schwieriger.

Aufgrund der dargelegten Bedeutung ethischer Aspekte, könnten sie als fünfte Anforderung in den im Abschnitt 2.2.3.2 dargelegten Anforderungskatalog aufgenommen werden. Die Anforderung wäre wie folgt zu formulieren:

Anforderung 5: Mystery Shopping soll ethisch durchgeführt werden.

Die erforderlichen Voraussetzungen zur Erfüllung dieser Anforderung ergeben sich einerseits aus den Richtlinien der Marktforschungsverbände und andererseits aus den in diesem Abschnitt gezeigten Interessen der betroffenen Mitarbeiter eines getesteten Unternehmens bzw. zur Harmonisierung der Wettbewerbssituation.

Voraussetzung 13: Mitarbeiter sollen in die Konzeptionsphase eingebunden werden.

Voraussetzung 14: Mitarbeiter sollen über die Verwendung der Daten aufgeklärt werden und müssen ihnen zustimmen.

Voraussetzung 15: Mystery Shopping soll bei Wettbewerbern nicht zur Erlangung von sensiblen Informationen eingesetzt werden.

³⁰² Vgl. Dawson/Hillier (1995: 423), Turner (2007: 343)

³⁰³ Vgl. Morrall (1994: 14), Dawson/Hillier (1995: 426), Wilson (1998a: 159)

³⁰⁴ Vgl. Shing/Spence (2002: 351)

2.3.4 Erfolgsfaktor Testerstab

Wie bereits nach Durchsicht der Forschungsbeiträge deutlich wurde, existieren keine gesicherten Informationen hinsichtlich der Zusammensetzung der einzusetzenden Feldmannschaft und ihrer Merkmale. Nachfolgend sollen zwei Aspekte, die einzusetzende Tester betreffen, näher betrachtet werden: Zum einen ist dies der Schulungsgrad der Tester und zum anderen deren soziodemografische Merkmale.

2.3.4.1 Schulungsgrad der Mystery Shopper

Für den Einsatz „im Feld“ können Tester auf unterschiedliche Weise geschult werden. Die eingesetzten Maßnahmen bestimmen dabei den Schulungsgrad. Die Maßnahmen reichen von einfachen schriftlichen Verhaltensanweisungen, deren Verständnis zusätzlich mit einem Fragebogen überprüft werden kann, über eine persönliche Schulung bis hin zu Rollenspielen mit dem Einsatz von Videokameras.³⁰⁵ Ziel der Schulungsmaßnahmen ist es, subjektive Einflüsse der Tester weitestgehend zu minimieren.³⁰⁶ Neben der Schulung für einen aktuell anstehenden Testfall wird teilweise vom Einsatz eines „Normcontrollerverfahrens“ berichtet. Bei diesem Verfahren werden die Testberichte regelmäßig gegen die Berichte eines speziell geschulten Testers gespiegelt, um festzustellen, ob der Tester gründlich arbeitet und ob er systematisch strenger oder milder bewertet.³⁰⁷ Die Rückmeldung der Abweichungen vom „geeichten“ Normcontroller an die Tester soll zukünftige Schwankungen reduzieren. Allerdings wird in einigen Beiträgen auch angeführt, dass auf eine Schulung verzichtet wird, „um kein für ‚normale‘ Konsumenten atypisches Wahrnehmen der Verkaufssituation zu verursachen.“³⁰⁸ Dabei blieben bei unerfahrenen und untrainierten Testern das natürliche Kundenverhalten sowie eine unbeeinflusste Wahrnehmung des Prozesses der Dienstleistungserstellung erhalten.³⁰⁹ Vor dem Hintergrund der Ausführungen unter Anforderung 1 weiter oben ist der vollständige Verzicht auf Schulungen jedoch anzuzweifeln.³¹⁰

Bei den praktischen Beiträgen wird zum Teil zwischen professionellen Testkunden und Realkunden-Panel unterschieden.³¹¹ Die beiden Typen unterscheiden sich hinsichtlich

³⁰⁵ Vgl. Sander (2000: 32f.), Drees/Schiller (2003: 169), Stücken (2003: 46ff.), Pergandé/Stücken (2004: 214ff.)

³⁰⁶ Vgl. hierzu die Ausführungen zur Anforderung 1 in Kapitel 2.2.3.2

³⁰⁷ Vgl. Deckers (1999: 206), Sander (2000: 33f.), Stücken (2003: 47f.)

³⁰⁸ Haas (2006: 651)

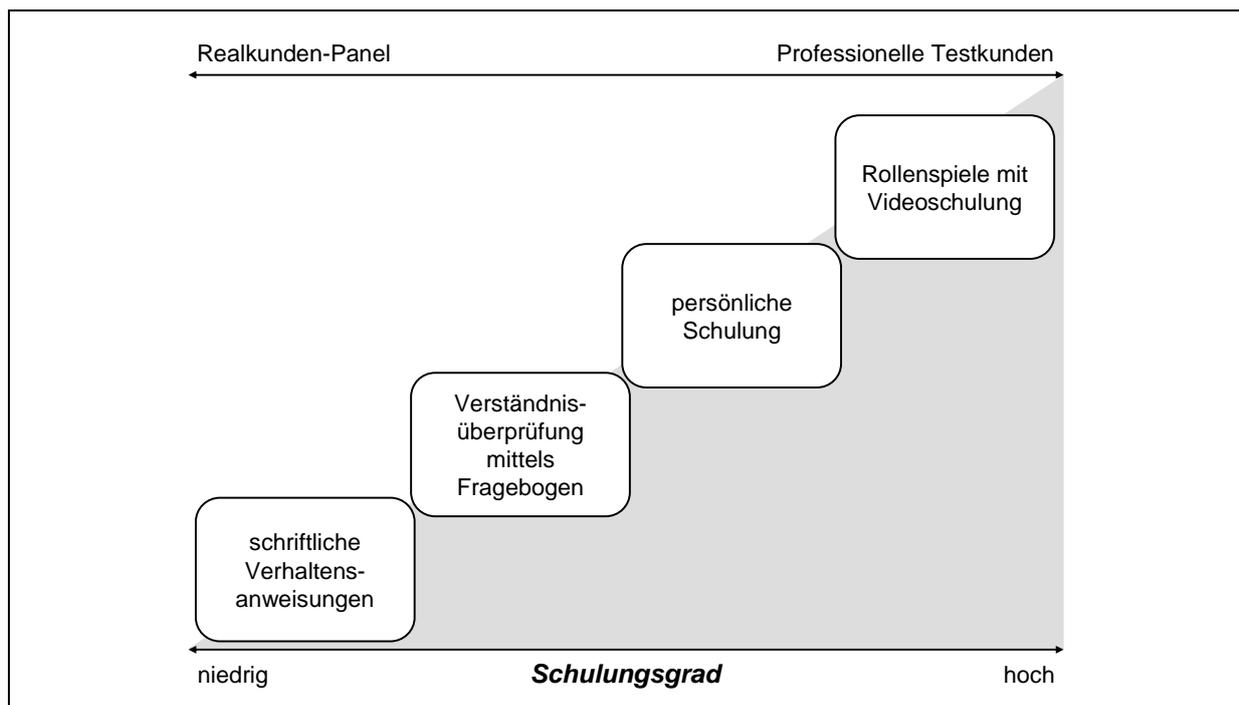
³⁰⁹ Vgl. Warmuth/Weinhold (2005: 35)

³¹⁰ Vgl. hierzu die Ausführungen zur Anforderung 1 in Kapitel 2.2.3.2

³¹¹ Vgl. hierzu und im Folgenden Sander (2000: 32), Stücken (2003: 46ff.), Keller/Stücken (2004: 530), Pergandé/Stücken (2004: 214f.),

Einsatzhäufigkeit, Schulungsgrad und Erfahrung mit Testkäufen, wobei die Ausprägung der Eigenschaften bei professionellen Testkunden hoch und bei denen aus dem Realkunden-Panel eher niedrig ist. Letztere werden vorzugsweise eingesetzt, sofern Auftraggeber bestimmte Anforderungen an Testermerkmale stellen (z.B. Besitzer eines bestimmten Autotyps, bestimmter Wohnort, bestimmtes Einkommen). Ein Realkunden-Panel besteht in der Regel aus über das Internet rekrutierten Personen, die in ihrer Freizeit weniger komplexe Testkäufe in der Nähe ihres Wohnortes durchführen und dafür lediglich schriftlich geschult werden. Der Zugriff auf derartig strukturierte Panels überrascht nicht, da mit zunehmendem Schulungsgrad ein Anstieg der Kosten einhergeht und da solche Tester in der Regel aufgrund einer Hauptbeschäftigung nur über begrenzte Zeit für aufwändige Schulungsmaßnahmen verfügen. Zudem gibt es in Deutschland nur wenige Personen, die ihren Lebensunterhalt mit Mystery Shopping bestreiten, so dass die Anfahrtswege beim Einsatz solcher professionellen Tester höhere Kosten verursachen würden.³¹² Dieser Zusammenhang ist in Abbildung 2-3-5 verdeutlicht.

Abbildung 2-3-5: Ausgewählte Schulungsmaßnahmen und Schulungsgrad bei Mystery Shopping



Quelle: Eigene Darstellung

³¹² Vgl. Semel (2006: 78); In den USA gibt es hingegen offenbar eine Vielzahl von Personen, die als Hauptbeschäftigung einer Tätigkeit als Mystery Shopper nachgehen, wie aus der Vielzahl von Publikationen zu diesem Thema geschlossen werden kann, vgl. z.B. Poynter (2002), Newhouse (2004), Hazlerig (2007); Ferner hat die MSPA 2006 ein Zertifizierungsprogramm für Mystery Shopper eingeführt, vgl. Turner (2007: 336); Da dieses im Übrigen kostenpflichtige Programm in seiner Basisvariante allerdings lediglich online erfolgt, kann der Erfolg eines solchen Programms nicht zwangsläufig als höher eingestuft werden als eine konkrete fallbezogene Schulung. Die aufbauende persönliche Schulungsvariante hat hingegen Kennzeichen einer professionellen allgemeinen Mystery Shopper-Schulung. Vgl. <http://www.mysteryshop.org/shoppers/certification.php> (Stand 30.10.2007)

Die unterschiedlichen Auffassungen bezüglich des erforderlichen Schulungsgrades in den praktischen und theoretischen Beiträgen lassen eine bestehende Ungewissheit vermuten.³¹³ Aufgrund der weiten Verbreitung des alleinigen Einsatzes von schriftlichen Verhaltensanweisungen – verbunden mit einer Eichung auf die Ausprägungen der Dimensionen – kann allerdings angenommen werden, dass ein solches Vorgehen ausreichend ist. Ziel einer Untersuchung könnte es allerdings sein, unterschiedliche Schulungsformen und -grade miteinander zu vergleichen. Dabei sollte auch zwischen einmaligem und wiederholtem Einsatz der Testpersonen differenziert werden. Die daraus gewonnenen Erkenntnisse könnten Unternehmen und Anbietern von Mystery Shopping bei der Konzeption von Untersuchungen und bei der Pflege ihres Testerbestands dienlich sein, um den Aufwand und den Nutzen der Maßnahmen gegeneinander abwägen zu können.

2.3.4.2 Soziodemografische und einstellungsbezogene Merkmale der Mystery Shopper

Bei den Überlegungen zu Anforderung 4 für den erfolgreichen Einsatzes der Methode Mystery Shopping wurde bereits ausgeführt, dass sich unterschiedliche Faktoren im Prozess von Informationsaufnahme, -speicherung und -abruf auf die Ergebnisqualität von Mystery Shopping auswirken können.³¹⁴ Morrison et al. fordern in diesem Zusammenhang ein Forschungsvorhaben, welches sich mit „effects of gender, age, and other demographic factors“ auseinandersetzt.³¹⁵ Bisher blieb die Frage nach den Einflüssen dieser Faktoren unbeantwortet. Die Untersuchung der Einflüsse könnte einerseits zu einem höheren theoretischen Erkenntnisstand führen und andererseits Unternehmen und Anbietern von Mystery Shopping Aufschluss darüber geben, ob bestimmte Gruppen bezogen auf ihre soziodemografischen Merkmale besonders geeignet sind, um als Tester eingesetzt zu werden und wie eine Feldmannschaft bei einem Mystery Shopping zusammensetzen wäre. Nachfolgend werden daher solche Faktoren aufgegriffen, die sich konkret auf die soziodemografischen Eigenschaften der Mystery Shopper beziehen.³¹⁶ Entsprechend werden folgende Hypothesen aufgestellt:

H1: *Männer und Frauen unterscheiden sich beim Mystery Shopping in ihrer Ergebnisqualität.*

³¹³ In diesem Zusammenhang ist anzumerken, dass in einem verwandten Feld, nämlich bei Leistungsbeurteilungen (z.B. von Mitarbeitern oder in einem Assessment Center), die Wirkung von Beurteilertrainings umstritten ist. Vgl. Schuler/Höft (2007: 326)

³¹⁴ Vgl. hierzu Abschnitt 2.2.3.2, Tabelle 2-2-1 sowie Morrison/Colman/Preston (1997)

³¹⁵ Morrison/Colman/Preston (1997: 359)

³¹⁶ Dabei wird berücksichtigt, dass die Zusammenhänge gemäß der obigen Ausführung zwar vermutet werden, bisher jedoch empirisch nicht nachgewiesen worden sind. Somit kann keine Aussage über die konkreten Zusammenhänge getroffen werden. Folglich sind die nachfolgend formulierten Hypothesen als Arbeitshypothesen zu verstehen. Die konkreten Zusammenhänge soll die Datenlage zeigen.

- H2:** *Die Ergebnisqualität von Mystery Shoppern unterschiedlichen Alters unterscheidet sich.*
- H3:** *Die Ergebnisqualität von Mystery Shoppern mit unterschiedlichen Bildungsabschlüssen unterscheidet sich.*
- H4:** *Die Ergebnisqualität von Mystery Shoppern unterscheidet sich bei unterschiedlichen (beruflichen) Tätigkeiten.*

Neben der Ermittlung des Einflusses soziodemografischer Faktoren der Mystery Shopper ist es interessant zu ergründen, wie sich solche Attribute auf die Ergebnisqualität auswirken, die unmittelbar mit der Tätigkeit als Mystery Shopper zusammenhängen, z.B. die Erfahrung mit Testkäufen sowie die Involvierung in das Testobjekt (z.B. die zu testende Dienstleistung oder das zu testende Produkt). Dies würde Mystery Shopping-Agenturen Aufschluss darüber geben, ob beispielsweise besonders erfahrene Tester mit der Durchführung der Testkäufe zu betrauen sind bzw. ob eine Affinität zum Testobjekt sinnvoll oder weniger sinnvoll wäre. Somit lauten die Hypothesen:

- H5:** *Die Ergebnisqualität von Mystery Shoppern mit unterschiedlicher Erfahrung hinsichtlich der Menge der Durchführung von Testkäufen unterscheidet sich.*
- H6:** *Die Ergebnisqualität von Mystery Shoppern mit unterschiedlicher Involvierung bzw. Einstellung zum Testobjekt unterscheidet sich.*

Im vorliegenden Kapitel wurden die Grundlagen von Mystery Shopping dargestellt. Im ersten Schritt wurde die Methode aus theoretischer Perspektive betrachtet. Dabei wurde das Verfahren eingeordnet, die historische Entwicklung aufgezeigt und vergangene Forschungsbemühungen skizziert. Im zweiten Schritt wurde Mystery Shopping aus praktischer Perspektive betrachtet. Dabei wurde nach einer Betrachtung des Marktes und Darstellung von Einsatzmöglichkeiten der Ablauf exemplarisch anhand einer Konzeption aufgezeigt. Im letzten Schritt wurde eine kritische Perspektive eingenommen und sich der Beantwortung unterschiedlicher Fragestellungen genähert. Dabei wurde die von Morrison et al. formulierte Forschungslücke hinsichtlich des Einflusses bestimmter (soziodemografischer) Faktoren auf die Ergebnisqualität von Mystery Shopping aufgegriffen und entsprechende Annahmen getroffen. Nachdem somit den ersten beiden Teilzielen Rechnung getragen wurde, soll sich das folgende Kapitel dem dritten Teilziel der Arbeit widmen. Dafür soll dargelegt werden, wie die hier formulierten Hypothesen empirisch überprüft werden können. Tabelle 2-3-3 fasst die Forschungshypothesen zusammen.

Tabelle 2-3-3: Forschungshypothesen

Gruppe	Hypothesen
Einfluss soziodemografischer Faktoren der Mystery Shopper	H1: Männer und Frauen unterscheiden sich beim Mystery Shopping in ihrer Ergebnisqualität.
	H2: Die Ergebnisqualität von Mystery Shoppern unterschiedlichen Alters unterscheidet sich.
	H3: Die Ergebnisqualität von Mystery Shoppers mit unterschiedlichen Bildungsabschlüssen unterscheidet sich.
	H4: Die Ergebnisqualität von Mystery Shoppern unterscheidet sich bei unterschiedlichen (beruflichen) Tätigkeiten.
Einfluss der Mystery Shopping-Erfahrung und Einstellung zum Testobjekt	H5: Die Ergebnisqualität von Mystery Shoppern mit unterschiedlicher Erfahrung hinsichtlich der Menge der Durchführung von Testkäufen unterscheidet sich.
	H6: Die Ergebnisqualität von Mystery Shoppern mit unterschiedlicher Involvierung bzw. Einstellung zum Testobjekt unterscheidet sich.

Quelle: Eigene Darstellung

3 Konzeption einer empirischen Untersuchung zur Ermittlung der Ergebnisqualität von Mystery Shoppern

Ziel des vorliegenden Kapitels ist die Konzeption einer empirischen Untersuchung, mit der die am Ende des vorangegangenen Kapitels formulierten Arbeitshypothesen überprüft werden können. Hierfür ist es zunächst erforderlich, zu spezifizieren, wie die Ergebnisqualität von Mystery Shoppern beurteilt werden kann. Dann sind ein Testszenario sowie ein entsprechender Beurteilungsbogen zu entwickeln, aus dem die Ergebnisqualität abgeleitet werden kann. Schließlich ist die Erfassung der Ergebnisqualität sowie der Gang der Felduntersuchung zu beschreiben.

3.1 Identifikation von Komponenten zur Beurteilung der Ergebnisqualität von Mystery Shoppern

Zur Beurteilung der Ergebnisqualität von Mystery Shoppern existieren in der Literatur bisher keine Beiträge.³¹⁷ Erstaunlich ist, dass auch die Literatur zur Marktforschung keinen Aufschluss über solche Interviewer oder Beobachter liefert, die aufgrund spezifischer Merkmale, wie beispielsweise Bildung oder Alter, besonders für derartige Tätigkeiten geeignet sind.³¹⁸ Allerdings lassen sich in der Literatur zur Organisations- und Personalpsychologie im Themenbereich der Leistungsbeurteilung, etwa durch Vorgesetzte oder im Rahmen von Assessment-Centern, Hinweise zu Ansätzen zur Ermittlung der Urteilsqualität finden.³¹⁹ Gleichwohl wird dort angeführt, dass „ein wirklich befriedigendes Instrumentarium, die Qualität von Beurteilungen quantitativ zu beschreiben, leider noch aussteht.“³²⁰ Es ist anzunehmen, dass sich auch die Ergebnisqualität von Mystery Shoppern nicht an einem „ultimativen Kriterium“ festmachen lässt. Vielmehr ist zu vermuten, dass sie durch verschiedene als Surrogat dienende Kriterien beschrieben werden könnte. Durch Kombination solcher Indikatoren und Komponenten lassen sich unterschiedliche Facetten des Konstrukts Ergebnisqualität berücksichtigen. Ziel dieses Abschnittes ist es daher, ein Schema zu präsentieren, welches sich aus unterschiedlichen Komponenten zusammensetzt und mit dem sich die Ergebnisqualität von Mystery Shoppern abbilden lässt. Nachfolgend soll dafür

³¹⁷ Es wird sich in unterschiedlichen Beiträgen jedoch mit Reliabilität und Validität des Instruments Mystery Shopping auseinandergesetzt; Vgl. Deckers (1999), Finn/Kayandé (1999), Dawes/Sharp (2000), Schmidt (2007)

³¹⁸ Schließlich setzen Marktforschungsunternehmen in Deutschland mehr als 2,2 Mrd. US\$ um, Vgl. ESOMAR (2007: 14)

³¹⁹ Vgl. für eine Übersicht zu Aspekten der Leistungsbeurteilung, der Qualität solcher Beurteilungen sowie zur Beurteilungsakkuratheit Schuler (2004), Marcus/Schuler (2006), Schuler/Höft (2007), Melchers/Kleinmann (2007) und für eine detaillierte Zusammenfassung der Aspekte Beurteilerübereinstimmung und Beurteilerreliabilität Wirtz/Caspar (2002) und die dort zitierte Literatur

³²⁰ Schuler (2004: 34)

zunächst das im Bereich der Leistungsbeurteilung verwendete quantitative Instrumentarium bestehend aus „Urteilstendenzen“ und „Beurteilungsakkuratheit“ dargestellt werden und in ein Konzept für den Zweck dieser Untersuchung übertragen werden. Anschließend widmet sich ein Abschnitt dem Themenbereich „Zuverlässigkeit von Beurteilungen“. Dann werden ergänzende pragmatische Überlegungen angestellt, wie sich die Ergebnisqualität von Mystery Shoppern abbilden lässt. Schließlich werden die Komponenten zur Abbildung der Ergebnisqualität in einer tabellarischen Übersicht zusammengefasst.

3.1.1 Urteilstendenzen und Beurteilungsakkuratheit

Urteilstendenzen nehmen in der Literatur zur Qualität von Beurteilungen einen breiten Raum ein.³²¹ Sie beziehen sich auf systematische Verzerrungen bei der Verwendung mehrstufiger Bewertungsskalen. Basierend auf den Ausführungen von BRANDSTÄTTER werden insbesondere drei Gruppen von Urteilstendenzen angeführt: Mittelwerts-, Streuungs- und Korrelationstendenzen.³²²

Unterschiedliche Beurteiler haben unterschiedliche Urteilsgehnheiten. So können Beurteiler zu eher milden oder eher strengen Urteilen neigen.³²³ Mit der Bezeichnung *Mittelwertstendenz* werden solche urteilerspezifischen Lokalisierungen von Urteilswerten beschrieben. Die Mittelwertstendenz bezieht sich nicht – wie gelegentlich beschrieben – auf die Neigung von Beurteilern, bei Einstufungen zum Skalenmittelwert zu tendieren.³²⁴ Es lassen sich allgemeine und differenzielle Mittelwertstendenz unterscheiden. Mit ersterer wird der durchschnittliche Mittelwert aller Beurteilungen eines Beurteilers beschrieben, während zweitere als einzelfallbezogene Abweichung genereller Urteilsgehnheiten des Beurteilers auftritt. Für Mystery Shopping-Ergebnisse sind Mittelwertstendenzen insofern interessant, als dass identifiziert werden kann, ob für bestimmte Gruppen spezifische Urteilsgehnheiten bestehen, beispielsweise besonders streng oder besonders mild. Zudem könnte ermittelt werden, ob der Grad einzelfallbezogener Abweichungen der Urteilsgehnheiten zwischen bestimmten Gruppen voneinander abweicht. Interessant erscheint allerdings nur erstgenannter Aspekt. Allerdings hat die Aufdeckung der Zusammenhänge lediglich deskriptiven Charakter. Die bei den Vergleichen der Mittelwerte aufgedeckten Unterschiede sind nicht als kausal zu interpretieren. Gleichwohl kann das Ergebnis Hinweise darauf geben, ob bestimmte Gruppen beispielsweise intensiver geschult werden sollten, denn im Rahmen einer Trainingsmaßnahme

³²¹ so z.B. bei Schuler (2004: 37ff.), Marcus/Schuler (2006: 454ff.), Schuler/Höft (2007: 325ff.)

³²² Vgl. Brandstätter (1970: 668ff.)

³²³ Vgl. hierzu und im Folgenden Schuler (2004: 37), Marcus/Schuler (2006: 454), Schuler/Höft (2007: 325)

³²⁴ Das beschriebene Phänomen wird mit „Tendenz zur Mitte“ oder als „Zentralstendenz“ bezeichnet und ist aufgrund der Häufung um einen Mittelwert eine Streuungstendenz; Vgl. Schuler (2004: 39ff.)

kann den betreffenden Testern Feedback bezüglich ihrer Urteilstendenzen gegeben werden, wobei im Vergleich mit den Urteilen anderer ein Lernprozess verfolgt werden kann.³²⁵

Mit *Streuungstendenz* wird die Ausnutzung der Skalenlänge bezeichnet.³²⁶ Während einige Beurteiler beispielsweise alle Skalenwerte für ihre Beurteilungen nutzen, verwenden andere nur wenige Skalenwerte. So können die Urteile verschiedener Beurteiler in unterschiedlichem Maße streuen. Ferner kann die Streuungstendenz auch von der Art der Kriterien abhängen, so dass verhaltens- und eigenschaftsbezogene Beurteilungen unterschiedliche Streuungen aufweisen können. Analog zu den Mittelwertstendenzen können Mystery Shopping-Ergebnisse hinsichtlich ihrer Streuungstendenzen untersucht werden. Dabei lässt sich identifizieren, ob für bestimmte Gruppen spezifische Urteilsgewohnheiten bestehen, z.B. vergleichsweise einheitlich mit geringer Streuung oder vergleichsweise uneinheitlich mit hoher Streuung. Dies ließe sich auch auf für verhaltens- oder eigenschaftsbezogene Kriterien ermitteln. Die Aufdeckung dieser Zusammenhänge hat ebenfalls einen deskriptiven Charakter und kann, wie bei den Mittelwertstendenzen, ein Indikator für erforderliche Schulungsmaßnahmen sein. Zudem ist bei zwei Gruppen mit unterschiedlicher Streuung anzunehmen, dass diejenige Gruppe mit geringerer Streuung bei gleichem Mittelwert der Gruppen für Mystery Shopping-Untersuchungen als vorteilhafter anzusehen ist, da vergleichsweise weniger Testfälle anzusetzen wären.

Die in der Literatur zur Organisations- und Personalpsychologie im Themenbereich der Leistungsbeurteilung als *Korrelationstendenz* bezeichnete Urteilstendenz³²⁷, beschreibt ein andernorts als Halo-Effekt oder Überstrahlung benanntes Phänomen.³²⁸ Der Effekt hat zur Folge, dass die Beurteilung einer Eigenschaft oder einer Dimension durch die vorangegangene Beurteilung einer anderen Eigenschaft oder Dimension überstrahlt wird. Bei einem Fragebogen kann eine Frage demnach eine andere derart beeinflussen, dass diese im Sinne einer vorangegangenen Frage und nicht mehr für sich selbst beantwortet wird. In der Literatur zur Dienstleistungsqualität besteht grundsätzlich Einigkeit über den Einfluss des Qualitätsurteils vorangehender Dienstleistungsepisoden auf nachfolgende.³²⁹ Ferner wird davon ausgegangen, dass sich das Gesamtqualitätsurteil aus dem Urteil einzelner Episoden und Konstrukten zusammensetzt.³³⁰ Allerdings wird die Stärke des Einflusses der einzelnen Episoden und Konstrukte unterschiedlich beurteilt.³³¹ Wie bei Mittelwerts- und

³²⁵ Vgl. Schuler (2004: 40)

³²⁶ Vgl. hierzu und im Folgenden Schuler (2004: 40f.), Marcus/Schuler (2006: 454f.), Schuler/Höft (2007: 325)

³²⁷ Vgl. Schuler (2004: 41f.), Marcus/Schuler (2006: 455.), Schuler/Höft (2007: 325)

³²⁸ Vgl. z.B. Bortz/Döring (2006: 183)

³²⁹ Vgl. Lingenfelder/Wieseke/Schmidt (2003: 287) sowie ihre Metaanalyse von Studien zur Wahrnehmung der Dienstleistungsqualität

³³⁰ Vgl. Stauss/Seidel (2006: 184)

³³¹ Vgl. ebenda (2006: 184)

Streuungstendenzen lassen sich unterschiedliche Gruppen von Testern hinsichtlich der Intensität des Halo-Effektes untersuchen. Vorstellbar ist, dass die Stärke des Halo-Effekts für verschiedene Gruppen von Mystery Shoppern unterschiedlich ist. Dies wäre, wie bei den anderen beiden Urteilstendenzen auch, ein Indikator für notwendige Schulungsmaßnahmen.³³² Ferner ist folgendes denkbar: Während ein starker positiver Effekt in einer echten Kundenkontaktsituation möglicherweise als vorteilhaft erachtet werden könnte, ist bei Mystery Shoppern ein starker Halo-Effekt eher als nachteilig zu bewerten: Mystery Shopping soll ein umfassendes Bild der Dienstleistungsqualität liefern. Dafür soll der Dienstleistungsprozess in seine einzelnen Phasen zerlegt werden.³³³ Überstrahlen nun einzelne Phasen oder Konstrukte das Urteil anderer Phasen, ist nicht von der Ermittlung eines umfassenden Bildes der Dienstleistungsqualität auszugehen, sondern von einer verzerrten. Folglich sind solche Mystery Shopping-Ergebnisse vorteilhaft, die keinem oder einem vergleichsweise geringen Halo-Effekt unterliegen, sondern die einzelnen Phasen und Konstrukte unabhängig voneinander bewerten. Die wesentlichen Merkmale der Urteilstendenzen sind in nachstehender Tabelle 3-1-1 zusammengefasst.

Tabelle 3-1-1: Urteilstendenzen

Tendenz	Wesen	Detaillierung
<i>Mittelwertstendenz</i>	Urteilerspezifische Lokalisierungen von Werten	Milde, Strenge im Urteil
<i>Streuungstendenz</i>	Ausnutzung der Skalenlänge	(Un-)Einheitlichkeit im Urteil
<i>Korrelationstendenz</i>	Korrelation unterschiedlicher Beurteilungsdimensionen	Überstrahlung des Urteils

Quelle: Eigene Darstellung

Es wird deutlich, dass die hier dargestellten Urteilstendenzen in erster Linie einen deskriptiven Charakter haben. Sind starke Abweichungen zu verzeichnen, kann dies ein Indikator für Schulungsmaßnahmen sein, um bei den Beurteilern einen Impuls für eine Vereinheitlichung des Verhaltens zu setzen. Allerdings ist die Wirkung solcher Beurteilertrainings umstritten.³³⁴ Für Mystery Shopping steht ein Nachweis der Wirkungszusammenhänge von Schulungsmaßnahmen noch aus.³³⁵ Untersucht man nun Mystery Shopping-Ergebnisse auf Urteilstendenzen, lassen sich unterschiedliche Gruppen von Mystery Shoppern unterschiedlich charakterisieren, sofern sie sich in ihren Urteilstendenzen unterscheiden. Die Charakterisierungen stellen einen Teil des

³³² Allerdings ist die Korrelationstendenz dem Beurteilertraining weniger leicht zugänglich als Mittelwerts- und Streuungstendenz; Vgl. Schuler (2004: 41)

³³³ Vgl. die Ausführungen zu Anforderung 2 und Voraussetzung 5 in Kapitel 2.2.3.2 dieser Arbeit

³³⁴ Vgl. Schuler/Höft (2007: 326)

³³⁵ wie in Kapitel 2.3.4.1 dieser Arbeit gezeigt wurde

Bewertungsverhaltens der Mystery Shopper dar und lassen einen Schluss auf die Ergebnisqualität zu. So ist die Ergebnisqualität von Mystery Shoppern, deren Ergebnisse von einer hohen Streuung und einem starken Halo-Effekt gekennzeichnet sind, als niedriger anzusehen als die von solchen, die Ergebnisse mit niedriger Streuung und schwachem Halo-Effekt aufweisen. Allerdings sind die so ermittelten Charakteristika lediglich als *ein* Indikator für die Ergebnisqualität zu verstehen. Weitere Indikatoren werden nachfolgend dargestellt.

Ein weiteres Maß, welches ebenfalls in der Organisations- und Personalpsychologie im Themenbereich der Leistungsbeurteilung zu finden ist, ist die *Beurteilungsakkuratheit*.³³⁶ Der Ansatz ermöglicht die Bestimmung der Genauigkeit von Urteilen, indem die Übereinstimmung mit bzw. Abweichung von einem wahren Leistungswert ermittelt wird. Werden mehrere Personen bezüglich mehrerer Dimensionen beurteilt, lassen sich vier verschiedene Aspekte der Akkuratheit unterscheiden.³³⁷ *Elevation* ist ein über alle beurteilten Personen und Merkmale gemittelter Abstand zwischen den Urteilen eines Beurteilers und dem wahren Wert. So lässt sich abbilden, ob ein Beurteiler über alle Beurteilten und Dimensionen hinweg durchschnittlich angemessen oder milder bzw. strenger urteilt. *Differential Elevation* bezeichnet die Genauigkeit der Einschätzung eines Beurteilten, gemittelt über die verschiedenen Urteilsdimensionen. Damit lässt sich feststellen, ob ein Beurteiler bezogen auf einen Beurteilten angemessen oder zu milde bzw. zu streng urteilt. *Stereotype Accuracy* beschreibt die Genauigkeit des Urteils einzelner Urteilsdimensionen gemittelt über alle Beurteilten. So lässt sich ermitteln, ob bestimmte Dimensionen angemessen oder besonders milde bzw. streng beurteilt werden. *Differential Accuracy* misst die Genauigkeit, mit der einzelne Beurteilte hinsichtlich einzelner Dimensionen beurteilt werden. Die beschriebenen unterschiedlichen Aspekte der Beurteilungsakkuratheit sind in Abbildung 3-1-1 verdeutlicht.

³³⁶ Vgl. hierzu und im Folgenden Schuler (2004: 42f.), Marcus/Schuler (2006: 453f.), Melchers/Kleinmann (2007: 561ff.) und die dort zitierte Literatur

³³⁷ Diese Unterscheidung geht zurück auf einen viel beachteten Beitrag von Cronbach (1955)

Abbildung 3-1-1: Cronbachs Aspekte der Beurteilungsakkuratheit

ELEVATION				DIFFERENTIAL ELEVATION			
	Dimension				Dimension		
	1	...	n		1	...	n
Beurteiler 1				Beurteiler 1			
Beurteiler ...				Beurteiler ...			
Beurteiler n				Beurteiler n			
STEREOTYPE ACCURACY				DIFFERENTIAL ACCURACY			
	Dimension				Dimension		
	1	...	n		1	...	n
Beurteiler 1				Beurteiler 1			
Beurteiler ...				Beurteiler ...			
Beurteiler n				Beurteiler n			

Quelle: Eigene Darstellung

Bezieht man die weiter oben beschriebenen Urteilstendenzen auf die Akkuratheitsmaße, entspricht Elevation am ehesten der allgemeinen und Differential Elevation der differenziellen Mittelwertstendenz. Stereotype und Differential Accuracy entsprechen am ehesten der Korrelationstendenz, wobei erstere sich auf eine Gruppe von Beurteilten und zweitere auf einzelne Beurteilte bezieht. Allerdings ist anzumerken, dass Urteilstendenzen sich auf empirische Verteilungen und Akkuratheitsmaße sich auf wahre Werte beziehen.³³⁸

Während der Vorteil des Maßes Beurteilungsakkuratheit in einer exakten Bestimmung der Übereinstimmung bzw. Abweichung besteht, liegt hierin gleichzeitig eine Restriktion, nämlich die Notwendigkeit der Kenntnis des wahren Leistungswertes. Bei Leistungsbeurteilungen (z.B. in Assessment Centern) wird der Referenzwert mittels einer Expertengruppe ermittelt, die beispielsweise anhand von Videoaufzeichnungen Leistungsverhalten beurteilen. Das Dilemma liegt darin, dass die Anwendung sich weitgehend auf Laborexperimente beschränkt. Im Übrigen können auch Expertenurteile von mangelnder Akkuratheit gekennzeichnet sein.³³⁹ Das Konzept der Beurteilungsakkuratheit ist sehr stark auf Leistungsbeurteilungen zurechtgeschnitten, was die Anwendung bei Felduntersuchungen mit vielen unterschiedlichen Beurteilern und Beurteilten, wie es bei Mystery Shopping-Erhebungen der Fall ist, schwierig gestaltet. Daher wird Beurteilungsakkuratheit nicht zur Ermittlung der Ergebnisqualität von Mystery Shoppern herangezogen.

³³⁸ Vgl. Schuler (2004: 43)

³³⁹ Sulsky/Balzer (1988: 497ff.)

3.1.2 Zuverlässigkeit: Beurteilerübereinstimmung und Beurteilerreliabilität

Neben den im voranstehenden Abschnitt dargestellten Maßen ist, vor dem Hintergrund der Fragestellung nach der Ergebnisqualität unterschiedlicher soziodemografischer Gruppen, der Grad der Zuverlässigkeit von Beurteilungen ein weiterer qualitätsdeterminierender Faktor. Schließlich sind die Ergebnisse von solchen Mystery Shoppern als vorteilhafter anzusehen, die eine höhere Zuverlässigkeit aufweisen.³⁴⁰ Zur Bestimmung der Zuverlässigkeit von Beurteilungen existieren unterschiedliche Maße, die allerdings in der Literatur zum Teil nicht (systematisch) unterschieden werden. Häufig werden die Begriffe Übereinstimmung und Reliabilität synonym verwendet.³⁴¹ Nachfolgend sollen die Begriffe unterschieden werden. Dabei wird die Anwendungsbedeutung im Kontext der Bestimmung der Ergebnisqualität von Mystery Shoppern verdeutlicht.³⁴²

Bei Beurteilungen gilt, dass die *Übereinstimmung* umso höher ist, je häufiger Beurteiler zu denselben Urteilen kommen und je seltener Inkongruenzen zwischen den Beurteilern festgestellt werden. Demnach lässt sich mit Maßen der Übereinstimmung eine Aussage darüber treffen, „inwiefern verschiedene Rater verschiedene Objekte jeweils exakt gleich beurteilen.“³⁴³ Sofern jedes Objekt von allen Beurteilern den gleichen Wert zugewiesen bekommt, liegt vollkommene Übereinstimmung vor. Für Mystery Shopping-Ergebnisse ist das Maß der Übereinstimmung insofern interessant, als dass ermittelt werden kann, ob für bestimmte Gruppen unterschiedliche Ausprägungen der Übereinstimmung vorliegen. Höhere Übereinstimmung lässt auf eine einheitlichere Beurteilung schließen als niedrigere Übereinstimmung. Allerdings ist die Erhebung der Übereinstimmung nicht immer sinnvoll, da sie nicht immer aussagekräftig ist. Denn insbesondere bei Urteilen auf kontinuierlichen Skalen wird eine exakte Übereinstimmung im Vergleich zu Nominalskalen nur selten festzustellen sein.³⁴⁴

³⁴⁰ Die Zuverlässigkeit kann mit der Messgenauigkeit von Waagen veranschaulicht werden, die bei jedem Wiegevorgang des gleichen Gegenstandes auch das gleiche Gewicht angeben sollen.

³⁴¹ Vgl. hierzu und im Folgenden Wirtz/Caspar (2002: 33), die eine umfassende und bislang konkurrenzlose Einführung in das Thema „Beurteilerübereinstimmung und Beurteilerreliabilität“ liefern.

³⁴² Vgl. auch Schmidt (2007), dessen Arbeit sich insbesondere den Aspekten der Reliabilität und Validität des Dienstleistungsqualitätsurteils von Testkunden in Reisebüros widmet. Dort werden u.a. Interraterreliabilität und Beurteilerübereinstimmung empirisch überprüft. Die theoretischen Grundlagen und die praktische Umsetzung der Messung dieser Gütekriterien werden ohne vergleichbares Beispiel von Wirtz/Caspar (2002) dargestellt. Schmidt überträgt dies auf den Einsatz bei Testkunden. In der vorliegenden Arbeit werden diese (neben anderen) als Komponenten der Ergebnisqualität herangezogen. Dabei erfolgt eine sehr kompakte Darstellung der theoretischen Grundlagen mit Verweis auf Wirtz/Caspar sowie auf Schmidts Leistungen.

³⁴³ Wirtz/Caspar (2002: 34)

³⁴⁴ Vgl. hierzu das Beispiel von Wirtz/Caspar, in dem Rater die durchschnittliche Dauer von Redeanteilen von Dialogpartnern einschätzen sollen. Hier könnte nicht die exakte Einschätzung sondern der Grad der Abweichung als Maß der Zuverlässigkeit angesehen werden. Sofern Rater nur 5 Sekunden abweichen, ist ihre Übereinstimmung als zuverlässiger anzusehen als solche, die 60 Sekunden abweichen. Vgl. Wirtz/Caspar (2002: 34)

Entgegen den Maßen der Übereinstimmung ist bei Maßen der *Reliabilität* keine exakte Gleichheit der beurteilten Merkmalsausprägungen erforderlich, sondern die relative Lage der Werte zum Mittelwert der Beurteiler muss für die Merkmalsausprägungen ähnlich sein.³⁴⁵ Maße der Interraterreliabilität quantifizieren das Ausmaß, in dem jedes Objekt von den verschiedenen Ratern ähnlich weit unter bzw. über dem Durchschnitt einer untersuchten Stichprobe liegend eingeschätzt wird.³⁴⁶

Die Begriffserklärungen der Maße implizieren, dass zur Bestimmung der Interraterreliabilität mindestens Ordinalskalenniveau erforderlich ist, während Übereinstimmungsmaße für Datensätze aller Skalenniveaus ermittelt werden können, da lediglich ihre Gleichheit bzw. Ungleichheit bestimmt wird. Allerdings wird die Datenbasis dabei unabhängig von ihrem Skalenniveau als nominalskaliert betrachtet, so dass bei höherem Skalenniveau Informationen verloren gehen können.

Es erscheint vorteilhafter, allein die Reliabilität als Maß der Zuverlässigkeit von Beurteilungen heranzuziehen. Allerdings kann es sinnvoll sein, sowohl die Übereinstimmung als auch die Reliabilität zu ermitteln, um einen umfassenden Zugang zur Datenstruktur zu erhalten.³⁴⁷ Denn hohe Reliabilität impliziert nicht, dass dieselben Urteile getroffen werden, und hohe Übereinstimmung ist nicht zwangsläufig mit hoher Reliabilität verknüpft. Allein wenn ergänzend zur Reliabilität auch die Übereinstimmung als nicht zufrieden stellend eingestuft wird, kann eindeutig auf die Unzuverlässigkeit der Beurteilung geschlossen werden. Insofern erscheint es auch für Mystery Shopping-Ergebnisse sinnvoll, sowohl die Übereinstimmung als auch die Reliabilität zu ermitteln, um die Zuverlässigkeit der Daten abzubilden. Sofern eine Gruppe weniger zufrieden stellende Werte bezogen auf diese Maße aufweist, ist sie als weniger geeignet einzustufen als eine mit hohen Werten.³⁴⁸

3.1.3 Ergänzende pragmatische Überlegungen zur Bestimmung der Ergebnisqualität

Neben den in den vorstehenden Abschnitten aufgeführten Möglichkeiten, die Ergebnisqualität von Mystery Shoppern basierend auf Urteilstendenzen sowie Beurteilerübereinstimmung und Beurteilerreliabilität darzustellen, sollen in diesem Abschnitt weitere Kriterien bestimmt werden, die allerdings auf pragmatischen Überlegungen basieren.

Mit der Retest-Reliabilität wird die Korrelation von Testwerten zu verschiedenen Messzeitpunkten erhoben. Damit wird vorwiegend die zeitliche Stabilität der erhobenen

³⁴⁵ Vgl. Wirtz/Caspar (2002: 157)

³⁴⁶ Vgl. ebenda (2002: 36)

³⁴⁷ Vgl. hierzu und im Folgenden ebenda (2002: 39f.)

³⁴⁸ Wobei weiter unten noch zu zeigen ist, was zufrieden stellende Werte sind.

Werte erfasst.³⁴⁹ Interessant wäre es zu ermitteln, ob die Ergebnisse von unterschiedlichen Mystery Shoppern nicht nur über Zeit stabil sind, sondern ob die Beurteilungen auch bereits innerhalb eines Mystery Shoppings identisch sind, sofern die gleiche Merkmalsausprägung zweimal abgefragt wird. Anzunehmen ist eine identische Antwort bei gewissenhafter Arbeit der Mystery Shopper. Insofern sind solche Mystery Shopping-Ergebnisse als vorteilhafter anzusehen, die keine Abweichungen aufweisen.

Ein weiterer Aspekt ist die Vollständigkeit des Ausfüllens der Beobachtungsbögen durch die eingesetzten Mystery Shopper. Werden Fragen ausgelassen, so könnte dies ein Indiz für eine niedrige Qualität für die betreffenden Mystery Shopper sein. Schließlich werden Mystery Shopper in der Regel für ihre Arbeit entlohnt, so dass ein vollständig ausgefüllter Beobachtungsbogen ohne fehlende Werte zu erwarten ist.³⁵⁰ Im Übrigen werden die Fragen eines Mystery Shoppings in der Regel im Rahmen eines Pretests auf Verständlichkeit und Praktikabilität überprüft, so dass fehlende Werte ausschließlich auf eine nicht zufrieden stellende Arbeitsleistung des betreffenden Testers zurückzuführen sein dürften.

Die beiden hier beschriebenen Komponenten sollen mit „Sorgfaltstendenz“ bezeichnet werden. Diese kann über eine hohe Ausprägung verfügen, sofern die doppelt aufgeführten Fragen identisch beantwortet werden und der Fragebogen vollständig ausgefüllt ist oder bei Nichterfüllung über eine niedrige Ausprägung.

Als Ergebnis der vorangestellten Überlegungen lässt sich konstatieren, dass die Ergebnisqualität von Mystery Shoppern nicht an einem „ultimativen Kriterium“ fest gemacht werden kann. Vielmehr lässt sie sich, wie in anderen Bereichen auch³⁵¹, durch verschiedene als Surrogat dienende Kriterien beschreiben. Durch die Kombination der aufgeführten Indikatoren und Komponenten lassen sich unterschiedliche Facetten des Konstruktes Ergebnisqualität berücksichtigen. Hier wurde auf in der Organisations- und Personalpsychologie verwendetes und bewährtes Instrumentarium zurückgegriffen und um eigene – wenngleich sehr pragmatische – Überlegungen ergänzt. Es bleibt festzustellen, dass sich die am Ende von Kapitel 2 vermutete Unterschiedlichkeit der Ergebnisqualität der einzelnen soziodemografischen Gruppen anhand unterschiedlicher Komponenten abbilden lässt. Die Ergebnisse der Überlegungen aus diesem Abschnitt sind in der Tabelle 3-1-2 zusammengefasst. Die Erfassung bzw. Berechnung der Komponenten wird weiter unten dargestellt.

³⁴⁹ Die Retest-Reliabilität wurde bei einer Mystery Shopping-Studie von Schmidt (2007) für Reisebüros nachgewiesen.

³⁵⁰ Im Gegensatz zu Kunden- oder Mitarbeiterbefragungen, bei denen das Ausfüllen weitgehend auf freiwilliger Basis erfolgt oder lediglich ein kleiner Anreiz, z.B. in Form eines kleinen Geschenkes, gestiftet wird. Allerdings erfolgt keine Vergütung.

³⁵¹ beispielsweise im angeführten Feld der Leistungsbeurteilung

Tabelle 3-1-2: Komponenten zur Abbildung der Ergebnisqualität von Mystery Shoppern

Komponente	Wesen
<i>Mittelwertstendenz</i>	Die Urteilstendenz können bezüglich milder oder strenger Bewertung spezifiziert werden.
<i>Streuungstendenz</i>	Die Urteilstendenz können bezüglich einheitlicher oder uneinheitlicher Bewertung bezogen auf die Streuung der Ergebnisse dargestellt werden.
<i>Korrelationstendenz</i>	Der Einfluss des Urteils vorangegangener auf nachgelagerte Episoden oder von Konstrukten auf das Globalurteil kann gezeigt werden.
<i>Beurteilerübereinstimmung</i>	Die Übereinstimmung der Ergebnisse kann dargelegt werden.
<i>Beurteilerreliabilität</i>	Die Zuverlässigkeit der Ergebnisse kann bestimmt werden.
<i>Gewissenhaftigkeit</i>	Es werden identische Fragen an unterschiedlichen Stellen des Beobachtungsbogens positioniert und die Gleichheit der Antwort überprüft.
<i>Vollständigkeit</i>	Die Vollständigkeit des Ausfüllens des Fragebogens wird überprüft.

Quelle: Eigene Darstellung

3.2 Konzeption eines Testszenarios zur Gegenüberstellung der Ergebnisqualität von Mystery Shoppern

Der vorliegende Abschnitt hat das Ziel, ein Testszenario und einen korrespondierenden Beobachtungsbogen zu entwickeln, so dass die im vorangegangenen Abschnitt festgelegte Ergebnisqualität abgebildet werden kann. Dafür ist ausgehend von der Auswahl der Mystery Shopper ein geeignetes Testobjekt auszuwählen. Daneben ist ein Beobachtungsbogen zu entwickeln.

3.2.1 Auswahl der Mystery Shopper und eines geeigneten Testobjektes

Ausgehend vom Kernziel der Untersuchung, den Einfluss soziodemografischer Faktoren auf die Ergebnisqualität zu untersuchen, sind gemäß der formulierten Arbeitshypothesen solche Mystery Shopper auszuwählen, die entsprechende Ausprägungen der soziodemografischen Merkmale aufweisen. Dies ist in Tabelle 3-2-1 dargestellt.³⁵²

Tabelle 3-2-1: Festlegung der soziodemografischen Merkmale und Ausprägungen der Mystery Shopper

Hypothesen	Soziodemografische Merkmale und ihre Ausprägungen	Anzahl
H1: Männer und Frauen unterscheiden sich beim Mystery Shopping in ihrer Ergebnisqualität.	Geschlecht: Männer, Frauen	2
H2: Die Ergebnisqualität von Mystery Shoppern unterschiedlichen Alters unterscheidet sich.	Alter: 18 bis 60 Jahre (je nach Datenlage)	-
H3: Die Ergebnisqualität von Mystery Shoppers mit unterschiedlichen Bildungsabschlüssen unterscheidet sich.	Höchster Bildungsabschluss: Hauptschule, Realschule, Abitur, Studium	4
H4: Die Ergebnisqualität von Mystery Shoppern unterscheidet sich bei unterschiedlichen (beruflichen) Tätigkeiten.	Berufliche Tätigkeit: Schüler/Student, voll berufstätig, Teilzeit berufstätig, Hausfrau/Rentner	4

Quelle: Eigene Darstellung

Daneben soll in der Untersuchung der Einfluss der Erfahrung der Mystery Shopper sowie ihre Involvierung bzw. Einstellung zum Testobjekt Berücksichtigung finden. Bei der Erfahrung

³⁵² Während Geschlecht und berufliche Tätigkeit nominal skaliert werden, ist der Bildungsabschluss ordinal zu skalieren. Das Alter lässt sich hingegen auf einer Intervallskala abbilden. Daher kann keine Anzahl von auswertbaren Gruppen spezifiziert werden. Gleichwohl wird bei der Auswahl der Mystery Shopper auf eine angemessene Verteilung des Alters geachtet, so dass eine Auswertung ermöglicht wird.

wird die Anzahl der durchgeführten Testkäufe zugrunde gelegt. Bei der Involvierung werden einerseits die Besuchshäufigkeit in einer Filiale des zu testenden Unternehmens und andererseits die Besuchshäufigkeit in einer Filiale anderer branchengleicher Unternehmen herangezogen. Bei der Einstellung wird die Affinität zum getesteten Unternehmen sowie zur Branche zugrunde gelegt. Wie bei den Hypothesen, die sich auf das Alter beziehen, ist auch hier im Vorwege keine Segmentierung in Größenklassen sinnvoll. Eine etwaige Einteilung soll basierend auf der Datenlage erfolgen (vgl. Tabelle 3-2-2).

Tabelle 3-2-2: Festlegung der Erfahrung und einstellungsbezogener Merkmale der Mystery Shopper

Hypothesen	Merkmale und ihre Ausprägungen	Anzahl
H5: Die Ergebnisqualität von Mystery Shoppern mit unterschiedlicher Erfahrung hinsichtlich der Menge der Durchführung von Testkäufen unterscheidet sich.	Häufigkeit eines Testkaufs (je nach Datenlage)	-
H6: Die Ergebnisqualität von Mystery Shoppern mit unterschiedlicher Involvierung bzw. Einstellung zum Testobjekt unterscheidet sich.	Besuchshäufigkeit Einstellung	je 5

Quelle: Eigene Darstellung

Bei der *Auswahl eines geeigneten Testobjektes* sind vor allem pragmatische Überlegungen anzustellen. Ausgangspunkt ist die insgesamt anzusetzende Fallzahl von 376 Testfällen.³⁵³ Das Testobjekt muss vor dem Hintergrund der Anzahl der Testfälle und der in Kapitel 2.2.3.2 formulierten Anforderungen und Voraussetzungen verschiedene Kriterien erfüllen: Sofern 376 Testkunden in einem eng gesteckten Zeitraum dasselbe Testobjekt mit demselben Testszenario aufsuchen, besteht die Gefahr der Enttarnung und hat möglicherweise eine Verhaltensänderung des Servicepersonals zur Folge. Diese Gefahr besteht auch, wenn das Testobjekt nicht für ein sehr breites Spektrum an Kundencharakteristika geeignet ist. Folglich eignet sich nur ein Testobjekt, bei dem 376 zusätzliche „Kunden“ mit dem gleichen Anliegen nicht auffallen und bei dem ein breites Spektrum unterschiedlicher Kundencharakteristika üblich ist. Die Wahl fällt hierbei auf die Branche der Schnellrestaurants. Zum einen verfügen Ketten wie McDonald's oder Burger King über zahlreiche Filialen und hohe Gästezahlen³⁵⁴, zum anderen „richten [sie] sich grundsätzlich an keine spezifische, d.h. eng definierte Bevölkerungsgruppe. Vielmehr kommen die Kunden aus allen gesellschaftlichen

³⁵³ Vgl. hierzu Kapitel 3.3.3 weiter unten

³⁵⁴ So wurden 2006 in Deutschland in 1.276 betriebenen McDonald's-Filialen 891 Mio. Gäste bewirtet. Vgl. McDonald's (2007: 1)

Bereichen.³⁵⁵ Darüber hinaus bieten sie als Systemgastronomen in allen Filialen ein konsequent gleiches Betriebssystem.³⁵⁶ Demnach ist davon auszugehen, dass der Prozess der Dienstleistungserstellung in jeder Filiale eines Schnellrestaurants gleich abläuft.³⁵⁷ Für die Kundenkontaktsituation soll in dieser Untersuchung ein vergleichsweise einfaches, wenig komplexes und für ein Fast Food-Restaurant typisches Testszenario ausgewählt werden:³⁵⁸ Die Mystery Shopper werden gebeten, ein einfaches Produkt Ihrer Wahl, z.B. einen Burger, Pommes Frites oder ein Getränk zu bestellen.

Nachdem in diesem Abschnitt die grundsätzlichen Auswahlkriterien für Tester und Testobjekt dargestellt wurde und das Testszenario skizziert wurde, widmet sich der nächste Abschnitt der Konzeption des Beobachtungsbogens.

3.2.2 Konzeption des Beobachtungsbogens zur Ermittlung der Ergebnisqualität von Mystery Shoppern

Die Konzeption von Beobachtungsbogen und Kundenkontaktsituation sind eng miteinander verknüpft. Ziel ist es, die wahrgenommene Dienstleistungsqualität umfassend abzubilden und mit den gewonnenen Daten die Ergebnisqualität von Mystery Shoppern zu ermitteln. Für die Abbildung der Dienstleistungsqualität bieten sich standardisierte Multiattributverfahren an.³⁵⁹ Ein weit verbreitetes Verfahren ist SERVPERF von CRONIN/TAYLOR.³⁶⁰ Allerdings wird der alleinige Einsatz standardisierter Instrumente als nicht ausreichend erachtet.³⁶¹ Damit der prozessuale Charakter von Dienstleistungen Berücksichtigung findet und Informationen gewonnen werden können, die standardisierte Multiattributverfahren mit ihrer begrenzten Anzahl abstrakter Items nicht erfassen können, wird für die weitere Konzeption des Beobachtungsbogens der Einsatz des „Service Blueprinting“ von SHOSTACK³⁶² vorgeschlagen. Dabei werden als Ergänzung zum SERVPERF-Fragebogen für die einzelnen Phasen des Dienstleistungsprozesses relevante Beobachtungskriterien definiert.³⁶³ Somit

³⁵⁵ TNS Emnid (2003: 6)

³⁵⁶ Vgl. Meyer/Hoffmann (1997: 32)

³⁵⁷ Zudem sind in der Literatur unterschiedliche Beiträge zu finden, die die Dienstleistungsqualität bei Filialisten in mehreren Filialen – und nicht nur in einer – erheben, die Ergebnisse aggregieren und vergleichen. Vgl. Finn/Kayandé (1999) mit n=16 Filialen, Dawes/Sharp (2000) mit n=450 Filialen, Schmidt (2007) mit n=86 Filialen

³⁵⁸ Vgl. Voraussetzung 1 in Kapitel 2.2.3.2

³⁵⁹ Vgl. Kapitel 2.1.2.3 sowie Anforderung 2 und Voraussetzung 5 in Kapitel 2.2.3.2

³⁶⁰ Vgl. Cronin/Taylor (1992), Cronin/Taylor (1994)

³⁶¹ Vgl. hierzu und im Folgenden die Ausführungen zu Anforderung 2 und Voraussetzungen 5 und 6 in Kapitel 2.2.3.2

³⁶² Vgl. Shostack (1984)

³⁶³ Vgl. Matzler/Pechlaner/Kohl (2000: 166f.)

erfährt die eher statische Ausrichtung des standardisierten Instrumentes SERVPERF eine Dynamisierung.

Der einzusetzende Beobachtungsbogen setzt sich daher (neben soziodemografischen Kennzahlen und Einstellungsfragen zum Testobjekt) aus zwei wesentlichen Teilen zusammen: Erstens soll der von CRONIN/TAYLOR entwickelte SERVPERF-Fragebogen eingesetzt werden.³⁶⁴ Dieser ist in zahlreichen Studien für Fast Food-Restaurants validiert worden.³⁶⁵ Zudem wurde der erfolgreiche Einsatz dieses Fragebogens explizit für Mystery Shopping nachgewiesen.³⁶⁶ Zweitens wird das „Service Blueprint“ eines Fast Food-Restaurantbesuchs verwendet, an dem sich weitere Teile des Beobachtungsbogens orientieren.³⁶⁷

Nachfolgend wird zunächst der SERVPERF-Ansatz vorgestellt, bevor das Vorgehen in der empirischen Untersuchung basierend auf dem Service-Blueprinting näher erläutert wird.³⁶⁸ Dabei werden die jeweiligen Teile des einzusetzenden Fragebogens, die aus diesen Ansätzen resultieren, präsentiert.

3.2.2.1 Statische Messung der Dienstleistungsqualität mittels SERVPERF

In diesem Abschnitt soll das SERVPERF-Konzept erläutert werden, welches als standardisiertes, attributorientiertes Instrument zur Messung der Dienstleistungsqualität weit verbreitet ist.³⁶⁹ Da SERVPERF die Weiterentwicklung des SERVQUAL-Konzeptes von PARASURAMAN/ZEITHAML/BERRY ist,³⁷⁰ wird zunächst auf ihr Modell eingegangen, bevor die Vorteilhaftigkeit von SERVPERF basierend auf den Defiziten von SERVQUAL dargelegt wird. Dann werden die Komponenten des SERVPERF-Ansatzes in einem Fragebogen operationalisiert.

Ausgangspunkt für die Entwicklung von SERVQUAL ist die generelle Empfehlung CHURCHILLS zur Entwicklung von Messinstrumenten für Marketingkonstrukte.³⁷¹ Er schlägt

³⁶⁴ Sowohl in dieser als auch in der Untersuchung von Schmidt (2007) wird der SERVPERF-Fragebogen eingesetzt. Dies erleichtert eine Vergleichbarkeit der ermittelten Ergebnisse im Rahmen zukünftiger Forschungsvorhaben. Zudem wurde SERVPERF bereits in vergangenen Mystery Shopping-Untersuchungen im Einzelhandel eingesetzt (Vgl. Lowndes/Dawes (1999) und Dawes/Sharp (2000)), so dass vier Arbeiten für solche Vergleichsstudien zur Verfügung stehen.

³⁶⁵ Vgl. Cronin/Taylor (1992), Cronin/Taylor (1994), Brady/Cronin/Brand (2002), Kim/Lee/Yun (2004)

³⁶⁶ Vgl. Lowndes/Dawes (1999) und Dawes/Sharp (2000) für den Einsatz im Einzelhandel sowie Schmidt (2007) für den Einsatz in Reisebüros

³⁶⁷ Eine ähnliche kombinierte Vorgehensweise wählt auch Schmidt (2007).

³⁶⁸ Die Darstellung der theoretischen Grundlagen ist aus Gründen der Vollständigkeit erforderlich und erfolgt in dieser Arbeit in ähnlicher Weise wie bei Schmidt (2007), allerdings deutlich kompakter.

³⁶⁹ Vgl. Lowndes (2000: 724).

³⁷⁰ Vgl. Cronin/Taylor (1992), Cronin/Taylor (1994). Die Kunstworte SERVPERF und SERVQUAL setzen sich aus den Begriffen „service“ und „performance“ bzw. „quality“ zusammen.

³⁷¹ Vgl. hierzu und im Folgenden Churchill (1979: 64ff.)

vor, theoretische Marketingkonstrukte durch eine Vielzahl an Kriterien zu operationalisieren und das Messinstrument durch einen Prozess aus abwechselnder Datensammlung und Reliabilitäts- sowie Validitätsprüfung in eine einsetzbare Form zu bringen. Im Rahmen eines solchen Prozesses in den Branchen Bank-, Kreditkarten-, Broker- und Reparaturdienstleistungen ermittelten PARASURAMAN/ZEITHAML/BERRY die folgenden fünf Dimensionen der Dienstleistungsqualität.³⁷²

- *Tangibles* (Ausstattung) umfasst das gesamte physische Umfeld einer Dienstleistung und schließt insbesondere die Räumlichkeiten, deren Einrichtung sowie das Erscheinungsbild des Personals mit ein.
- *Reliability* (Verlässlichkeit) stellt die Fähigkeit des Dienstleistungsanbieters dar, die angebotene Leistung zuverlässig und akkurat ausführen zu können.
- *Responsiveness* (Einsatzbereitschaft) meint die Bereitschaft und Schnelligkeit des Dienstleistungsanbieters bei der Lösung von Kundenproblemen.
- *Assurance* (Vertrauenswürdigkeit) umfasst Wissen, Höflichkeit und Vertrauenswürdigkeit der Servicekräfte.
- *Empathy* (Einfühlungsvermögen) ist die Bereitschaft des Dienstleistungsanbieters, auf die individuellen Bedürfnisse der Kunden einzugehen.

Die fünf Dimensionen werden von den Autoren den bereits in Abschnitt 2.1.2.1 erläuterten Dimensionen einer Dienstleistung zugeordnet.³⁷³ Demnach können *Tangibles* und *Assurance* der Potenzialdimension zugeordnet werden, während *Reliability* mit der Ergebnisdimension korrespondiert und *Responsiveness* und *Empathy* sich auf die Prozessdimension beziehen.

Ergänzend zu den fünf Dimensionen präsentieren PARASURAMAN/ZEITHAML/BERRY ein Modell zur Erklärung der Entstehung des Dienstleistungsqualitätsurteils eines Kunden.³⁷⁴ Dabei wird das Urteil als Differenz („gap“) zwischen Kundenerwartung und Kundenwahrnehmung begriffen: „a form of attitude, related but not equivalent to satisfaction, that results from a comparison of expectations with perceptions of performance.“³⁷⁵ Das Modell soll anhand der Abbildung 3-2-1 verdeutlicht werden.

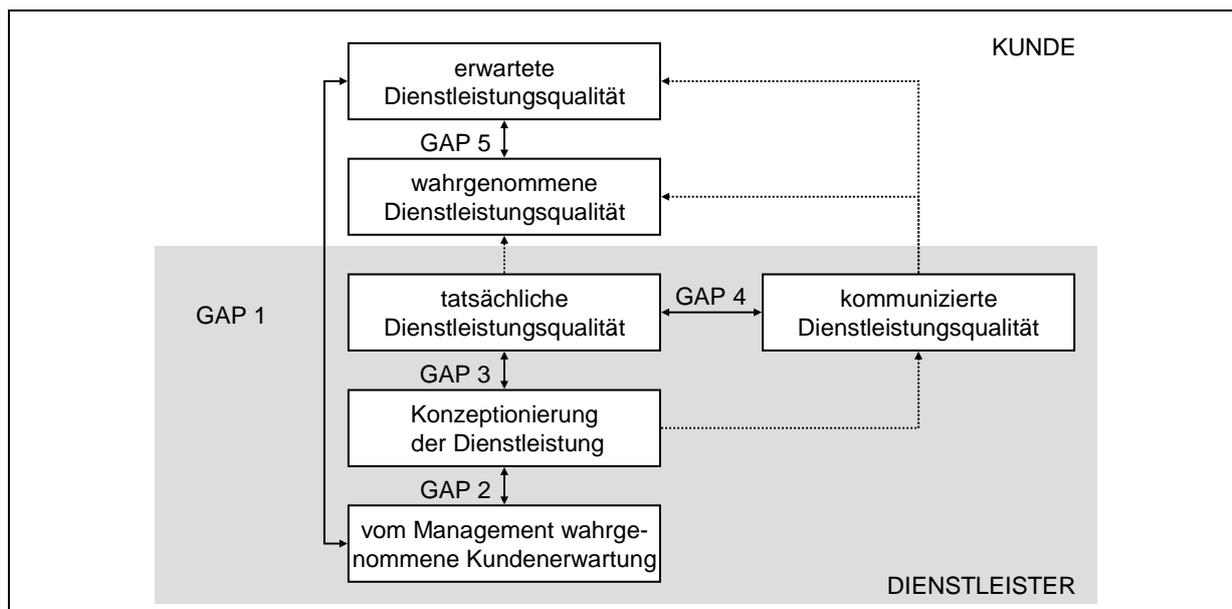
Abbildung 3-2-1: GAP-Modell der Dienstleistungsqualität

³⁷² Vgl. hierzu und im Folgenden Parasuraman/Zeithaml/Berry (1985: 41 ff.), Parasuraman/Zeithaml/Berry (1988: 12 ff.), Hentschel (1990: 230 ff.), Hentschel (2000: 307)

³⁷³ nämlich Potenzial-, Prozess- und Ergebnisdimension

³⁷⁴ Vgl. hierzu und im Folgenden Parasuraman/Zeithaml/Berry (1985: 41 ff.), Parasuraman/Zeithaml/Berry (1988: 12 ff.)

³⁷⁵ Parasuraman/Zeithaml/Berry (1988: 15)



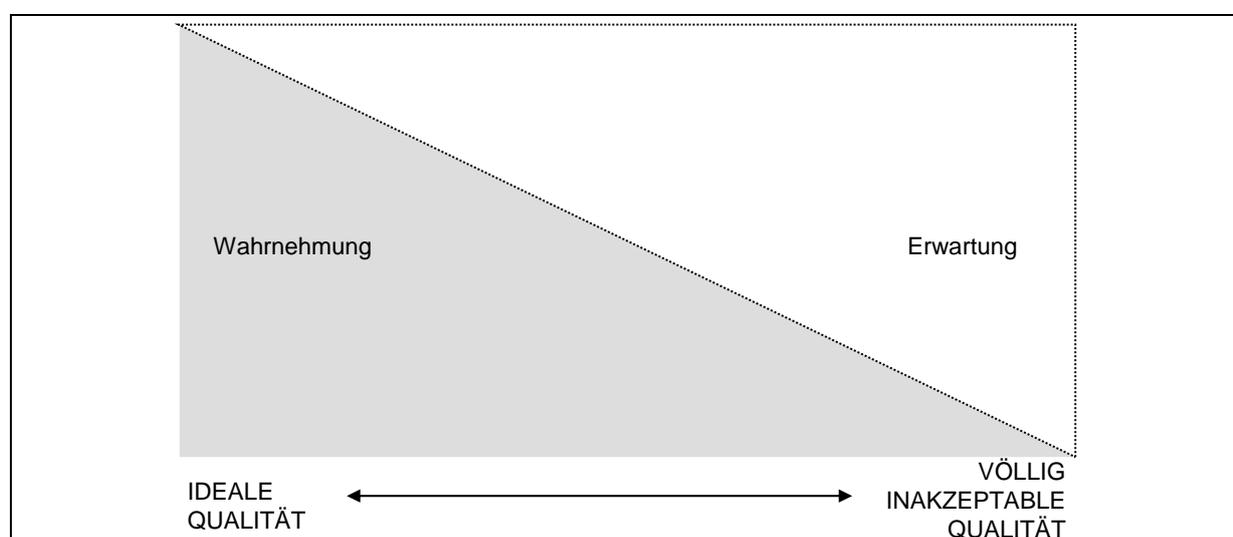
Quelle: Eigene Darstellung in Anlehnung an Parasuraman/Berry/Zeithaml (1985: 44)

Im Prozess von Erstellung bis Inanspruchnahme der Dienstleistung durch den Kunden entstehen Lücken, die sich verzerrend auf das Kundenurteil auswirken können. Einerseits kann eine Diskrepanz zwischen der vom Management wahrgenommenen Kundenerwartung und der tatsächlichen Erwartung des Kunden bestehen (Gap 1), so dass bereits vor der Konzeptionsphase falsche Annahmen seitens des Managements getroffen werden. Diese falschen Annahmen fließen sodann in die Konzeptionierung der Dienstleistung ein (Gap 2). Ferner kann es zu einer Diskrepanz zwischen gewollter und tatsächlicher Dienstleistungsqualität kommen (Gap 3). Daraus resultierend wird die gewollte Dienstleistungsqualität kommuniziert, die wiederum nicht mit der tatsächlichen Dienstleistungsqualität übereinstimmen kann (Gap 4). Letztlich kann es zu einer Diskrepanz zwischen Kundenerwartung und Kundenwahrnehmung hinsichtlich der Dienstleistungsqualität kommen (Gap 5). Es wird deutlich, dass Gap 5 die Summe der vorangegangenen Gaps darstellt. Das Ausmaß dieser zentralen Leistungslücke wird mit SERVQUAL ermittelt.³⁷⁶ Dafür wurden insgesamt 22 Items in einem standardisierten Fragebogen zusammengefasst, die die oben aufgeführten fünf Dimensionen repräsentieren. Dabei werden zu jedem Item zwei Aussagen formuliert, die einerseits die generelle Erwartung des Kunden hinsichtlich eines Aspektes abfragt („so sollte es sein“) und andererseits seinen Erfüllungsgrad erhebt („so ist es“). Soll-Standard und Ist-Leistung werden auf einer mehrstufigen Skala erfasst, so dass eine Differenzbildung erfolgen kann: Die Dienstleistungsqualität wird als umso höher angesehen, je größer der Wert der ermittelten Differenz ist.

³⁷⁶ Vgl. hierzu und im Folgenden Parasuraman/Zeithaml/Berry (1985: 41ff.), Parasuraman/Zeithaml/Berry (1988: 12ff.), Hentschel (2000: 307ff.)

Der Vorgehensweise des SERVQUAL-Konzepts zur Bestimmung der Dienstleistungsqualität wird in der Literatur mit breiter Kritik begegnet.³⁷⁷ Insbesondere die Verwendung der Doppelskala wird massiv kritisiert.³⁷⁸ Ein Widerspruch ist darin zu sehen, dass ein Kundenzufriedenheitsurteil nicht alleinig als mathematische Differenz zwischen Erwartung und Wahrnehmung angesehen werden kann. Das Vergleichsergebnis zwischen Soll-Standard und Ist-Leistung kann auf einem Kontinuum zwischen den Extremen „ideale Qualität“ und „völlig inakzeptable Qualität“ liegen. Übertrifft die wahrgenommene Qualität die erwartete, so tendiert die wahrgenommene Qualität in Richtung „idealer Qualität“, im umgekehrten Falle in Richtung „völlig inakzeptabler Qualität.“ Der Zusammenhang ist in Abbildung 3-2-2 veranschaulicht.

Abbildung 3-2-2: Ergebnisvarianten des SERVQUAL-Vergleichsprozesses



Quelle: Eigene Darstellung

PARASURAMAN/ZEITHAML/BERRY unterstellen, dass der Erwartungswert einer fiktiven Idealleistung entspricht. Vor dem Hintergrund des zuvor dargestellten Zusammenhanges führt dies jedoch dazu, dass das Dienstleistungsangebot besser als ideal sein muss, um als qualitativ befriedigend bewertet werden zu können. Denn in ihrem Konzept ist es nicht möglich, dass die erlebte Leistung die erwartete übertrifft. Dabei kann es aufgrund der Doppelskala wegen mangelnder Plausibilität zu fragwürdigen Interpretationen kommen: So ist es möglich, dass ein Kunde einen hohen Erwartungswert hinsichtlich der Freundlichkeit des Personals hat (Erwartungswert = 5) und dieser auch erfüllt wird (Wahrnehmungswert = 5). Einem anderen

³⁷⁷ Vgl. hierzu und im Folgenden Hentschel (1990), Cronin/Taylor (1992), Hentschel (1992: 399ff.), Cronin/Taylor (1994), Hentschel (2000: 309ff.), Corsten (2001: 313), Brady/Cronin/Brand (2002) sowie Dyke/Prybutok/Kappelman (1999: 878ff.) und die dort zitierte Literatur für einen umfangreichen Überblick der identifizierten Einzelprobleme

³⁷⁸ Vgl. hierzu auch die Diskussion des Einsatzes von Ein- vs. Zweikomponentenansätzen in Abschnitt 2.1.2.3 dieser Arbeit, in der u.a. ausgeführt wird, dass auf eine doppelte Erhebung von Soll- und Ist-Leistung zu verzichten ist, da sich das Urteil nicht aus einer mathematische Differenz, sondern aus einer vom Kunden empfundenen Diskrepanz ergibt.

Kunden ist die Freundlichkeit weniger wichtig (Erwartungswert = 1), nimmt die ihm entgegengebrachte Freundlichkeit allerdings als sehr positiv wahr (Wahrnehmungswert = 5). Da sich beim ersten Kunden eine Differenz von 0 und beim zweiten eine Differenz von +4 ergibt, bewertet der zweite Kunde gemäß der SERVQUAL-Logik das Kriterium Freundlichkeit mit einer höheren Dienstleistungsqualität, was unplausibel ist. Dieser Effekt könnte durch das Phänomen von Deckeneffekten, also einer Bewertung nahe dem Skalenmaximum, oder durch Anspruchsinflation bei der Einschätzung des Erwartungswertes („alles ist wichtig“) verstärkt werden. In diesem Zusammenhang ist auch die Möglichkeit der Beeinflussung durch Ausstrahlungseffekte (Halo-Effekte) zu nennen, wobei beispielsweise bestimmte als positiv bewertete Eigenschaften auf die Bewertung anderer Eigenschaften ausstrahlen.

Als Folge dieser breiten Kritik an der verwendeten Doppelskala haben CRONIN/TAYLOR als SERVQUAL-Hauptkritiker in ihrem SERVPERF-Ansatz die Erhebung des Erwartungswertes eliminiert und erheben einzig die wahrgenommene Dienstleistungsqualität.³⁷⁹ Dabei greifen sie auf die Skala des Wahrnehmungswertes der SERVQUAL-Skala zurück. Neben der Vorteilhaftigkeit der Verwendung Einfachskala gegenüber der Doppelskala, bietet SERVPERF weitere Vorteile gegenüber SERVQUAL, die in zahlreichen empirischen Studien nachgewiesen worden sind.³⁸⁰ Demnach ist SERVPERF überlegen, da mit diesem Ansatz ein größerer Varianzanteil des Dienstleistungsqualitätsgesamturteils aufgeklärt werden kann als mit SERVQUAL. Ferner konnte die ursprüngliche Faktorstruktur überwiegend reproduziert werden, was bei entsprechenden Replikationsstudien für SERVQUAL nicht gelang. Somit sind mit SERVPERF handlungsrelevantere Ergebnisse zu generieren als mit SERVQUAL.

Vor dem Hintergrund der massiven Kritik an SERVQUAL soll in der vorliegenden Arbeit der leistungsfähigere SERVPERF-Ansatz eingesetzt werden, der sich im Übrigen – wie bereits gezeigt wurde – bei vergangenen Mystery Shopping-Untersuchungen bewährt hat. In Tabelle 3-2-3 werden die Komponenten Tangibles, Reliability, Responsiveness, Assurance und Empathy mit den entsprechenden Items operationalisiert.³⁸¹

³⁷⁹ Vgl. Cronin/Taylor (1992), Cronin/Taylor (1994)

³⁸⁰ Vgl. hierzu und im Folgenden Cronin/Taylor (1992), Cronin/Taylor (1994) sowie den Überblick entsprechender empirischer SERVPERF-Studien bei Schmidt (2007: 122f.) und den Überblick entsprechender SERVQUAL-Studien bei Dyke/Kappelman/Prybutok (1997: 202)

³⁸¹ Anzumerken ist, dass in dieser Arbeit eine eigene Überarbeitung von Hentschels (2000: 308) deutschsprachiger Übersetzung eingesetzt wird, die sich sprachlich genauer am Originaltext orientiert. Ferner werden die Items in dieser Untersuchung – entgegen der teilweise negativ formulierten Form in der Literatur (z.B. bei Cronin/Taylor (1992: 65f.) – durchgängig positiv formuliert. Wie Lowndes für eine Mystery Shopping-Untersuchung unter Einsatz des SERVPERF-Kataloges zeigt, existiert allerdings kein „negative statement effect“, so dass die hier gewählte Vorgehensweise als praktikabel anzusehen ist. Vgl. Lowndes (2000: 724ff.)

Tabelle 3-2-3: Operationalisierung der SERVPERF-Dimensionen am Beispiel von McDonald's

Items	Bewertungsskala
I. Tangibles (Ausstattung)	
1. Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	
2. Die Einrichtung ist ansprechend gestaltet.	
3. Die Angestellten machen einen ordentlichen Eindruck.	
4. Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	
II. Reliability (Verlässlichkeit)	
5. Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	
6. Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	
7. McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	
8. Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	
9. McDonald's liefert fehlerfreie Belege.	
III. Responsiveness (Einsatzbereitschaft)	
10. Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	Stimme voll zu
11. Der Gast erhält von den Angestellten einen prompten Service.	–
12. Die Mitarbeiter sind immer willens zu helfen.	Stimme überhaupt nicht zu
13. Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	
IV. Assurance (Vertrauenswürdigkeit)	
14. Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	
15. Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	
16. Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	
17. Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	
V. Empathy (Einfühlungsvermögen)	
18. McDonald's gibt Ihnen individuelle Aufmerksamkeit.	
19. McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	
20. McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	
21. McDonald's liegen Ihre Interessen am Herzen.	
22. Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	

Quelle: In Anlehnung an Cronin/Taylor (1992: 65f.), eigene Übersetzung

3.2.2.2 Dynamisierung der Messung der Dienstleistungsqualität mit der Kontaktpunkt-Analyse

Nach Erläuterung des SERVPERF-Ansatzes und Präsentation eines Teils des Fragebogens wird die statische Ausrichtung des vorangegangenen Abschnitts nachfolgend um eine dynamische Komponente erweitert, indem der prozessuale Charakter einer Dienstleistung Berücksichtigung findet. Wie oben bereits begründet wurde, wird die ausschließliche Verwendung standardisierter, merkmalsorientierter Konzepte zur Beschreibung des Qualitätsgesamturteils nicht als ausreichend erachtet, sondern bedarf einer Ergänzung um ereignisorientierte Elemente, um dem Prozesscharakter einer Dienstleistung zu berücksichtigen.³⁸²

Bei Dienstleistungen ist der Kunde am Erstellungsprozess beteiligt, was zur Folge hat, dass die Leistung nicht vollständig standardisierbar ist. Neben Art und Umfang der Kundenbeteiligung ist der Prozess von situativen Umständen sowie von der Person des Dienstleistenden abhängig. Zudem erfolgt die Qualitätswahrnehmung aufgrund der Intangibilität von Dienstleistungen in einem vergleichsweise hohen Maße während des Prozesses der Dienstleistungserstellung.³⁸³ Daher wird die Kundenkontaktsituation in der Literatur häufig mit der Metapher „moment of truth“ bezeichnet.³⁸⁴ Diese Augenblicke der Wahrheit können u.a. in personalbezogene und nicht personalbezogene Kontaktpunkte aufgeteilt werden.³⁸⁵ Während als personalbezogen bezeichnete Kontaktpunkte alle Interaktionen zwischen Dienstleistungsunternehmen und seinen Kunden (z.B. der Bestellvorgang in einem Restaurant) einschließt, sind nicht personalbezogene Kontaktpunkte solche, die tangible Elemente des Umfeldes (z.B. die Wahrnehmung oder Beurteilung der Geschäftsausstattung) umfassen. Der Dienstleistungsprozess ist demnach eine Abfolge solcher Kontaktpunkte. Für die dynamische Analyse eines Dienstleistungsprozesses ist es daher sinnvoll, ihn in unterschiedliche Teilprozesse zu zerlegen. Im Übrigen resultiert das Urteil der Dienstleistungsqualität durch die prozessuale Wahrnehmung und kognitive Verarbeitung aus episodischen Einzelurteilen und nicht aus zeitunabhängigen und übergreifenden Dimensionen.³⁸⁶ Mit dem von SHOSTACK vorgeschlagenen „Service Blueprinting“ bietet sich eine geeignete Methode zur Strukturierung und Visualisierung eines Prozesses der Dienstleistungsnutzung.³⁸⁷ Mit Hilfe einer grafischen Darstellung werden die Teilprozesse einer Dienstleistungssituation in ihrer Abfolge in einem Ablaufdiagramm

³⁸² Vgl. hierzu die Ausführungen zu Anforderung 2 und Voraussetzungen 5 und 6 in Kapitel 2.2.3.2

³⁸³ Vgl. Stauss (2000: 323f.)

³⁸⁴ Vgl. für viele Carlzon (1987)

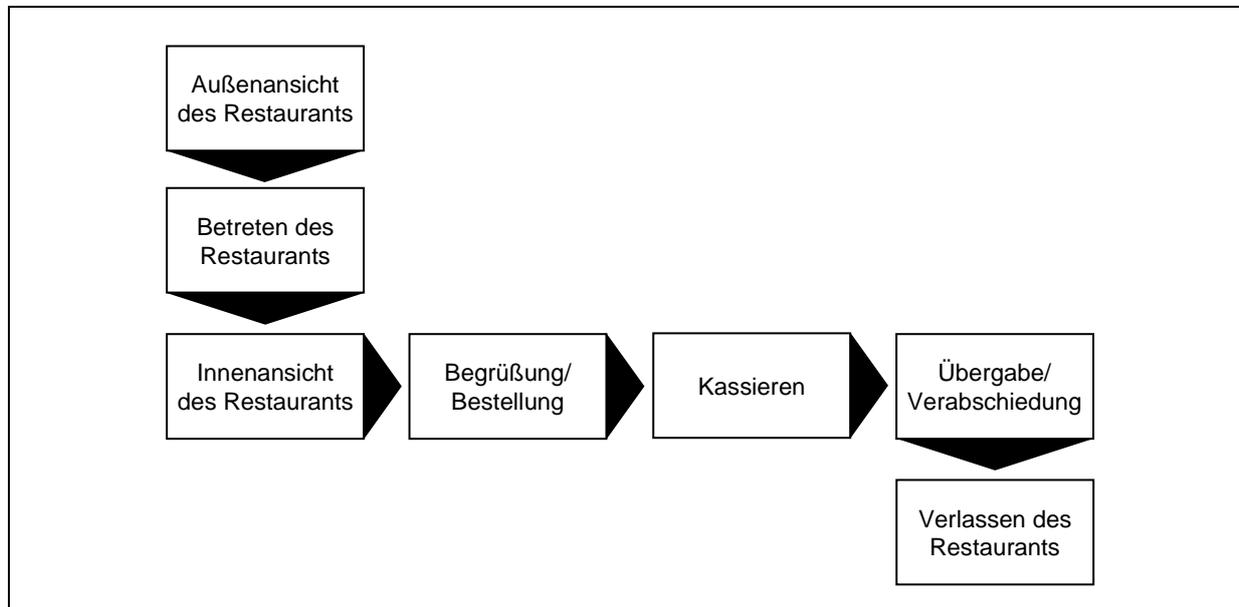
³⁸⁵ Vgl. hierzu und im Folgenden Stauss (2000: 324f.)

³⁸⁶ für eine ausführliche Darstellung, Vertiefung und Fundierung vgl. Siefke (1998: 74f.)

³⁸⁷ Vgl. Shostack (1984: 133ff.)

abgebildet. Sofern zwischen für den Kunden sichtbaren und unsichtbaren Prozessen unterschieden wird, kann der sichtbare Teil in einer so genannten „Line of Visibility“ deutlich gemacht werden. In Abbildung 3-2-3 ist der sichtbare Teil eines Service Blueprints eines Fast Food-Restaurantbesuchs dargestellt.

Abbildung 3-2-3: Service Blueprint der Line of Visibility eines Fast Food-Restaurantbesuchs



Quelle: Eigene Darstellung in Anlehnung an Olavarria (1999: 178) und Stauss (2000: 328)

Basierend auf dem Service Blueprint aus Abbildung 3-2-3 lassen sich beim Besuch eines Fast Food-Restaurants unterschiedliche Kontaktpunkte unterscheiden, die für die empirische Analyse zu operationalisieren sind. Während „Außenansicht“ und „Innenansicht des Restaurants“ einzeln abgefragt werden, werden die Items der Kontaktpunkte „Begrüßen/Bestellen“, „Kassieren“ sowie „Übergabe/Verabschiedung“ unter dem übergeordneten Punkt „Kundengespräch“ zusammengefasst. Daneben werden Items zu Mitarbeitereindrücken abgefragt, die unter personalbezogene Kontaktpunkte fallen, aber nicht explizit in die Kontaktpunktanalyse basierend auf dem Service Blueprint einzuordnen sind. In den wissenschaftlichen Beiträgen zur Bedeutungsgewichtung einzelner Items hat sich bisher kein Ansatz zur Aggregation von Qualitätseinzelnmerkmalen durchgesetzt. Daher werden in dieser Untersuchung die Qualitätsurteile der einzelnen Kontaktpunkte sowie das Gesamtqualitätsurteil separat erhoben.³⁸⁸ Die Items sind in Tabelle 3-2-4 dargestellt.

³⁸⁸ Vgl. Lingenfelder/Wieseke/Schmidt (2003) für das gleiche Vorgehen einer separaten Erhebung der Qualitätsurteile

Tabelle 3-2-4: Operationalisierung der Kontaktpunkte am Beispiel von McDonald's

Items	Bewertungsskalen
I. Außenansicht	
1. Wie ist der Sauberkeitszustand der näheren Umgebung des Restauranteingangs (Bürgersteig) und gegebenenfalls von der Terrasse?	
2. Wie ist der Zustand der Fenster und Fassaden?	
3. Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	
II. Zustand der Innenräume	
4. Wie ist der Sauberkeitszustand des Bodens?	
5. Wie ist der Zustand des Sitzbereichs?	
6. Wie ist der Zustand der Servicestation?	
7. Wie ist der Zustand des Thekenbereichs?	
8. Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	Sauber – extrem verschmutzt
III. Kundengespräch (Begrüßen/Bestellen, Kassieren, Übergabe/Verabschiedung)	
9. Beim Herantreten an die Theke werde ich freundlich begrüßt.	
10. Ich werde freundlich nach meinem Bestellwunsch gefragt.	1 – 5
11. Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	
12. Der Mitarbeiter nennt mir in höflicher Weise den Preis.	
13. Der Bezahlvorgang verläuft reibungslos.	Stimme voll zu – Stimme überhaupt nicht zu
14. Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	
15. Die Mitarbeiter sind durchweg höflich zu mir.	
16. Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	
IV. Mitarbeiter	
17. Die Mitarbeiter wirken ordentlich.	
18. Die Mitarbeiter wirken hektisch und überlastet.	
19. Die Mitarbeiter machen einen freundlichen Eindruck.	
20. Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	
21. Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	
V. Gesamteindruck	
22. Welchen Eindruck macht der Besuch insgesamt auf Sie?	

Quelle: Eigene Darstellung

3.3 Vorgehensweise bei der empirischen Untersuchung

Nachdem im ersten Abschnitt dieses Kapitels die Komponenten der Ergebnisqualität von Mystery Shoppern erörtert wurden, ist im zweiten Abschnitt ein TestszENARIO konzipiert worden. Dieser Abschnitt widmet sich der Darstellung der Vorgehensweise bei der empirischen Untersuchung. Dafür wird insbesondere die Erfassung der Komponenten der Ergebnisqualität dargelegt, es wird auf ihre Bewertung eingegangen und schließlich wird der Ablauf der Erhebung geschildert.

3.3.1 Erfassung der Komponenten der Ergebnisqualität

In der Untersuchung werden verschiedene Komponenten der Ergebnisqualität von Mystery Shoppern erfasst. Konkret sind dies die zu Beginn dieses Kapitels aufgeführten Urteilstendenzen, dargestellt durch Mittelwerts-, Streuungs- und Korrelationstendenz, die Zuverlässigkeit, dargestellt durch Beurteilerübereinstimmung und Beurteilerreliabilität sowie die Sorgfaltstendenz, dargestellt durch die Gewissenhaftigkeit und Vollständigkeit des Ausfüllens. Ihre Erfassung bzw. Berechnung wird in den nachfolgenden Abschnitten kurz vorgestellt.

3.3.1.1 Erfassung der Urteilstendenzen

Eine Darstellung der Erfassung der Urteilstendenzen *Mittelwertstendenz* und *Streuungstendenz* erfolgt hinreichend an vielen Stellen in der Literatur³⁸⁹ und wird daher nicht näher erläutert.³⁹⁰ Die *Korrelationstendenz* nimmt in der Literatur nur wenig Raum ein und soll deshalb kurz erläutert werden. Wie weiter oben dargelegt wurde, soll zur Abbildung der Korrelationstendenz für die einzelnen soziodemografischen Gruppen ermittelt werden, welchen Effekt das Dienstleistungsqualitätsurteil vorangehender Phasen auf dasjenige nachgelagerter Phasen ausübt bzw. welchen Effekt die Konstrukte auf das Gesamtzufriedenheitsurteil haben. Die Korrelationstendenz ist komplexer in ihrer Erfassung und Darstellung als Mittelwerts- und Streuungstendenz, da nicht vergleichsweise einfach darzustellende Mittelwerts- oder Streuungsprofile untersucht werden. Zur Ermittlung des beschriebenen Effekts eignet sich die Kausalanalyse.³⁹¹ Mit Kausalmodellen können Abhängigkeiten zwischen latenten Variablen, die sich einer direkten Messung verschließen,

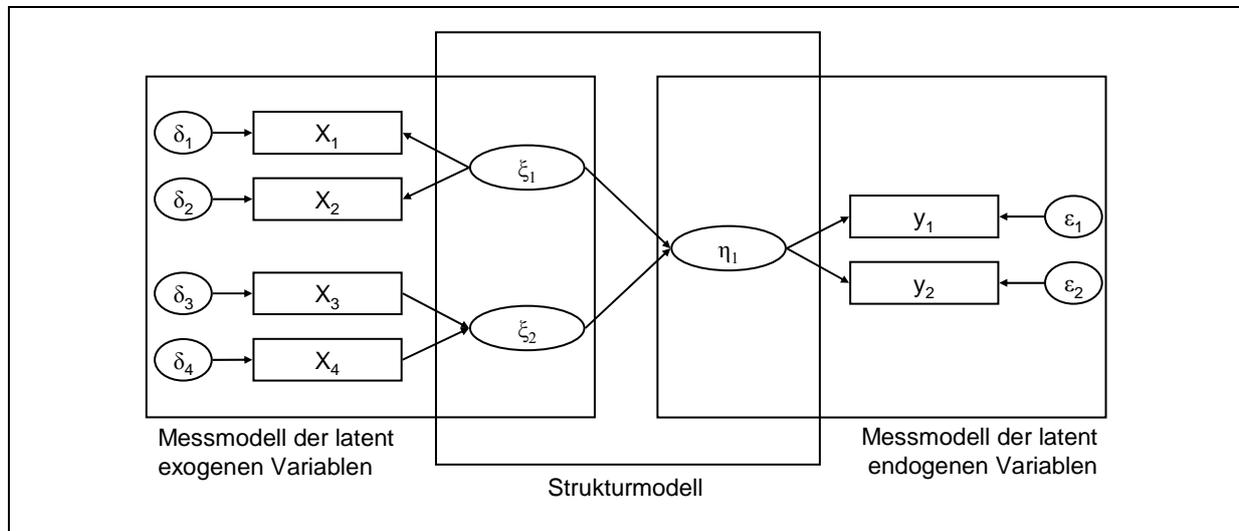
³⁸⁹ siehe stellvertretend für viele Reichardt/Reichardt (2000) und Zöfel (2003)

³⁹⁰ Der Vollständigkeit halber ist die Erfassung im Anhang 1 zu finden.

³⁹¹ Die Kausalanalyse wird auch als Strukturgleichungsanalyse bezeichnet, vgl. z.B. Homburg/Pflesser (2000: 633ff.), Backhaus/Erichson/Plinke/Weiber (2006: 337ff.). Für anwendungsorientierte Beispiele der Kausalanalyse zur Abbildung des Effektes von vorgelagerten auf nachgelagerte Dienstleistungsphasen vgl. z.B. Siefke (1998) und zur Abbildung des Effektes von unterschiedlichen Konstrukten aufeinander bei der Erstellung von (komplexen) Dienstleistungen vgl. z.B. Güthoff (1995)

abgebildet werden, indem sie über Indikatoren ihren empirischen Bezug erhalten.³⁹² Abbildung 3-3-1 soll ein Strukturgleichungsmodell veranschaulichen.

Abbildung 3-3-1: Pfaddiagramm eines Strukturgleichungsmodells



Quelle: Eigene Darstellung

Reflektive Indikatoren (z.B. x_1 und x_2 bei ξ_1 in Abbildung 3-3-1) können die Ausprägung einer latenten Variablen widerspiegeln, während formative Indikatoren auf ihre Ausprägung einwirken (ξ_2).³⁹³ Strukturgleichungsmodelle können mit kovarianzbasierten und varianzbasierten Verfahren geschätzt und modelliert werden.³⁹⁴ Zu den varianzbasierten Verfahren zählt der Ansatz der Partial Least Squares (PLS)-Pfadmodellierung. Während seine Anwendung zu Gunsten kovarianzbasierter Strukturgleichungsmodelle in der Vergangenheit eher eine untergeordnete Rolle spielte, sind Publikationen in jüngerer Zeit zahlreich.³⁹⁵ Dabei wird der Einsatz insbesondere aufgrund der Vorteile gegenüber den kovarianzbasierten Modellen propagiert, die aus unterschiedlichen Merkmalen der Verfahren resultieren. Ausgewählte im Schrifttum angeführte Unterschiede der Ansätze sind in Tabelle 3-3-1 zusammengefasst.

³⁹² Vgl. für viele Herrmann/Huber/Kressmann (2006: 35); Aufgrund der häufigen Verwendung der Kausalanalyse in der empirischen Forschung kann auf eine detaillierte Darstellung des Verfahrens verzichtet werden, da dies bereits an anderer Stelle ausführlich erfolgt ist, vgl. z.B. Backhaus/Erichson/Plinke/Weiber (2006: 337ff.). Das Verfahren soll hier daher lediglich skizziert werden.

³⁹³ Vgl. z.B. Herrmann/Huber/Kressmann (2006: 35)

³⁹⁴ Vgl. z.B. Ringle (2004a: 5)

³⁹⁵ Vgl. Beiträge in „Zfbf“, z.B. Albers/Hildebrandt (2006), Fassott (2006), Herrmann/Huber/Kressmann (2006), in „Die Betriebswirtschaft“, z.B. Götz/Liehr-Gobbers (2004), Eberl (2006), Huber/Heitmann/Herrmann (2006) sowie weitere Beiträge in derselben Ausgabe zu Strukturgleichungsmodellen und Beiträge mit dem Schwerpunkt der Anwendungsorientierung in „Das Wirtschaftsstudium“, z.B. Ringle/Boysen/Wende/Will (2006), Ringle/Spreen (2007).

Tabelle 3-3-1: PLS vs. kovarianzbasierte Strukturgleichungsmodelle

Merkmal	PLS	Kovarianzbasiertes Modell
<i>Stichprobengröße</i>	Kleine Stichprobengröße möglich, abhängig von größter Prädiktorzahl	Empfehlungen für Mindestgrößen liegen zwischen 200 und 800 Fällen
<i>Verteilung der Stichprobe</i>	Keine Vorgaben	Normalverteilung
<i>Schwerpunkt der Zielsetzung</i>	Vorhersage der Datenmatrix	Replikation der Kovarianzstruktur der Ausgangsdatenmatrix
<i>Unsinnige Werte</i>	Nicht möglich	Möglich
<i>Modellierung formativer Indikatoren</i>	Problemlos modellierbar	Eingeschränkt modellierbar
<i>Modellkomplexität</i>	Hoch (z.B. 100 Konstrukte und 1000 Indikatoren)	Niedrig (weniger als 100 Indikatoren)

Quelle: In Anlehnung an Chin/Newsted (1999: 314)

Für den Zweck der vorliegenden Untersuchung erscheint der PLS-Ansatz insbesondere aufgrund der wenigen Restriktionen und geringeren Anforderungen an die Daten (vergleichsweise geringe Stichprobengröße, Verteilung) als vorteilhafter. Zudem lassen sich formative Indikatoren problemlos modellieren. Komplexitätsreduzierend wirkt sich dabei aus, dass Indikatorvariablen in formativen Messmodellen im Gegensatz zu reflektiven Messmodellen keinen Residualwert haben.³⁹⁶ Die Schätzung der Modellparameter soll mit der Software SmartPLS 2.0 erfolgen.³⁹⁷ Die Gütebeurteilung für das PLS-Verfahren fällt hinsichtlich der Zahl möglicher Gütemaße im Vergleich zur Kovarianzstrukturanalyse wesentlich geringer aus.³⁹⁸ Bisher existiert kein allgemein anerkanntes globales Gütemaß. Die Messmodelle der latenten Variablen und des Strukturmodells müssen isoliert betrachtet werden. Einen Überblick der Gütemaße gibt Tabelle 3-3-2.

³⁹⁶ Diamantopoulos/Winklhofer (2001: 271)

³⁹⁷ Vgl. Ringle/Wende/Will (2005) für die leicht zu bedienende statistische Software mit grafischer Modellierungsoberfläche. SmartPLS 2.0 hat sich im Vergleich mit anderen Software-Produkten als leistungsfähig erwiesen, vgl. z.B. Temme/Kreis/Hildebrandt (2006)

³⁹⁸ Vgl. hierzu und im Folgenden Ringle (2004b)

Tabelle 3-3-2: Gütemaße zur Beurteilung von PLS-Modellen

Bezeichnung	Wesen		Anforderung
<i>Strukturmodell</i>			
<i>Bestimmtheitsmaß</i> R^2	Anteil der erklärten Varianz der endogenen Variablen	(K-1) $R^2 = \frac{\sum_{i \in I} (\hat{y}_i - \bar{y})^2}{\sum_{i \in I} (y_i - \bar{y})^2}$	0,67 substanzuell 0,33 durchschnittlich 0,19 schwach
<i>Effektstärke</i> f^2	Einfluss der exogenen latenten Variablen auf die endogene latente Variable	(K-2) $f^2 = \frac{R^2_{included} - R^2_{excluded}}{1 - R^2_{included}}$	0,02 geringer Einfluss 0,15 mittlerer Einfluss 0,35 großer Einfluss
<i>Stone-Geisser-Test-Kriterium</i> Q^2	Vorhersagegüte tatsächlich beobachteter Werte	(K-3) $Q^2 = 1 - \frac{\sum_D E_D}{\sum_D O_D}$	> 0
<i>Reflektive Messmodelle</i>			
<i>Indikatorreliabilität</i>	Anteil der durch die latente Variable erklärten Varianz in Relation zum Messfehler	$DEV = \frac{\sum_i \lambda_i^2}{\sum_i \lambda_i^2 + \sum_i \text{var}(\varepsilon_i)}$ Durchschn. Erfasste Varianz	> 0,5, d.h. Faktorladungen > 0,7
<i>Diskriminanzvalidität</i>	Ausmaß, mit dem die latenten Variablen tatsächlich eigenständige Konstrukte darstellen		DEV > quadrierte Korrelationen aller latenten Variablen
<i>Konstruktreliabilität</i>	Messgüte eines Konstruktes durch seine Items	(K-4) $p_c = \frac{(\sum_i \lambda_i)^2}{(\sum_i \lambda_i)^2 + \sum_i \text{var}(\varepsilon_i)}$	> 0,7
<i>Formative Messmodelle</i>			
<i>Signifikanz der Gewichte</i>	Mit dem Bootstrapping-Verfahren wird ein t-Test für die Signifikanz der Gewichte zwischen manifester und latenter Variable durchgeführt.		Zweiseitig > 1,98

Quelle: In Anlehnung an Ringle/Boysen/Wende/Will (2006: 87) und die dort zitierte Literatur

Zum Vergleich der verschiedenen Gruppen von Mystery Shoppern sollen die Pfadkoeffizienten zwischen den latenten Konstrukten verglichen werden. Bedeutend ist, ob sich die errechneten Koeffizienten signifikant voneinander unterscheiden oder ob keine Unterschiede ausgemacht werden können. Hierfür lassen sich vier Fälle unterscheiden:³⁹⁹ Sofern sich die Wirkungsbeziehung eines Pfades im Strukturmodell in beiden Gruppen als nicht signifikant erweist, kann auch kein Unterschied zwischen den Gruppen festgestellt werden. Ist die Wirkungsbeziehung in einer Gruppe signifikant und in der anderen nicht, so

³⁹⁹ Vgl. hierzu und im Folgenden Huber/Herrmann/Meyer/Vogel/Vollhardt (2007: 122ff.)

unterscheiden sie sich signifikant voneinander. Tritt in beiden Gruppen eine signifikante Wirkungsbeziehung auf und sind beide Pfadkoeffizienten gleich groß, ist von der Gleichheit der Strukturparameter auszugehen. Der vierte mögliche Fall tritt dann auf, wenn die Pfadkoeffizienten unterschiedlich sind und jeweils eine signifikante Wirkungsbeziehung in den Gruppen festgestellt wird. Dieser Fall lässt allerdings keine einfache Augenscheinahme des Unterschieds zu, sondern erfordert eine Bestimmung der dazu benötigten t-Werte für die Differenz zwischen den Schätzern. Der Wert folgt einer t-Verteilung mit $m+n-2$ Freiheitsgraden und wird wie folgt ermittelt:⁴⁰⁰

$$(K-5) \quad t = \frac{p_x^1 - p_x^2}{\sqrt{\frac{(m-1)^2}{(m+n-2)} \cdot (\sigma(p_x^1))^2 + \frac{(n-1)^2}{(m+n-2)} \cdot (\sigma(p_x^2))^2 \cdot \left(\sqrt{\frac{1}{m} + \frac{1}{n}}\right)}}$$

Dabei sind m und n die Größen der zu vergleichenden Stichproben, p der Schätzer des Originalsamples bezüglich der zu untersuchenden Modellassoziation in der jeweiligen Stichprobe und σ der jeweilige Standardfehler des für das Modell generierten Bootstrap-Samples.

3.3.1.2 Erfassung der Zuverlässigkeit

Nachdem im vorangegangenen Abschnitt dargestellt wurde, wie die Korrelationstendenz erfasst werden soll, wird in diesem Abschnitt gezeigt, wie die Zuverlässigkeit unter Heranziehung von Beurteilerübereinstimmung und Beurteilerreliabilität erfasst werden soll.

Die *Beurteilerübereinstimmung* kann mit Hilfe der prozentualen Übereinstimmung erfasst werden. Sie ist das einfachste Maß der Übereinstimmung. Zur Berechnung wird die Anzahl der von allen Ratern gleich beobachteten Merkmale durch die Gesamtanzahl der Merkmale geteilt.⁴⁰¹ Allerdings setzt dies voraus, dass dieselben Merkmale von allen Ratern beobachtet werden. Im Rahmen dieser Untersuchung ist dies allerdings nur schwer realisierbar. Damit die Übereinstimmung dennoch beobachtet werden kann, besteht die Möglichkeit, Paare von Mystery Shoppern gemeinsam ihre Beobachtungen durchführen zu lassen, um im Anschluss getrennt voneinander die Bewertungen durchzuführen.⁴⁰² Die prozentuale Übereinstimmung über alle Raterpaare kann ermittelt werden, indem die Häufigkeit identischer Urteile h_{ij} durch

⁴⁰⁰ Vgl. und Huber/Herrmann/Meyer/Vogel/Vollhardt (2007: 118f.), deren Formel auf Chin (2004) basiert.

⁴⁰¹ Vgl. Wirtz/Caspar (2002: 47)

⁴⁰² Vom paarweisen Einsatz von Mystery Shoppern berichten bereits Becker (1973: 108), Sutton/Rafaeli (1988: 467), Rafaeli/Sutton (1990: 628), Deckers (1999: 196ff.), Dawes/Sharp (2000: 36) sowie Schmidt (2007: 139).

die Anzahl der Raterpaare multipliziert mit der Anzahl der zu beurteilenden Objekte N geteilt wird:⁴⁰³

$$(Z-1) \quad P\ddot{U} = \frac{\sum_{i=1}^s h_{ij}}{\frac{R \cdot (R-1)}{2} \cdot N} \cdot 100\%$$

Wenngleich die prozentuale Übereinstimmung eine sehr anschauliche Interpretation ermöglicht, ist sie nur beschränkt aussagekräftig, da die durch den Zufall zu erwartende Übereinstimmung keine Berücksichtigung findet. Cohens κ hingegen berücksichtigt das Verhältnis des Anteils an beobachteten Übereinstimmungen P_o und den rein zufällig zu erwartenden Übereinstimmungen P_e .⁴⁰⁴

$$(Z-2) \quad \kappa = \frac{P_o - P_e}{1 - P_e}$$

Der Wertebereich für κ liegt zwischen -1,0 (vollständig unterschiedliche Übereinstimmung) und 1,0 (perfekte Übereinstimmung), wobei Werte zwischen 0,4 und 0,6 als akzeptabel, zwischen 0,6 und 0,75 als gut und ein $\kappa > 0,75$ als Indikator für sehr gute Übereinstimmung angesehen wird. Die Signifikanz des Wertes wird unter Verwendung der t-Statistik ermittelt, die bei kleinen Stichproben zu exakteren Ergebnissen führt als die Verwendung des Chi-Quadrat-Tests.⁴⁰⁵

Zur Bestimmung der *Beurteilerreliabilität* existieren unterschiedliche Maße.⁴⁰⁶ Für die Erfassung der Beurteilerreliabilität von Mystery Shoppern haben sich die Intraklassen-Korrelationskoeffizienten (ICC) als am geeignetsten gegenüber anderen Maßen erwiesen.⁴⁰⁷ Dies begründet sich beispielsweise darin, dass mehr als zwei Rater eingesetzt werden können. Selbst wenn dies nicht der Fall ist, ist der ICC der Produkt-Moment-Korrelation ($r_{x,y}$) vorzuziehen, sofern die Beurteiler den Variablen x und y nicht eindeutig zugeordnet werden können. Ferner kann flexibel entschieden werden, ob Mittelwertsunterschiede als Fehlerquellen berücksichtigt werden oder nicht.⁴⁰⁸ Der unjustierte, einfaktorielle ICC ist einzusetzen, wenn nicht alle Merkmalsträger von denselben Ratern beurteilt wurden.⁴⁰⁹ Ist

⁴⁰³ Vgl. Wirtz/Caspar (2002: 48)

⁴⁰⁴ Vgl. hierzu und nachfolgend ebenda (2002: 55ff.)

⁴⁰⁵ Vgl. ebenda (2002: 75)

⁴⁰⁶ Für einen detaillierten Überblick unterschiedlicher Maße zur Bestimmung der Beurteilerreliabilität (und -Übereinstimmung) vgl. Wirtz/Caspar (2002)

⁴⁰⁷ Vgl. hierzu die sehr ausführliche Betrachtung der unterschiedlichen Koeffizienten zur Bestimmung der Interraterreliabilität von Mystery Shoppern bei Schmidt (2007: 37ff.)

⁴⁰⁸ Vgl. Wirtz/Caspar (2002: 232)

⁴⁰⁹ Entgegen den justierten und unjustierten zweifaktoriellen ICC, die dies nicht zulassen. Diesen ICC setzt auch Schmidt ein, Vgl. Schmidt (2007: 158)

MS_{zw} die Varianz zwischen den Beobachtungsfällen und MS_{inn} die Varianz innerhalb der Beobachtungsfälle und k die Anzahl der Rater, berechnet sich Reliabilität ρ wie folgt:⁴¹⁰

$$(Z-3) \quad \rho_{unjust, einfakt} = ICC_{unjust, einfakt} = \frac{MS_{zw} - MS_{inn}}{MS_{zw} + (k-1) \cdot MS_{inn}}$$

Der Wertebereich für den ICC liegt zwischen 0 (keine Interraterreliabilität) und 1, wobei ein Wert von 0,4 die Untergrenze einer brauchbaren Schätzung der Interraterreliabilität durch den ICC markiert.⁴¹¹ Die Signifikanz der Abweichung der $ICC_{unjust, einfakt}$ von $\rho_0 = 0$ wird mittels der F-verteilten Größe $F_0 = MS_{zw} / MS_{inn}$ getestet. Der zugehörige F-Wert ist in der Flächentabelle der F-Verteilung nachzulesen und mit dem empirischen Wert zu vergleichen. Sofern der empirische Wert den theoretischen übersteigt, ist der ermittelte ICC als statistisch gesichert anzusehen.⁴¹²

3.3.1.3 Erfassung der Sorgfaltstendenz

Letztlich ist die Sorgfaltstendenz zu erfassen, die sich aus zwei Komponenten zusammensetzt. Erstens wird die *Gewissenhaftigkeit* der Mystery Shopper ermittelt, indem die identische Beantwortung inhaltlich gleicher Fragen überprüft wird. Dafür sind zwei inhaltlich gleiche Items an unterschiedlichen Stellen im Fragebogen positioniert worden. Die Gewissenhaftigkeit wird erfasst, indem die übereinstimmenden und nicht übereinstimmenden Itempaare ermittelt werden und einer der Ausprägungen „Übereinstimmung“ oder „keine Übereinstimmung“ zugeordnet werden. Zweitens wird die *Vollständigkeit* des Ausfüllens des Fragebogens ermittelt. Diese wird ebenfalls durch Abzählen und zwar der ausgelassenen Items erfasst. Die Fälle werden dann einer der Ausprägungen „Felder ausgelassen“ oder „vollständig ausgefüllt“ zugeordnet. Für beide Teilkomponenten werden die Ausprägungen jeweils für die einzelnen Merkmalsausprägungen der Gruppen gegenübergestellt. Dies erfolgt mit Hilfe einer Kreuztabelle. Die Überprüfung der Signifikanz des Unterschieds erfolgt bei den 2x2-Kreuztabellen (z.B. Geschlecht) mit dem Fisher-Test und bei allen anderen mit dem Chiquadrat Test. Ferner werden die Zusammenhänge der Teilkomponenten in Bezug auf die Merkmalsträger mit einem Korrelationskoeffizienten überprüft. Dafür wird der Rangkorrelationskoeffizient nach Spearman r_s eingesetzt. Dieser setzt keine Normalverteilung voraus und ist auch bei kleinen Fallzahlen anwendbar.⁴¹³ Als symmetrischer Test sind die Variablen weder als abhängig noch als unabhängig zu bezeichnen. Der Wertebereich für r_s liegt zwischen -1,0 (positiver linearer Zusammenhang)

⁴¹⁰ Vgl. Wirtz/Caspar (2002: 174)

⁴¹¹ Vgl. vor allem Schmidt (2007: 44) und die dort zitierte Literatur

⁴¹² Vgl. Wirtz/Caspar (2002:176)

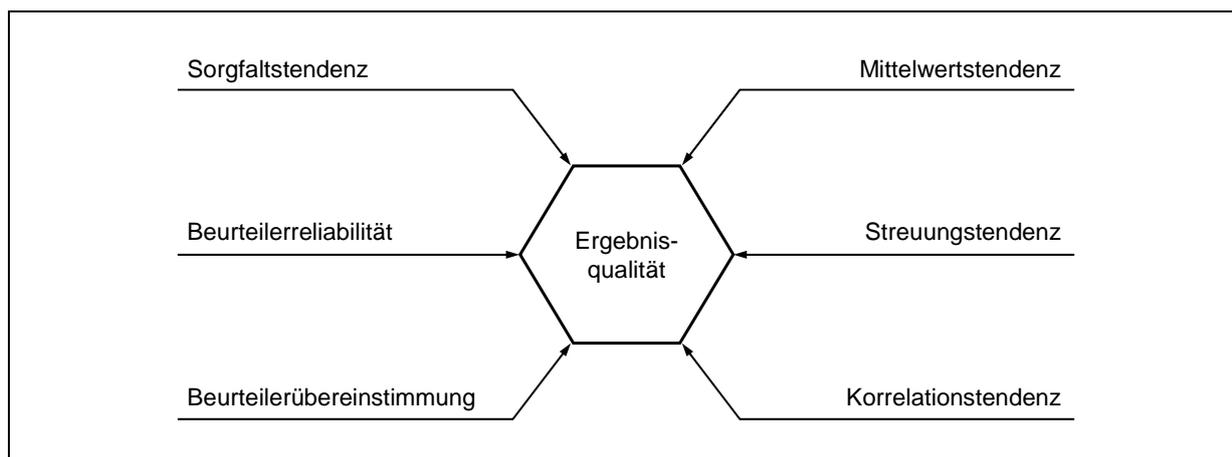
⁴¹³ Vgl. hierzu und nachfolgend Zöfel (2003: 156f.)

und 1,0 (negativer linearer Zusammenhang). Ein Wert von 0 hingegen deutet auf einen nicht vorhandenen linearen Zusammenhang hin.

3.3.2 Bewertung der Komponenten der Ergebnisqualität

In den voranstehenden Abschnitten wurde gezeigt, wie die einzelnen Komponenten, die in Abbildung 3-3-2 zusammengefasst sind, erfasst bzw. berechnet werden.

Abbildung 3-3-2: Komponenten der Ergebnisqualität von Mystery Shoppern



Quelle: Eigene Darstellung

Es stellt sich die Frage, wie die einzelnen Komponenten der Ergebnisqualität zu bewerten sind. Konkret ist zu beantworten, ob den Komponenten unterschiedliches Bedeutungsgewicht zukommen sollte. Ferner stellt sich die Frage, ob die aufgeführten Komponenten in Ermangelung eines „ultimativen Kriteriums“ zu einem solchen aggregiert werden sollten.

Wie gezeigt wurde, sind die einzelnen Komponenten bezüglich der zu erfassenden bzw. zu berechnenden Größen grundsätzlich verschieden voneinander. Eine Unterscheidungsmöglichkeit liegt in ihrer Darstellungsform: Während die Mittelwertstendenz beispielsweise den Unterschied zwischen zwei Gruppen bezüglich Milde und Strenge in der Urteilstendenz abbildet, liefern Beurteilerübereinstimmung und Interraterreliabilität einen konkreten Wert über die Zuverlässigkeit der Beurteilung, der auch unabhängig von einem Vergleich mit einer anderen Gruppe aussagekräftig ist. Wie oben bereits gezeigt wurde, lassen sich mit Urteilstendenzen Beurteilungsgewohnheiten beschreiben, die ein Indikator für Schulungsmaßnahmen sein können, während mit den Komponenten Zuverlässigkeit und Sorgfaltstendenz konkrete Aussagen über die Qualität der abgelieferten Arbeit eines Mystery Shoppers getroffen werden können. Insofern gibt es solche Kennziffern (nämlich Zuverlässigkeit und Sorgfaltstendenz), die für sich allein betrachtet zu einer Aussage führen und solche (nämlich die Urteilstendenzen), die für eine Aussage im Vergleich zu einer anderen Gruppe dargestellt werden müssen. Allerdings lässt diese Unterscheidung keinen Schluss auf das mögliche Bedeutungsgewicht der einzelnen Komponenten zu, sondern ist vielmehr von der Zielsetzung der Untersuchung abhängig. Sollen Beurteilungsgewohnheiten

beschrieben werden, sind die Urteilstendenzen als wichtiger anzusehen, während bei einer Erhebung der Arbeitsqualität von Mystery Shoppern Zuverlässigkeit und Sorgfalt eine größere Bedeutung zukommt. Gleichwohl gestaltet sich eine Bedeutungsgewichtung innerhalb einer der beiden Zielsetzungen als schwierig, wie bereits oben gezeigt wurde. Daher werden die Komponenten einzeln berichtet und eine entsprechende Interpretation erfolgt auf Basis der Datenlage.

Die Aggregation zu einem „ultimativen Kriterium“, etwa in Gestalt eines messbaren Wertes, gestaltet sich auch neben den angeführten Gründen als schwierig. Einerseits sind die zu erfassenden bzw. zu berechnenden Größen bzgl. ihrer Einheit sehr unterschiedlich. So gestaltet es sich schwierig, die Anzahl abgezahlter Fehler mit einer identifizierten milden Beurteilung in Beziehung zu setzen. Andererseits wird durch die Beschreibung am Anfang des Kapitels deutlich, dass die Größen sehr unterschiedliche Merkmale beschreiben, die einzeln betrachtet möglicherweise aussagekräftiger sind als ein aggregierter Wert. Daher werden die Größen einzeln dargestellt und nicht zu einem Wert aggregiert.

Damit dennoch veranschaulicht werden kann, wie die ermittelten Daten zur Einstufung verwendet werden können, wird hier vorgeschlagen, die gewonnenen Erkenntnisse in einer Matrix zu visualisieren (siehe Abbildung 3-3-3).

Abbildung 3-3-3: Arbeitsqualität und Schulungsbedarf

	<i>hoch</i>	ungeeignet	eher ungeeignet	bedingt geeignet
Schulungsbedarf	<i>mittel</i>	eher ungeeignet	bedingt geeignet	geeignet
	<i>niedrig</i>	bedingt geeignet	geeignet	sehr geeignet
		<i>niedrig</i>	<i>mittel</i>	<i>hoch</i>
		Arbeitsqualität		

Quelle: Eigene Darstellung

Bei der Einstufung wären die Merkmalsträger einer Gruppe miteinander zu vergleichen. Der Grad des Schulungsbedarfs könnte ermittelt werden, indem bei Abweichungen in allen drei Urteilstendenzen ein hoher Schulungsbedarf für die Merkmalsträger besteht, während bei keinen Abweichungen ein niedriger Schulungsbedarf besteht. Der Grad der Arbeitsqualität könnte hingegen bestimmt werden, indem jeweils ein Niveau für Zuverlässigkeit, Vollständigkeit und Gewissenhaftigkeit festgelegt wird, welches die Einordnung in niedrige,

mittlere oder hohe Arbeitsqualität ermöglicht. Vom Autor dieser Arbeit werden Richtwerte vorgeschlagen, die der Tabelle 3-3-3 zu entnehmen sind.

Tabelle 3-3-3: Bewertung der Arbeitsqualität

Komponente	Bewertung	
<i>Zuverlässigkeit</i>	niedrig:	weniger als 80% der Items werden zuverlässig beurteilt
	mittel:	mindestens 80% der Items werden zuverlässig beurteilt
	hoch:	mindestens 90% der Items werden zuverlässig beurteilt
<i>Vollständigkeit</i>	niedrig:	weniger als 80% der Merkmalsträger haben den Beobachtungsbogen vollständig ausgefüllt
	mittel:	mindestens 80% der Merkmalsträger haben den Beobachtungsbogen vollständig ausgefüllt
	hoch:	mindestens 90% der Merkmalsträger haben den Beobachtungsbogen vollständig ausgefüllt
<i>Gewissenhaftigkeit</i>	niedrig:	weniger als 80% der Merkmalsträger haben das doppelt vorhandene Item identisch beantwortet
	mittel:	mindestens 80% der Merkmalsträger haben das doppelt vorhandene Item identisch beantwortet
	hoch:	mindestens 90% der Merkmalsträger haben das doppelt vorhandene Item identisch beantwortet

Quelle: Eigene Darstellung

Mit der hier vorgeschlagenen Einstufung würden sowohl durch Urteilstendenzen gewonnene Erkenntnisse über erforderliche Schulungsmaßnahmen als auch Indikatoren der Arbeitsqualität Berücksichtigung finden. Demnach sind Mystery Shopper mit einer hohen Arbeitsqualität und niedrigem Schulungsbedarf sehr geeignet, Mystery Shopper mit einer niedrigen Arbeitsqualität und hohem Schulungsbedarf ungeeignet und die anderen Konstellationen geeignet, bedingt geeignet bzw. eher ungeeignet.

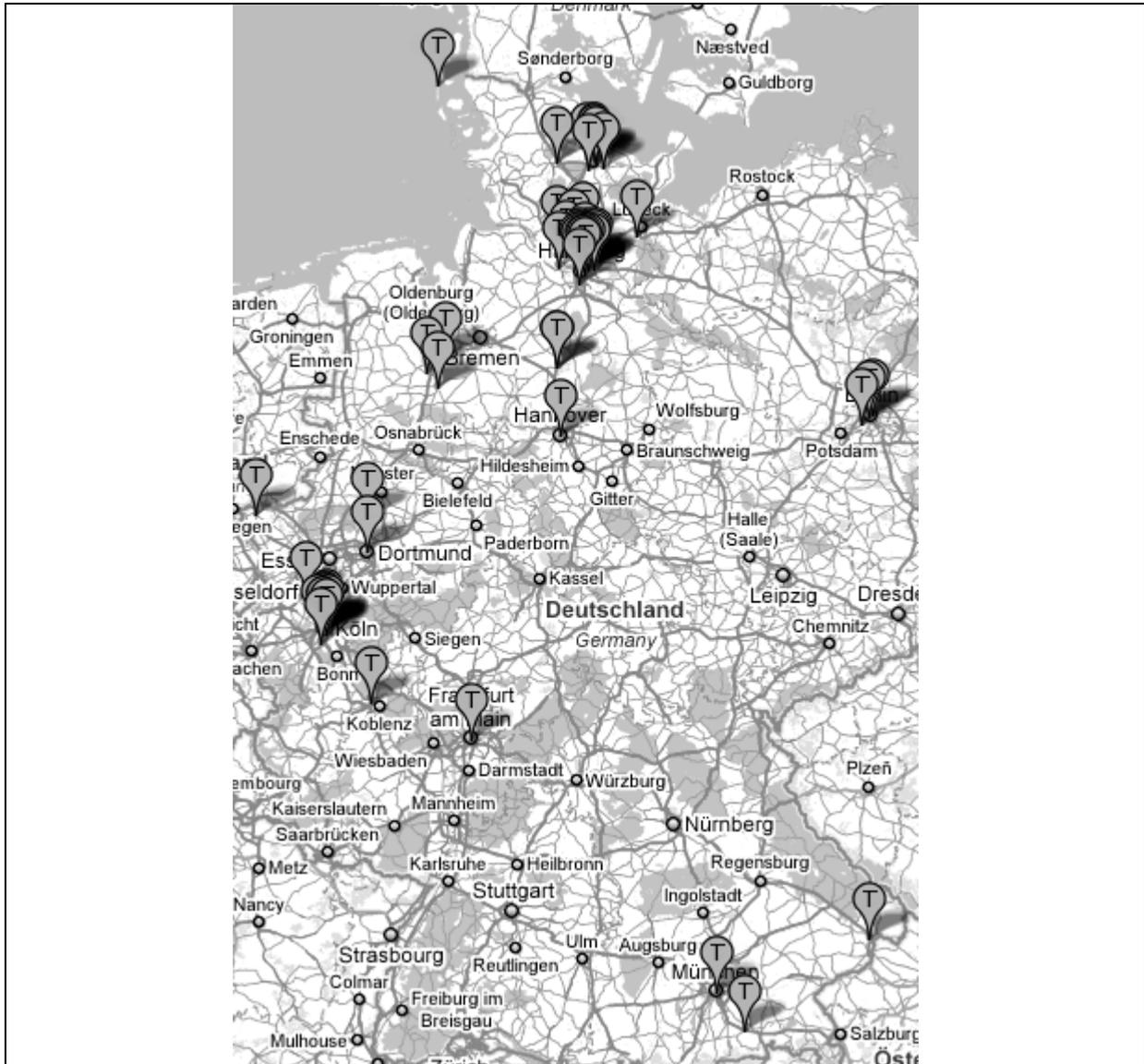
Sofern lediglich Merkmalsträger einer Gruppe verglichen werden sollen, könnte die Eignung auch auf Basis eines direkten Vergleichs entschieden werden. Demnach wären solche Merkmalsträger einer Gruppe geeigneter, die häufiger einen vorteilhafteren Wert aufweisen. Es ist zu betonen, dass die hier vorgeschlagene Einstufung auf pragmatischen Überlegungen beruht.

3.3.3 Aufbau und Ablauf der Erhebung

Abschließend sollen die Kerndaten zum Aufbau und Ablauf der Erhebung geschildert werden. Bei der Auswahl der zu besuchenden Fast Food-Restaurants, die als Testobjekt dienen sollen,

wurde eine Stichprobe gewählt, die sich über das Bundesgebiet verteilt und städtische wie ländliche Gebiete umfasst, wobei die Schwerpunkte in Norddeutschland und Nordrhein-Westfalen liegen (Vgl. Abbildung 3-3-4⁴¹⁴).

Abbildung 3-3-4: Orte der Testdurchführung



Quelle: Eigene Darstellung mit Google Maps

Bei der Durchführung des Mystery Shoppings im Januar 2008 wurden die Tester paarweise eingesetzt, wobei die Paare sich zur Auswertung der Beurteilerübereinstimmung und –reliabilität in mindestens einem Merkmal (z.B. zwei Männer) glichen. Für die Zusammensetzung der Stichprobe wurden Personen ausgewählt, die über die geforderten soziodemografischen Merkmale verfügen, um einen Vergleich der Ergebnisqualität der unterschiedlichen Merkmalsgruppen zu ermöglichen. Da eine große Anzahl an verschiedenen

⁴¹⁴ Die auf Basis von Google Maps erstellte Karte soll lediglich der Visualisierung dienen, denn sie gibt nicht alle Standorte zuverlässig wieder.

Testern einzusetzen war, wurde bei der Auswahl auf den Testerbestand eines Mystery Shopping-Anbieters⁴¹⁵ zugegriffen. Daneben wurden vom Autor eigenständig Mystery Shopper rekrutiert. Die Zusammensetzung der Stichprobe ist in Tabelle 3-3-4 dargestellt.

Für die Anzahl der Testerpaare liegen ganz unterschiedliche Werte vor. Die Anzahl schwankt zwischen null Paaren (bei Affinität McDonald's = sehr gern) und 110 Paaren (Mystery Shopper-Erfahrung = keine). Damit eine auswertbare Größe generiert werden kann, sind die Klassen entweder zusammenzuziehen, sofern dies sinnvoll ist, oder sie werden nicht in die Auswertung einbezogen.

Die eingesetzten Tester sind anhand von konkreten schriftlichen und mündlichen Verhaltensanweisungen mit dem Ablauf der Testsituation vertraut gemacht worden. Dabei wurde sichergestellt, dass sich die Mystery Shopper in der Testsituation wie normale Kunden verhalten. Die Aufzeichnung der Ergebnisse auf dem Beobachtungsbogen erfolgte unmittelbar nach der Testsituation, allerdings jeweils unabhängig voneinander.

Abschließend wird in Tabelle 3-3-5 die Erfüllung der in Kapitel 2.2.3.2 ausgeführten Voraussetzungen für den erfolgreichen Einsatz von Mystery Shoppern für diese Untersuchung zusammengefasst.

Nachdem im vorangegangenen Kapitel den ersten beiden Teilzielen der Arbeit Rechnung getragen wurde, wurde in diesem Kapitel das dritte Teilziel erreicht. Das nachfolgende Kapitel widmet sich dem vierten Teilziel der Arbeit und stellt die Ergebnisqualität von Mystery Shoppern mit unterschiedlichen Merkmalsausprägungen gegenüber.

⁴¹⁵ dies war der Testerstab der ABH Marketingservice GmbH, Köln.

Tabelle 3-3-4: Zusammensetzung der Stichprobe

Merkmal	Absolut	Relativ	Paare
<i>Stichprobenumfang</i>	376	100%	188
<i>Geschlecht</i>			
- männlich	193	51,3%	43
- weiblich	183	48,7%	38
<i>Alter</i>			
- 18-30 Jahre	102	27,1%	27
- 31-45 Jahr	206	54,8%	76
- 46-59 Jahre	38	10,1%	8
- >60 Jahre	30	8,0%	11
<i>Höchster Bildungsabschluss</i>			
- Hauptschule	14	3,7%	3
- Realschule	62	16,5%	15
- Abitur	103	27,4%	28
- Studium	197	52,4%	70
<i>Tätigkeit</i>			
- Schüler/Student	51	13,6%	21
- Hausfrau/Rentner	31	8,2%	9
- voll berufstätig	255	67,8%	109
- Teilzeit berufstätig	39	10,4%	6
<i>Mystery Shopping-Erfahrung</i>			
- noch nie	277	73,7%	110
- 1-5 Testkäufe	81	21,5%	15
- >5 Testkäufe	18	4,8%	3
<i>Affinität Fast Food-Restaurants²</i>			
- sehr gern	19	5,1%	0
- eher gern	73	19,5%	6
- teils-teils	163	43,5%	36
- weniger gern	82	21,9%	13
- gar nicht gern	31	10,1%	4
<i>Affinität McDonald's³</i>			
- sehr gern	24	6,5%	3
- eher gern	99	26,7%	17
- teils-teils	143	38,5%	30
- weniger gern	74	19,9%	13
- gar nicht gern	31	8,4%	2
<i>Anzahl Besuche in einem Fast Food Restaurant⁴</i>			
- keinmal	14	3,7%	2
- 1-12 Besuche	186	49,5%	58
- 13-52 Besuche	150	39,9%	41
- >52 Besuche	26	6,9%	1
<i>Anzahl Besuche bei McDonald's⁴</i>			
- keinmal	36	9,6%	5
- 1-12 Besuche	255	67,8%	90
- 13-52 Besuche	80	21,3%	12
- >52 Besuche	5	1,3%	1

¹ n=374; bei den Paaren haben die beiden Mystery Shopper das gleiche Geschlecht² Losgelöst von Ihrer heutigen Aufgabe als Testkunde, wie gern essen Sie in einem Fast Food-Restaurant? (n=375)³ Wie gerne essen Sie bei McDonald's? (n=371)⁴ Besuche in den letzten zwölf Monaten**Quelle:** Eigene Darstellung

Tabelle 3-3-5: Voraussetzungen für den erfolgreichen Einsatz von Mystery Shoppern und ihre Erfüllung

Voraussetzung	Erfüllung
1. Die Kundenkontaktsituation soll möglichst realitätsnah konstruiert werden.	Die Mystery Shopper sollen ein typisches Produkt, z.B. ein Getränk, bestellen. Die Entwicklung des Szenarios erfolgte mittels Service Blueprinting und wurde im Rahmen eines Pretests überprüft.
2. Die Mystery Shopper sollen die Charakteristika typischer Kunden aufweisen.	Fast Food-Restaurants bedienen die gesamte Breite der Bevölkerung.
3. Mystery Shopper sollen sich wie normale Kunden verhalten.	Die Mystery Shopper erhalten entsprechende schriftliche Anweisungen.
4. Mystery Shopper sollen hinsichtlich des Ablaufs der Testkaufsituation und des Verhaltens eines realen Kunden geschult werden.	Siehe 3.
5. Der Dienstleistungsprozess soll in seine einzelnen Phasen zerlegt werden, für die Beobachungskriterien festzulegen sind.	Der Prozess wurde mittels Blue Printing in seine einzelnen Phasen zerlegt, für die Beobachungskriterien definiert wurden.
6. Der Beobachtungsbogen soll konkrete Erlebnisinformation und nicht abstrakte Dienstleistungsdimensionen beinhalten.	Der abstrakte Beobachtungskatalog aus Tabelle 3-2-3 wurde um Erlebnisinformation aus Tabelle 3-2-4 ergänzt.
7. Die Erlebnisse sind unmittelbar nach der Testkaufsituation aufzuzeichnen, um einen Informationsverlust zu verhindern.	Die Aufzeichnung erfolgt individuell unmittelbar nach dem Testbesuch.
8. Der Beobachtungsbogen ist klar zu strukturieren und das Szenario klar zu beschreiben.	Siehe 3.
9. Individuelle und situationsspezifische Faktoren der Mystery Shopper sind bei der Untersuchungsplanung zu berücksichtigen.	Die Berücksichtigung erfolgt über die Auswahl und die Anweisungen.
10. Beobachtungsspielräume sind durch ausreichende Operationalisierung der Konstrukte einzuschränken.	Die Operationalisierung erfolgt gemäß Skalen wie in den Tabellen 3-2-3 und 3-2-4 dargestellt.
11. Es ist auf eine angemessene Anzahl an Bewertungskategorien zu achten, um die Erinnerung der Mystery Shopper nicht zu überfordern.	Der Umfang wurde in einem Pretest als anforderungsgerecht beurteilt.
12. Es sind ausreichend hohe Fallzahlen zu erbringen.	Es wurde eine Fallzahl von insgesamt 376 Fällen erbracht, wobei die Tester paarweise eingesetzt werden.

Quelle: Eigene Darstellung

4 Die Ergebnisqualität von Mystery Shoppern im Spiegel empirischer Befunde

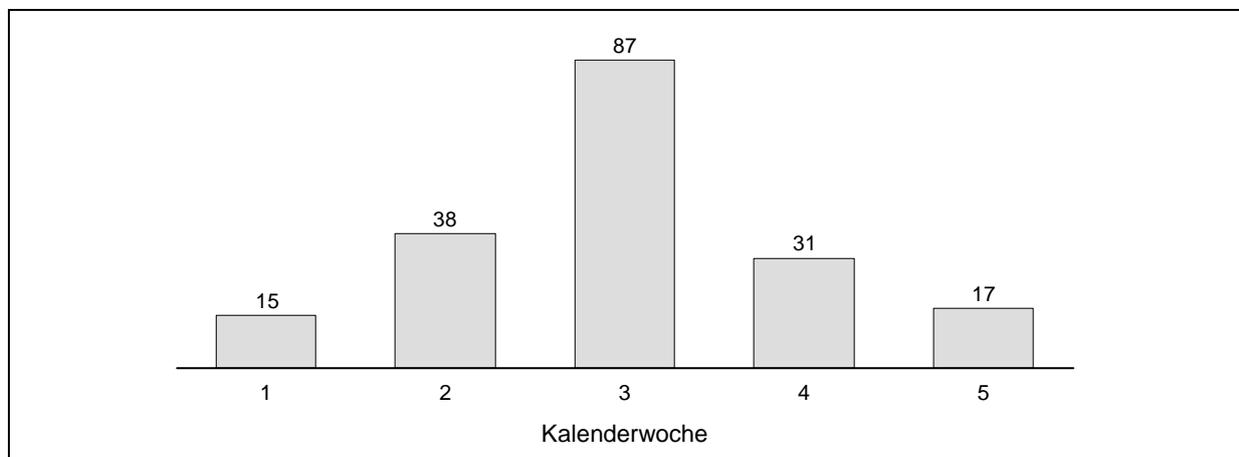
In diesem Kapitel werden die Ergebnisse der empirischen Untersuchung präsentiert. Im ersten Abschnitt wird das Teilnehmerverhalten beschrieben. Im zweiten Teil werden die Ergebnisse für die einzelnen Komponenten und die unterschiedlichen Gruppen dargelegt. Im dritten Abschnitt werden die Kernaussagen der Untersuchung wiedergegeben.

4.1 Teilnehmerverhalten der Erhebung

Den 188 Testpaaren wurde zu Beginn der ersten Kalenderwoche 2008 der Fragebogen zugesendet. Sie wurden gebeten, die ausgefüllten Bögen bis zum Beginn der vierten Kalenderwoche zurückzusenden.⁴¹⁶ Nach Rücksprache wurde der Zeitraum auf das Ende der Kalenderwoche fünf verlängert. Alle 188 Bögen wurden bis dahin zurückgesandt. Die Rücklaufquote von 100% ist allerdings wenig verwunderlich, da Testkunden in der Regel entlohnt werden und daher ein Anreiz zur Auftrags Erfüllung besteht.

Wie Abbildung 4-1-1 zu entnehmen ist, liegt der Schwerpunkt der durchgeführten Testkäufe in der dritten Kalenderwoche, während die Kalenderwochen 1 und 5 sowie 2 und 4 ähnliche Werte aufweisen.

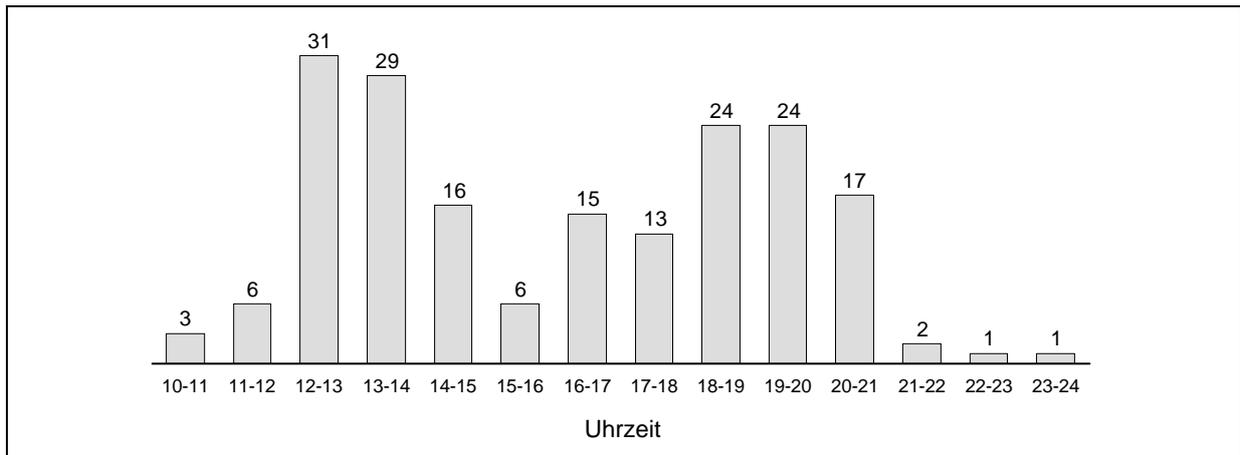
Abbildung 4-1-1: Verteilung der paarweise durchgeführten Testkäufe auf die Kalenderwochen 2008



Quelle: Eigene Darstellung

Die Testkäufe wurden von den Paaren zu unterschiedlichen Tageszeitpunkten durchgeführt. Wie in Abbildung 4-1-2 dargestellt, liegt die Spitze um die Mittagszeit und eine weitere Spitze am frühen Abend. Morgens und am späten Abend sind die wenigsten Testkäufe durchgeführt worden.

⁴¹⁶ Dabei wurde den Teilnehmern drei Optionen zur Wahl gestellt: Postweg, per Email oder per Fax. Eine Aufschlüsselung der Nutzungsfrequenz der unterschiedlichen Rücksendewege erfolgte nicht.

Abbildung 4-1-2: Verteilung der paarweise durchgeführten Testkäufe auf Tageszeiten

Quelle: Eigene Darstellung

4.2 Ergebnisse der Komponenten zur Bestimmung der Ergebnisqualität von Mystery Shoppern

In diesem Abschnitt werden die Ergebnisse der im vorangegangenen Kapitel festgelegten Komponenten präsentiert. Dies sind die Urteilstendenzen Mittelwertstendenz, Streuungstendenz und Korrelationstendenz, die Übereinstimmungsmaße Beurteilerreliabilität und Beurteilungsübereinstimmung, die mit Zuverlässigkeit bezeichnet werden, sowie die Sorgfaltstendenzen Gewissenhaftigkeit und Vollständigkeit. Bei der Präsentation der Werte werden die Ausprägungen für die unterschiedlichen Merkmalsgruppen jeweils gegenübergestellt.

4.2.1 Mittelwertstendenz von Mystery Shoppern

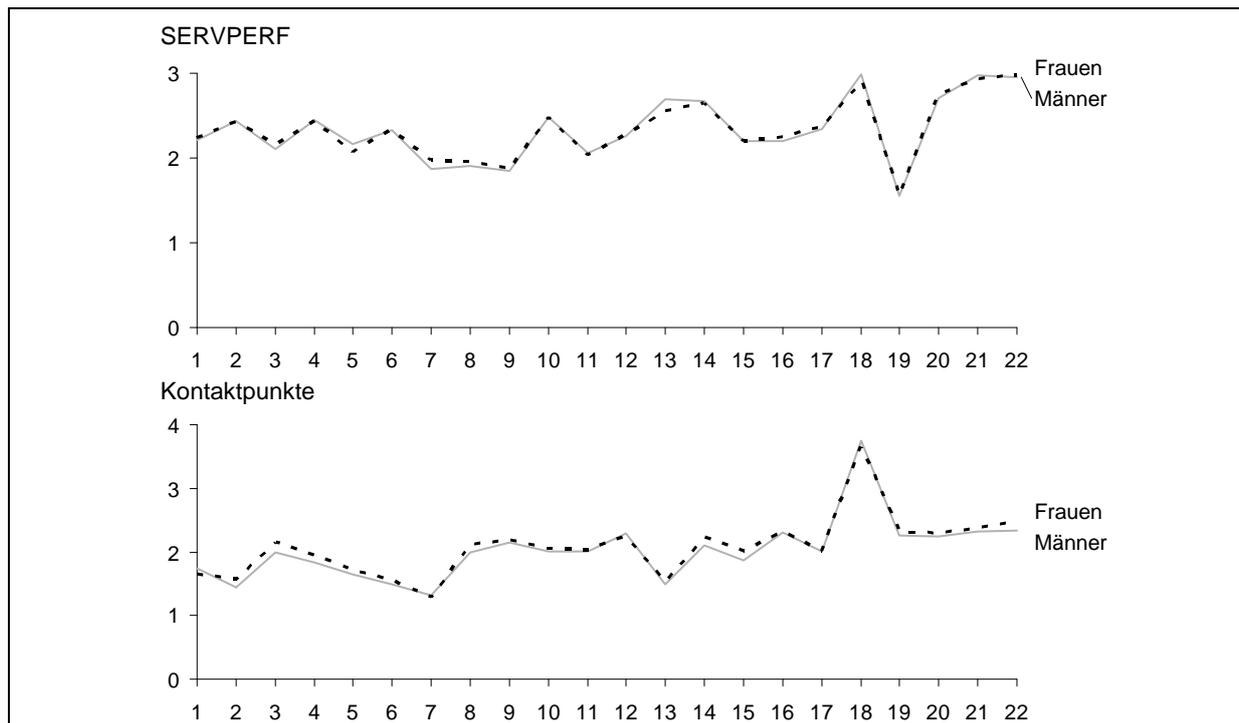
Die Ergebnisse der Komponente Mittelwertstendenz soll für die einzelnen Merkmalsgruppen jeweils in zwei Teilen berichtet werden. Im jeweiligen ersten Teil eines jeden Abschnitts wird die Mittelwertstendenz für die Operationalisierung der SERVPERF-Dimensionen dargelegt⁴¹⁷, im zweiten Teil des jeweiligen Abschnitts wird die Mittelwertstendenz der Operationalisierung der Kontaktpunkte wiedergegeben⁴¹⁸. Der Übersichtlichkeit halber werden die Mittelwertprofile, die auf einer Skala von 1 bis 5 liegen, grafisch dargestellt. Eine tabellarische Übersicht der Werte für Mittelwert, Standardabweichung und Varianz für die untersuchten Merkmalsausprägungen findet sich im Anhang.

4.2.1.1 Mittelwertstendenz bezüglich des Geschlechtes der Mystery Shopper

Zur Ermittlung der Komponente Mittelwertstendenz bezüglich des Geschlechts der Mystery Shopper wurden Frauen und Männer hinsichtlich ihrer Mittelwerte bei den einzelnen Items gegenübergestellt. Das Ergebnis ist in Abbildung 4-2-1-1 dargestellt.

⁴¹⁷ Vgl. Tabelle 3-2-3

⁴¹⁸ Vgl. Tabelle 3-2-4

Abbildung 4-2-1-1: Komponente Mittelwertstendenz bezüglich Geschlecht

Quelle: Eigene Darstellung

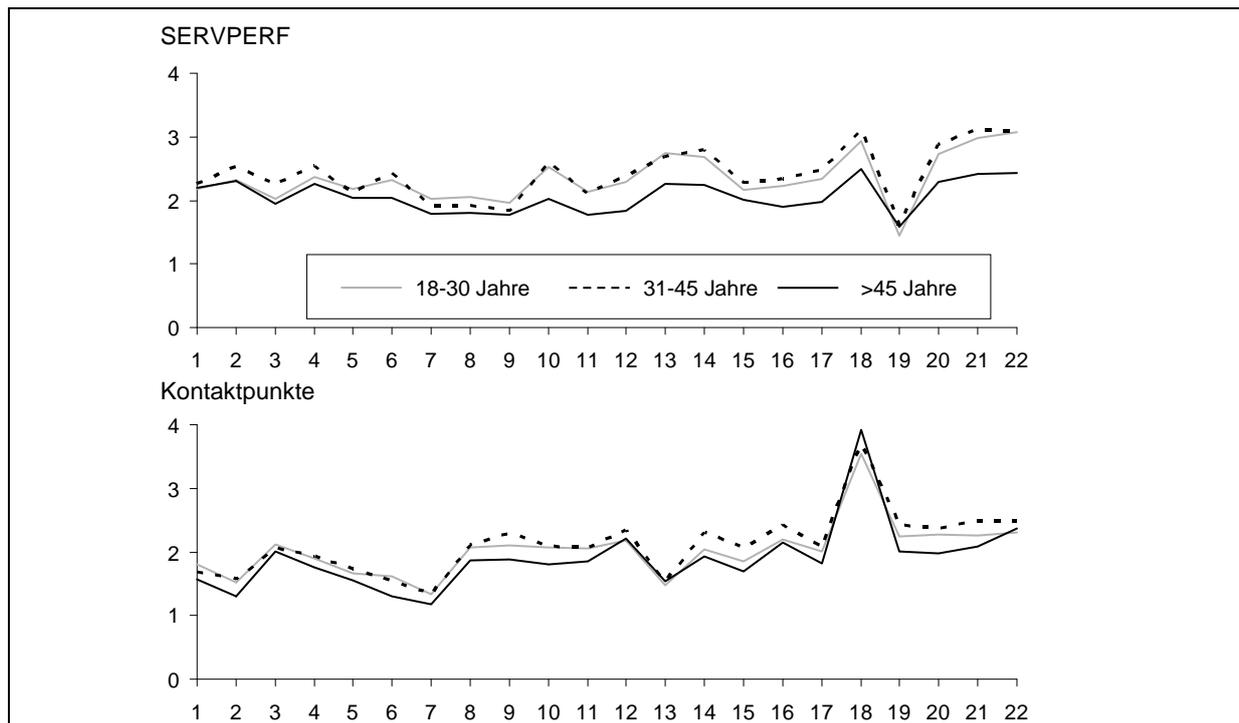
Wie der Abbildung zu entnehmen ist, weisen Männer und Frauen sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte ein ähnliches Mittelwertprofil auf. Die Annahme eines nicht vorhandenen Unterschieds wird auch durch den H-Test untermauert, der für keinen Wert eine Irrtumswahrscheinlichkeit von $p < 0,05$ annimmt.⁴¹⁹ Insofern ist die Mittelwertstendenz bei Männern und Frauen nicht signifikant verschieden.

4.2.1.2 Mittelwertstendenz bezüglich des Alters der Mystery Shopper

Basierend auf Altersangaben der Mystery Shopper wurden drei Altersklassen für die Auswertung der Komponente Mittelwertstendenz gebildet. Die ursprüngliche Einteilung in vier Altersklassen erschien bei näherer Betrachtung der Fallzahlen nicht sinnvoll.⁴²⁰ Die Ergebnisse der Gegenüberstellung der Mittelwertprofile der unterschiedlichen Altersklassen sind in Abbildung 4-2-1-2 dargestellt.

⁴¹⁹ Allerdings gibt es einen Wert, der hier nachrichtlich genannt werden soll: Das Item „Welchen Eindruck macht der Besuch insgesamt auf Sie?“ nimmt den Wert $p = 0,070$ an (Mittelwert Männer: 2,34; Frauen: 2,48). Dies ist ein schwacher Indikator für eine Tendenz zur Milde (Strenge) der Männer (Frauen) beim Gesamtzufriedenheitsurteil.

⁴²⁰ Siehe Tabelle 3-3-4 für die Anzahl der Fälle bei einer Aufteilung in vier Altersklassen.

Abbildung 4-2-1-2: Komponente Mittelwertstendenz bezüglich Altersklassen

Quelle: Eigene Darstellung

Grafisch gesehen weisen die unterschiedlichen Altersklassen sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte ein unterschiedliches Mittelwertprofil auf. Die Signifikanz des Unterschieds wird für einen Teil der Items durch den H-Test bestätigt, wie den Tabellen 4-2-1-2-1 und 4-2-1-2-2 zu entnehmen ist. Sofern eine Signifikanz nach dem H-Test festgestellt wurde, wurde sukzessive paarweise ein U-Test durchgeführt.

Tabelle 4-2-1-2-1: SERVPERF-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der Altersklassen

	Signifikanz			
	H-Test	Mann-Whitney-U-Test		
		18-30 vs. 31-45 Jahre	18-30 vs. >45 Jahre	31-45 vs. >45 Jahre
Die Angestellten machen einen ordentlichen Eindruck.	0,002**	0,013*	0,451	0,002**
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	0,034*	0,065	0,654	0,022*
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	0,002**	0,367	0,016*	0,000**
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	0,000**	0,492	0,002**	0,000**
Der Gast erhält von den Angestellten einen prompten Service.	0,010**	0,924	0,016*	0,003**
Die Mitarbeiter sind immer willens zu helfen.	0,000**	0,325	0,000**	0,000**
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	0,002**	0,568	0,001**	0,002**
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	0,000**	0,350	0,002**	0,000**
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	0,003**	0,411	0,012*	0,001**
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	0,000**	0,191	0,003**	0,000**
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	0,000**	0,199	0,006**	0,000**
McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	0,000**	0,284	0,003**	0,000**
McDonald's liegen Ihre Interessen am Herzen.	0,000**	0,205	0,000**	0,000**
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	0,000**	0,866	0,000**	0,000**

* Signifikanzniveau 5%, ** Signifikanzniveau 1%

Quelle: Eigene Darstellung

Bei 14 von 22 SERVPERF-Operationalisierungen wurde mittels H-Test ein signifikanter Unterschied zwischen den Altersklassen nachgewiesen, wobei 13 Items eine Irrtumswahrscheinlichkeit von $p < 0,01$ aufweisen. Vergleicht man die Altersklassen untereinander mit dem Mann-Whitney-U-Test wird die Signifikanz des Unterschieds zwischen den einzelnen Klassen deutlicher. Zwischen der Altersklasse der 18-30jährigen Mystery Shopper und der nächst höheren Altersklasse ist nur in einem Fall ein signifikanter Unterschied mit $p < 0,05$ nachzuweisen. Die 18-30jährigen und die Mystery Shopper älter als 45 Jahre weisen bis auf zwei Items sowie die 31-45jährigen und die >45 Jahre alten Mystery Shopper bei allen 14 Items signifikante bis hoch signifikante Unterschiede auf. Da die Mittelwerte der Tester >45 Jahre bei allen 14 Items am niedrigsten sind, ist dieser Altersklasse im Vergleich zu den anderen beiden für diese Items eine Tendenz zur Milde zu unterstellen. Anzumerken ist ferner, dass alle Items aus dem Bereich Responsiveness sowie vier von fünf Items aus dem Bereich Empathy diesen signifikanten Unterschied aufweisen.

Tabelle 4-2-1-2-2: Kontaktpunkt-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der Altersklassen

	Signifikanz			
	H-Test	Mann-Whitney-U-Test		
		18-30 vs. 31-45 Jahre	18-30 vs. >45 Jahre	31-45 vs. >45 Jahre
Wie ist der Zustand der Servicestation?	0,004**	0,778	0,004**	0,002**
Beim Herantreten an die Theke werde ich freundlich begrüßt.	0,009**	0,110	0,153	0,003**
Ich werde freundlich nach meinem Bestellwunsch gefragt.	0,037*	0,694	0,048*	0,011*
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	0,028*	0,082	0,337	0,016*
Die Mitarbeiter sind durchweg höflich zu mir.	0,019*	0,065	0,392	0,010**
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	0,020*	0,024*	0,695	0,026*
Die Mitarbeiter wirken ordentlich.	0,041*	0,405	0,096	0,013*
Die Mitarbeiter wirken hektisch und überlastet.	0,031*	0,205	0,012*	0,052
Die Mitarbeiter machen einen freundlichen Eindruck.	0,003**	0,081	0,095	0,001**
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	0,002**	0,468	0,017*	0,000**
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	0,001**	0,017*	0,200	0,000**

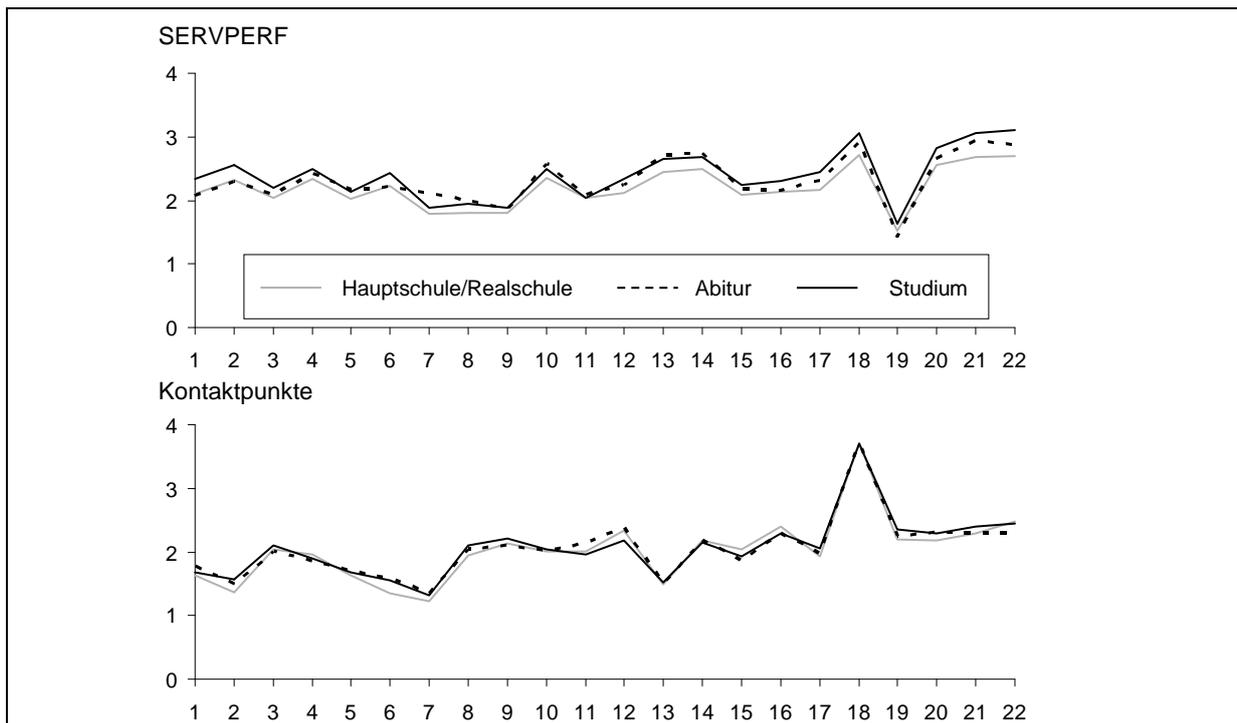
* Signifikanzniveau 5%, ** Signifikanzniveau 1%

Quelle: Eigene Darstellung

Bei der Hälfte der 22 Kontaktpunkt-Operationalisierungen wurde mittels H-Test ein signifikanter oder hoch signifikanter Unterschied der Altersklassen nachgewiesen. Vergleicht man die Signifikanz des Unterschieds bei den Altersklassen mit dem U-Test lässt sich zwischen der Altersklasse der 18-30jährigen und der 31-45jährigen bei zwei Items ein signifikanter Unterschied mit $p < 0,05$ nachweisen. Die 18-30jährigen und die Mystery Shopper älter als 45 Jahre weisen bei vier Items einen signifikanten Unterschied auf. Die beiden hohen Altersklassen unterscheiden sich bis auf eines der elf Items signifikant bis hoch signifikant. Da die höchste Altersklasse bei allen elf Items die niedrigsten Mittelwerte aufweist, kann ihr im Vergleich zur nächst niedrigeren Altersklasse für diese Items eine Tendenz zur Milde unterstellt werden. Die signifikanten Unterschiede betreffen fünf von acht Items aus dem Bereich Kundengespräch und alle Items aus dem Bereich Mitarbeiter.

4.2.1.3 Mittelwertstendenz bezüglich des höchsten Bildungsabschlusses der Mystery Shopper

Analog zur Ermittlung der Mittelwertstendenz bezüglich des Alters der Mystery Shopper wurden die ursprünglichen vier Bildungsklassen in Ermanglung ausreichender Fallzahlen zu drei Klassen zusammengeführt. Die Mittelwertprofile der Altersklassen sind in Abbildung 4-2-1-3 dargestellt.

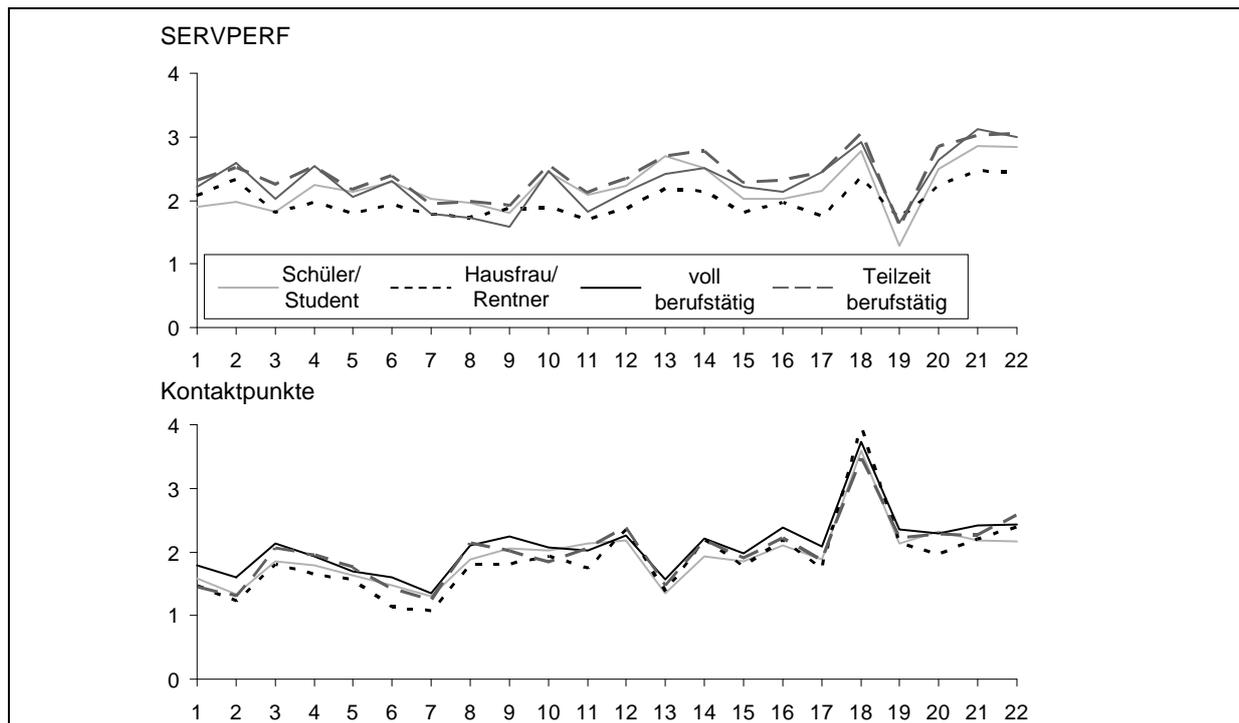
Abbildung 4-2-1-3: Komponente Mittelwertstendenz bezüglich des höchsten Bildungsabschlusses

Quelle: Eigene Darstellung

Die drei Bildungsklassen weisen sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte minimal unterschiedliche Mittelwertprofile auf, wie die Grafik verdeutlicht. Die Signifikanz des Unterschieds wird für fünf SERVPERF-Items und für ein Kontaktpunkt-Item durch den H-Test bestätigt. Nach Durchführung von paarweisen U-Tests zum Vergleich der Bildungsklassen zeigt sich bei vier Items zwischen Hauptschule/Realschule und Abitur sowie bei drei Items zwischen Abitur und Studium ein hoch signifikanter Unterschied. Auf eine detaillierte Darstellung wird an dieser Stelle verzichtet, da es sich um lediglich vier von 44 Items handelt und diese Anzahl als unbedeutend angesehen wird. Es ist festzustellen, dass bis auf diese Ausnahme kein Unterschied in der Mittelwertstendenz bei unterschiedlich hohem Bildungsabschluss der Mystery Shopper besteht.

4.2.1.4 Mittelwertstendenz bezüglich Tätigkeit der Mystery Shopper

In der nachfolgenden Abbildung 4-2-1-4 sind die Mittelwertprofile unterschiedlicher Tätigkeitsgruppen der Mystery Shopper dargestellt.

Abbildung 4-2-1-4: Komponente Mittelwertstendenz bezüglich Tätigkeit

Quelle: Eigene Darstellung

Die vier Tätigkeitsgruppen weisen sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte unterschiedliche Mittelwertprofile auf. Die Signifikanz des Unterschieds wird bei 18 der 22 SERVPERF-Items und bei acht der 22 Kontaktpunkt-Items durch den H-Test bestätigt, wie den Tabellen 4-2-1-4-1 und 4-2-1-4-2 zu entnehmen ist. Sofern eine Signifikanz nach dem H-Test festgestellt wurde, wurde sukzessive paarweise ein U-Test durchgeführt.

Tabelle 4-2-1-4-1: SERVPERF-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der Tätigkeit

	Signifikanz						
	H-Test	U-Test					
		Schüler/Student vs. Hausfrau/Rentner	Schüler/Student vs. voll berufstätig	Schüler/Student vs. Teilzeit berufstätig	Hausfrau/Rentner vs. voll berufstätig	Hausfrau/Rentner vs. Teilzeit berufstätig	voll berufstätig vs. Teilzeit berufstätig
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	0,031*	0,863	0,008**	0,051	0,111	0,217	0,757
Die Einrichtung ist ansprechend gestaltet.	0,003**	0,191	0,001**	0,002*	0,261	0,154	0,511
Die Angestellten machen einen ordentlichen Eindruck.	0,000**	0,962	0,000**	0,180	0,001**	0,172	0,061
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	0,002**	0,208	0,040**	0,075	0,001**	0,003**	0,708
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	0,021*	0,049*	0,440	0,777	0,002**	0,009**	0,686
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	0,009**	0,029*	0,387	0,753	0,001**	0,014*	0,677
Der Gast erhält von den Angestellten einen prompten Service.	0,012*	0,048*	0,747	0,197	0,004**	0,410	0,041*
Die Mitarbeiter sind immer willens zu helfen.	0,005**	0,030*	0,268	0,643	0,001**	0,067	0,113
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	0,018*	0,030*	0,833	0,130	0,008**	0,400	0,049*
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	0,001**	0,060	0,037*	0,953	0,000**	0,104	0,069
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	0,014*	0,321	0,055	0,285	0,004**	0,048*	0,711
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	0,020*	0,962	0,013*	0,406	0,041*	0,452	0,196
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	0,000**	0,026*	0,031*	0,133	0,000**	0,000**	0,977
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	0,006*	0,070	0,104	0,685	0,001**	0,052	0,403
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	0,010**	0,003**	0,006**	0,007**	0,205	0,682	0,410

McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	0,001**	0,193	0,028*	0,693	0,000**	0,110	0,141
McDonald's liegen Ihre Interessen am Herzen.	0,017*	0,092	0,260	0,163	0,004**	0,004**	0,563
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	0,015*	0,092	0,189	0,467	0,002**	0,020*	0,726

* Signifikanzniveau 5%, ** Signifikanzniveau 1%

Quelle: Eigene Darstellung

Signifikante Unterschiede sind bei der SERVPERF-Operationalisierung mit 15 Items insbesondere zwischen den Gruppen Hausfrau/Rentner und voll berufstätig auszumachen. Die Gruppen Schüler/Student und voll berufstätig unterscheiden sich bei neun Items signifikant voneinander. Bei allen Items bewertet die Gruppe der voll Berufstätigen strenger als die anderen beiden Gruppen. Insofern kann ihr eine Tendenz zur Strenge unterstellt werden. Der signifikante Unterschied zu der Gruppe Hausfrau/Rentner tritt bei allen Items aus den Bereichen Responsiveness und Assurance sowie bei vier von fünf Items aus dem Bereich Empathy auf. Der signifikante Unterschied zur Gruppe Schüler/Student besteht bei allen Items des Bereiches Tangibles und drei von vier Items aus dem Bereich Assurance.

Tabelle 4-2-1-4-2: Kontaktpunkt-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der Tätigkeit

	Signifikanz						
	H-Test	U-Test					
		Schüler/Student vs. Hausfrau/Rentner	Schüler/Student vs. voll berufstätig	Schüler/Student vs. Teilzeit berufstätig	Hausfrau/Rentner vs. voll berufstätig	Hausfrau/Rentner vs. Teilzeit berufstätig	voll berufstätig vs. Teilzeit berufstätig
Wie ist der Sauberkeitszustand der näheren Umgebung des Resturanteingangs?	0,006**	0,322	0,088	0,409	0,013*	0,803	0,011*
Wie ist der Zustand von Fenstern und Fassaden?	0,000**	0,455	0,004**	0,977	0,002**	0,535	0,022*
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	0,021*	0,865	0,008**	0,302	0,034*	0,337	0,502
Wie ist der Zustand der Servicestation?	0,000**	0,001**	0,302	0,698	0,000**	0,002**	0,169
Wie ist der Zustand des Thekenbereichs?	0,010**	0,011*	0,436	0,729	0,001**	0,022*	0,257
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	0,042*	0,601	0,034	0,121	0,034*	0,106	0,731
Die Mitarbeiter wirken ordentlich.	0,010**	0,390	0,045	0,952	0,010**	0,374	0,075

Welchen Eindruck macht der Besuch insgesamt auf Sie?	0,032*	0,244	0,019*	0,006**	0,676	0,207	0,128
--	--------	-------	--------	---------	-------	-------	-------

* Signifikanzniveau 5%, ** Signifikanzniveau 1%

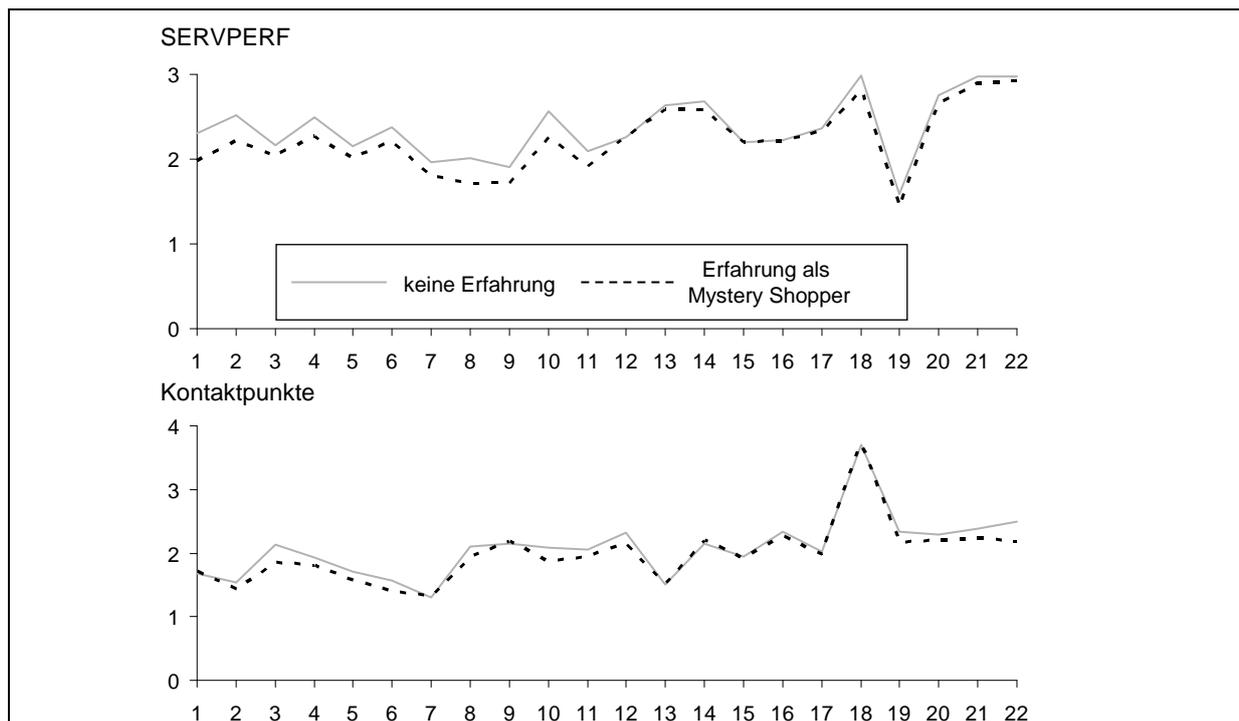
Quelle: Eigene Darstellung

Die Signifikanz des Unterschieds zwischen den Gruppen voll berufstätig und Hausfrau/Rentner spiegeln auch die Ergebnisse der Kontaktpunkt-Analyse wieder (sieben Items mit $p < 0,05$). Allerdings sind die Unterschiede eher bei tangiblen Bewertungsobjekten, wie dem Sauberkeitszustand zu finden. Auch hier bewerten die voll Berufstätigen strenger als die Gruppe Hausfrau/Rentner.

4.2.1.5 Mittelwertstendenz bezüglich Erfahrung als Mystery Shopper

Zur Beurteilung der Mittelwertstendenz bezüglich der Erfahrung als Mystery Shopper wurden zwei Gruppen gebildet: Solche mit Einsatzerfahrung als Tester und solche ohne. Die Mittelwertprofile werden in der Abbildung 4-2-1-5 grafisch veranschaulicht.

Abbildung 4-2-1-5: Komponente Mittelwertstendenz bezüglich Erfahrung als Mystery Shopper



Quelle: Eigene Darstellung

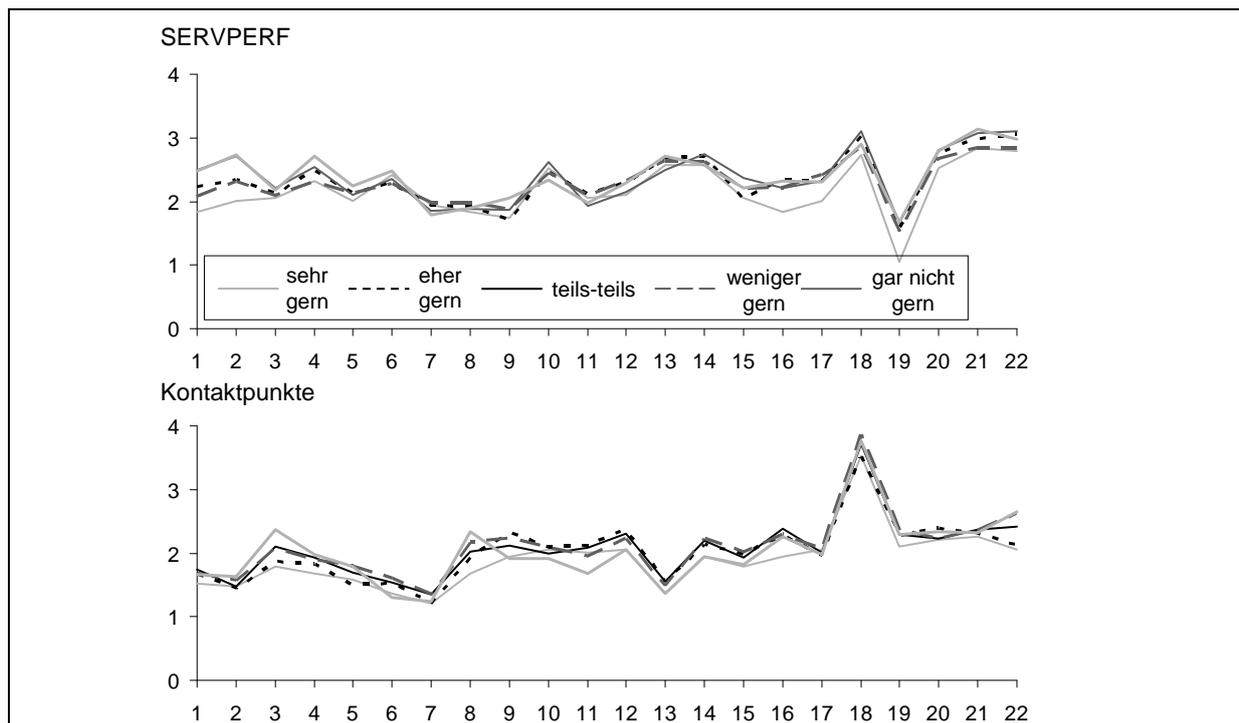
Wie der Abbildung zu entnehmen ist, weisen Tester mit und ohne Mystery Shopping-Erfahrung sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte nur minimal unterschiedliche Mittelwertprofile auf. Die Signifikanz des Unterschieds wird durch den H-Test nur für sechs SERVPERF-Items und vier Kontaktpunktitems mit einer Irrtumswahrscheinlichkeit von $p < 0,05$ nachgewiesen. Die Items stammen alle aus unterschiedlichen Bereichen. Daher sowie aufgrund der geringen Anzahl wird auf eine nähere Darstellung verzichtet. Es ist festzustellen, dass bis auf die

beschriebene Ausnahme kein Unterschied in der Mittelwertstendenz bei unterschiedlichem Erfahrungsstand der Mystery Shopper besteht.

4.2.1.6 Mittelwertstendenz bezüglich der Fast Food-Affinität der Mystery Shopper

Zur Ermittlung der Mittelwertstendenz bezüglich der Fast Food-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen für die einzelnen Items gegenübergestellt.⁴²¹ Das Ergebnis ist in Abbildung 4-2-1-6 dargestellt.

Abbildung 4-2-1-6: Komponente Mittelwertstendenz bezüglich Fast Food-Affinität



Quelle: Eigene Darstellung

Die Abbildung zeigt, dass sich die Mittelwertprofile bei unterschiedlicher Fast Food-Affinität der Mystery Shopper unterscheiden. Allerdings erweist sich der Unterschied nach Durchführung des H-Tests als nicht signifikant. Selbst bei Reduzierung auf drei Klassen in sehr gern/ eher gern, teils-teils sowie weniger gern/ gar nicht gern ist nur bei drei SERVPERF-Items und bei vier Kontaktpunkt-Items ein signifikanter Unterschied auszumachen. Aufgrund der geringen Anzahl soll auf eine detaillierte Darstellung verzichtet werden. Genannt werden soll jedoch das Item, das die Gesamtzufriedenheit abfragt. Bei Einteilung in die beschriebenen drei Gruppen sind die Unterschiede bei diesem Item zwischen allen drei Gruppen hoch signifikant ($p < 0,01$). Die Gruppe mit der niedrigsten Fast Food-Affinität bewertet am strengsten, die Gruppe mit der höchsten Fast Food-Affinität bewertet am mildesten. Eine Erklärung könnte sein, dass die Fast Food-Affinität der Tester auf ihr

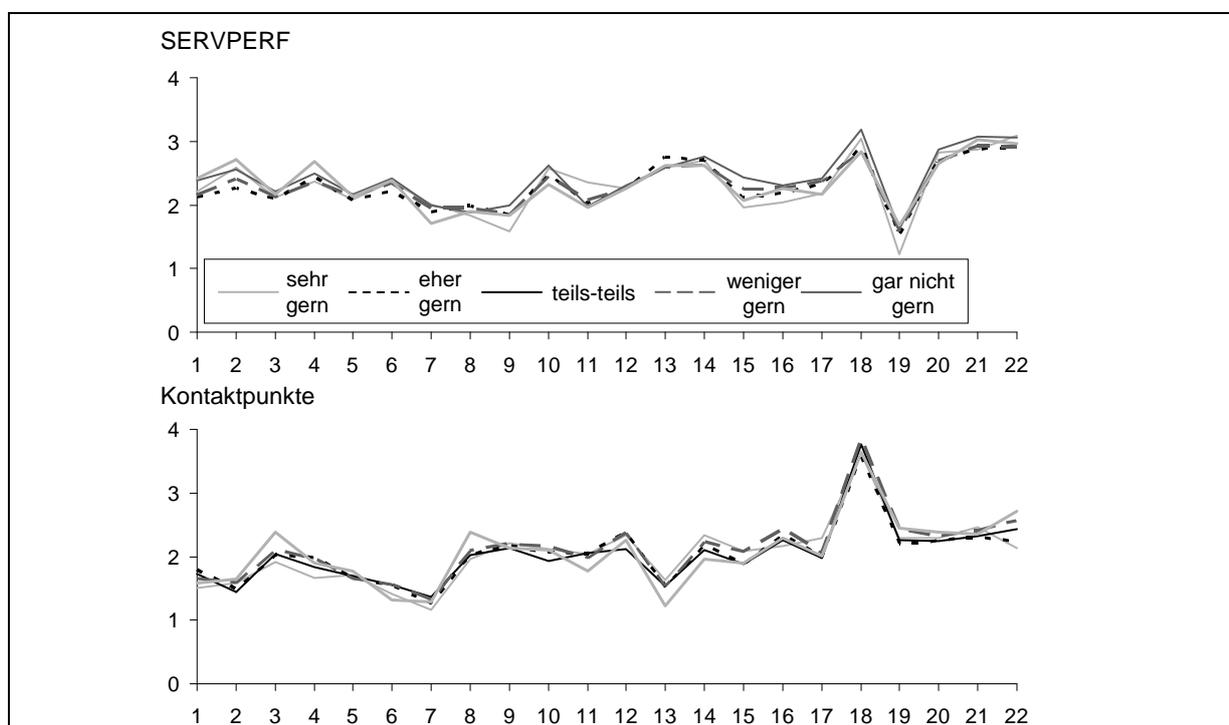
⁴²¹ Die Fast Food-Affinität wurde mit dem Item „Losgelöst von Ihrer heutigen Aufgabe als Testbesucher, wie gern essen Sie in einem Fast Food Restaurant?“ erfragt.

Gesamtzufriedenheitsurteil abstrahlt. Dieser Erklärungsversuch soll bei der Untersuchung der Korrelationstendenz wieder aufgegriffen werden. Es ist festzustellen, dass bis auf diese Ausnahme kein Unterschied in der Mittelwertstendenz bei unterschiedlich hoher Fast Food-Affinität der Mystery Shopper besteht.

4.2.1.7 Mittelwertstendenz bezüglich der Restaurant-Affinität der Mystery Shopper

Zur Ermittlung der Mittelwertstendenz bezüglich der Restaurant-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen für die einzelnen Items gegenübergestellt.⁴²² Das Ergebnis ist in Abbildung 4-2-1-7 dargestellt.

Abbildung 4-2-1-7: Komponente Mittelwertstendenz bezüglich Restaurant-Affinität



Quelle: Eigene Darstellung

Die Abbildung zeigt, dass sich die Mittelwertprofile bei unterschiedlicher Restaurant-Affinität der Mystery Shopper unterscheiden. Allerdings erweist sich der Unterschied nach Durchführung des H-Tests als nicht signifikant. Selbst bei Reduzierung auf drei Klassen in sehr gern/ eher gern, teils-teils sowie weniger gern/ gar nicht gern ist bei keinem SERVPERF-Item und bei nur einem Kontaktpunkt-Item ein signifikanter Unterschied auszumachen. Dieser signifikante Unterschied ($p < 0,01$) ist wie bei der Fast Food-Affinität bei der Abfrage der Gesamtzufriedenheit zu finden. Auch hier bewertet die Gruppe mit der niedrigsten Ausprägung am strengsten und die Gruppe mit der höchsten Ausprägung am mildesten. Folgt man dem Erklärungsansatz, der bei der Fast Food-Affinität geäußert wurde, so könnte auch

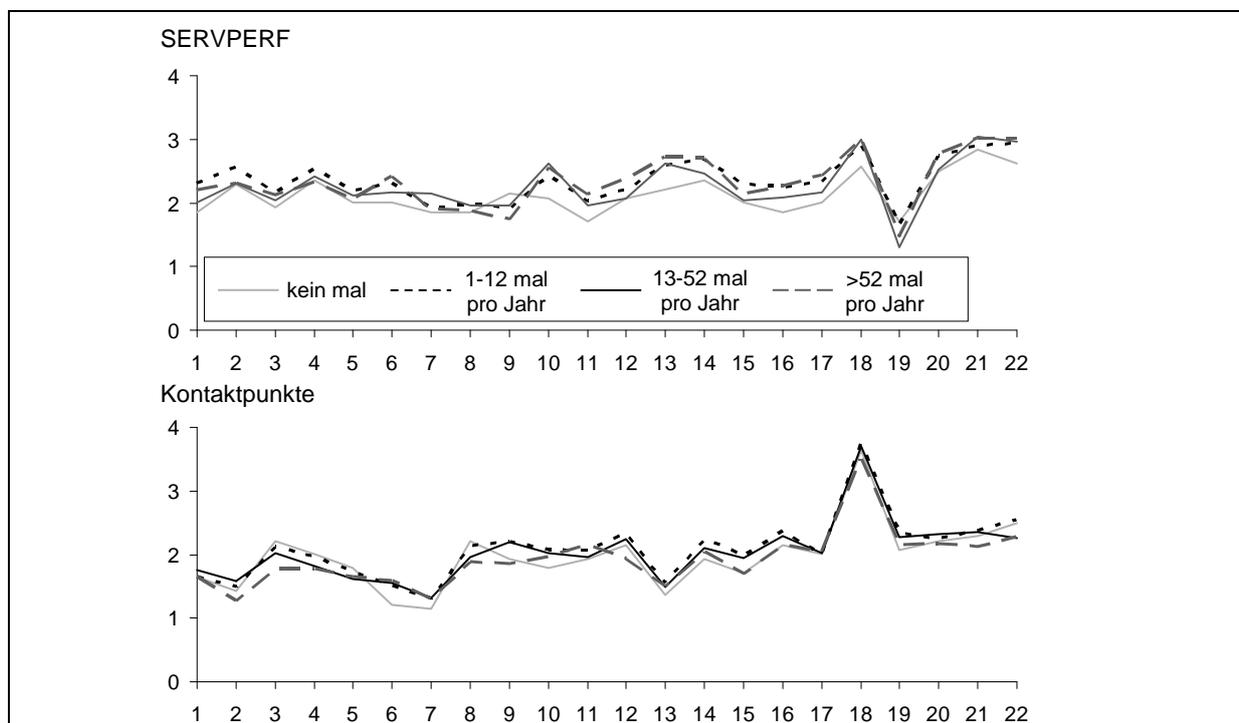
⁴²² Die Restaurant-Affinität wurde mit dem Item „Und wie gern essen Sie in diesem Fast Food-Restaurant?“ erfragt. Dabei war nicht die spezielle Filiale sondern irgendeine Filiale dieser Kette gemeint.

die Restaurant-Affinität auf das Gesamtzufriedenheitsurteil abstrahlen. Es ist festzustellen, dass bis auf diese Ausnahme kein Unterschied in der Mittelwertstendenz bei unterschiedlich hoher Restaurant-Affinität der Mystery Shopper besteht.

4.2.1.8 Mittelwertstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs

Zur Ermittlung der Komponente Mittelwertstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs der Mystery Shopper wurden unterschiedliche Häufigkeitssausprägungen bei den Items miteinander verglichen.⁴²³ Die Mittelwertprofile sind in Abbildung 4-2-1-8 dargestellt.

Abbildung 4-2-1-8: Komponente Mittelwertstendenz bezüglich Häufigkeit eines Fast Food-Restaurantbesuchs



Quelle: Eigene Darstellung

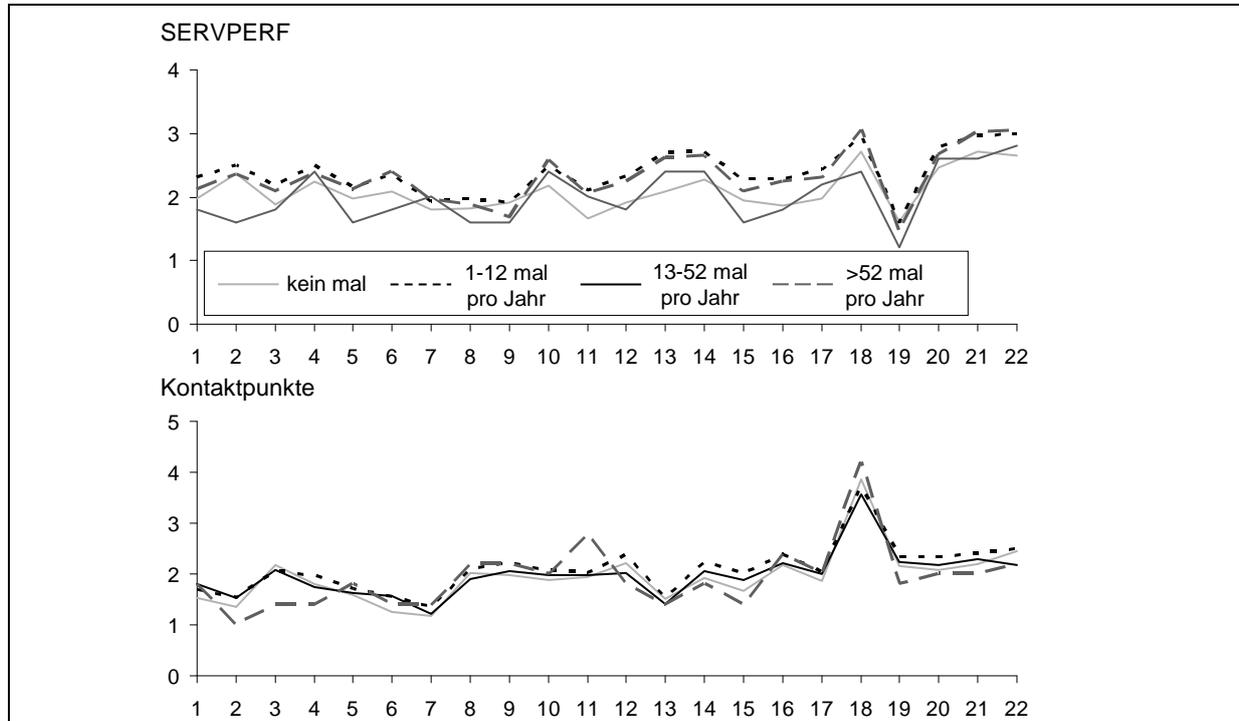
Die Unterschiedlichkeit, die anhand der grafischen Abbildung vermutet wird kann durch den H-Test nicht als signifikant bestätigt werden. Einzig das Item, welches die Gesamtzufriedenheit abfragt, weist einen Wert von $p < 0,01$ auf. Allerdings wird mittels U-Test lediglich ein signifikanter Unterschied zwischen der Besuchshäufigkeit 1-12 mal pro Jahr und 13-52 mal im Jahr von $p = 0,002$ ermittelt. Insofern besteht bis auf diese Ausnahme kein Unterschied in der Mittelwertstendenz bei unterschiedlicher Besuchshäufigkeit in einem Fast Food-Restaurant.

⁴²³ Die Häufigkeit eines Fast Food-Restaurantbesuchs wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit unterschiedlicher Fast Food-Restaurants addiert und in unterschiedlichen Klassen zusammengefasst wurden.

4.2.1.9 Mittelwertstendenz bezüglich der spezifischen Restaurant-Besuchshäufigkeit

Schließlich wurde die Komponente Mittelwertstendenz bezüglich der Häufigkeit eines Besuchs in einer Filiale der getesteten Fast Food-Kette ermittelt.⁴²⁴ Die Mittelwertprofile sind in Abbildung 4-2-1-9 dargestellt.

Abbildung 4-2-1-9: Komponente Mittelwertstendenz bezüglich spezifischer Restaurant-Besuchshäufigkeit



Quelle: Eigene Darstellung

Wie der Abbildung zu entnehmen ist, weisen Tester mit unterschiedlicher Besuchshäufigkeit in einer Filiale der getesteten Restaurantkette sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte unterschiedliche Mittelwertprofile auf. Allerdings wird die Signifikanz des Unterschieds durch den H-Test nur für sieben SERVPERF-Items und einem Kontaktpunkt-Item mit einer Irrtumswahrscheinlichkeit von $p < 0,05$ nachgewiesen. Trotz der geringen Anzahl erscheint der Unterschied interessant, denn bei den SERVPERF-Items sind drei von vier Items aus dem Bereich Assurance und alle Items aus dem Bereich Responsiveness vertreten.

⁴²⁴ Die Häufigkeit eines Besuchs in dem getesteten Restaurant wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit in dem spezifischen Fast Food-Restaurant addiert und in unterschiedlichen Klassen zusammengefasst wurden.

Tabelle 4-2-1-9: SERVPERF-Operationalisierung: Signifikante H-Werte und korrespondierender U-Werte bei der Mittelwertstendenz der spezifischen Restaurant-Besuchshäufigkeit.

	Signifikanz		
	H-Test	U-Test	
		keine Besuche vs. 1-12 mal pro Jahr	keine Besuche vs. 13-52 mal pro Jahr
Der Gast erhält von den Angestellten einen prompten Service.	0,028*	0,003**	0,027*
Die Mitarbeiter sind immer willens zu helfen.	0,014*	0,003**	0,034*
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	0,005**	0,000**	0,007**
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	0,048*	0,007**	0,080
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	0,013*	0,020**	0,465
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	0,029*	0,005**	0,038*
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	0,030*	0,004**	0,054

* Signifikanzniveau 5%, ** Signifikanzniveau 1%

Quelle: Eigene Darstellung

Wie vorstehende Tabelle 4-2-1-9 verdeutlicht, besteht ein signifikanter Unterschied zwischen der Bewertung der Gruppe der Mystery Shopper, die einer Filiale im vergangenen Jahr keine Besuche abgestattet hat und den Gruppen, die 1-12 mal (sieben signifikante Unterschiede) sowie 13-52 mal (vier signifikante Unterschiede) in einer Filiale zu Gast waren. Da die Gruppe, die keine Besuche durchgeführt hatte, bei diesen Items milder bewertet hat, kann ihr für die aufgeführten Items der Bereiche Assurance und Responsiveness eine Tendenz zur Milde unterstellt werden. Bei der Kontaktpunkt-Operationalisierung erweist sich erneut das Item, welches die Gesamtzufriedenheit abfragt, als signifikant unterschiedlich zwischen den Gruppen. Allerdings kann mittels U-Test lediglich ein signifikanter Unterschied zwischen der Besuchshäufigkeit 1-12 mal pro Jahr und 13-52 mal im Jahr von $p=0,001$ nachgewiesen werden. Insofern bestehen zum Teil Unterschiede bei der Komponente Mittelwertstendenz in Abhängigkeit von der Besuchshäufigkeit in einer Filiale der getesteten Restaurantkette.

4.2.1.10 Zusammenfassung der Ergebnisse der Mittelwertstendenz

Für die Komponente Mittelwertstendenz ist zu konstatieren, dass kein signifikanter Unterschied zwischen Männern und Frauen, zwischen Mystery Shoppern mit unterschiedlich hohem Bildungsabschluss, zwischen Mystery Shoppern mit unterschiedlichem Erfahrungsstand sowie auch nicht zwischen Mystery Shoppern, die in einem Zeitraum von einem Jahr vor Testdurchführung ein Fast Food Restaurant besucht haben, festzustellen ist. Signifikante Unterschiede wurden in fünf Bereichen festgestellt, die folgende Aussagen zulassen: Erstens weisen Mystery Shopper aus der höchsten Altersklasse insbesondere in den Bereichen Responsiveness und Empathy sowie in den Bereichen Kundengespräch und Mitarbeiter eine Tendenz zur Milde auf. Zweitens verfügen Mystery Shopper, die Vollzeit einem Beruf nachgehen im Vergleich zu Schülern und Studenten in den Bereichen Tangibles und Assurance über eine Tendenz zur Strenge. Im Vergleich zur Gruppe Hausfrau/Rentner

konnte in den Bereichen Assurance, Responsiveness, Empathy, Außenzustand und Zustand der Innenräume eine Tendenz zur Strenge festgestellt werden. Drittens weisen Mystery Shopper mit einer hohen Fast Food-Affinität im Gesamtzufriedenheitsurteil eine Tendenz zur Milde und mit einer niedrigen Affinität eine Tendenz zur Strenge auf. Diese Feststellung konnte viertens analog für die spezifische Restaurant-Affinität festgestellt werden. Fünftens verfügen Mystery Shopper, die einer Filiale der getesteten Restaurantkette im vergangenen Jahr keinen Besuch abgestattet haben, über eine Tendenz zur Milde in den Bereichen Assurance und Responsiveness.

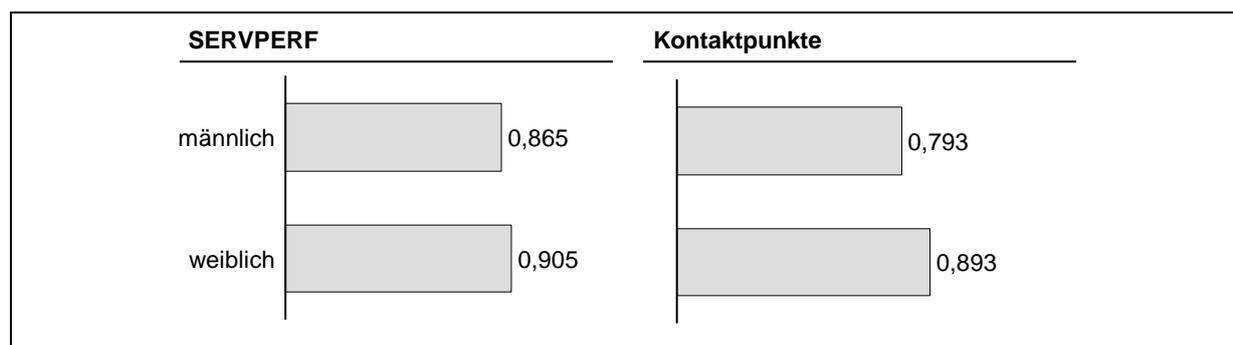
4.2.2 Streuungstendenz von Mystery Shoppern

In diesem Abschnitt werden die Ergebnisse der Komponente Streuungstendenz dargestellt. Dabei werden die Werte der Ausprägungen der einzelnen Merkmalsgruppen jeweils separat für die Operationalisierung der SERVPERF-Dimensionen⁴²⁵ und die Operationalisierung der Kontaktpunkte wiedergegeben⁴²⁶. Soweit es möglich ist, wird versucht, übersichtliche Grafiken für die Ausprägungen zu zeigen. Eine itemspezifische tabellarische Übersicht der ermittelten Werte für Mittelwert, Standardabweichung und Varianz findet sich im Anhang.

4.2.2.1 Streuungstendenz bezüglich des Geschlechtes der Mystery Shopper

Zur Ermittlung der Komponente Streuungstendenz bezüglich des Geschlechts der Mystery Shopper wurden Frauen und Männer hinsichtlich ihrer durchschnittlichen Standardabweichung gegenübergestellt. Das Ergebnis ist in Abbildung 4-2-2-1 dargestellt.

Abbildung 4-2-2-1: Durchschnittliche Standardabweichung (Geschlecht)



Quelle: Eigene Darstellung

Der aus der Abbildung anzunehmende Unterschied bestätigt sich nach Durchführung des H-Tests für keins der Items als signifikant. Insofern unterscheiden sich Männer und Frauen nicht signifikant hinsichtlich ihrer Streuungstendenz.

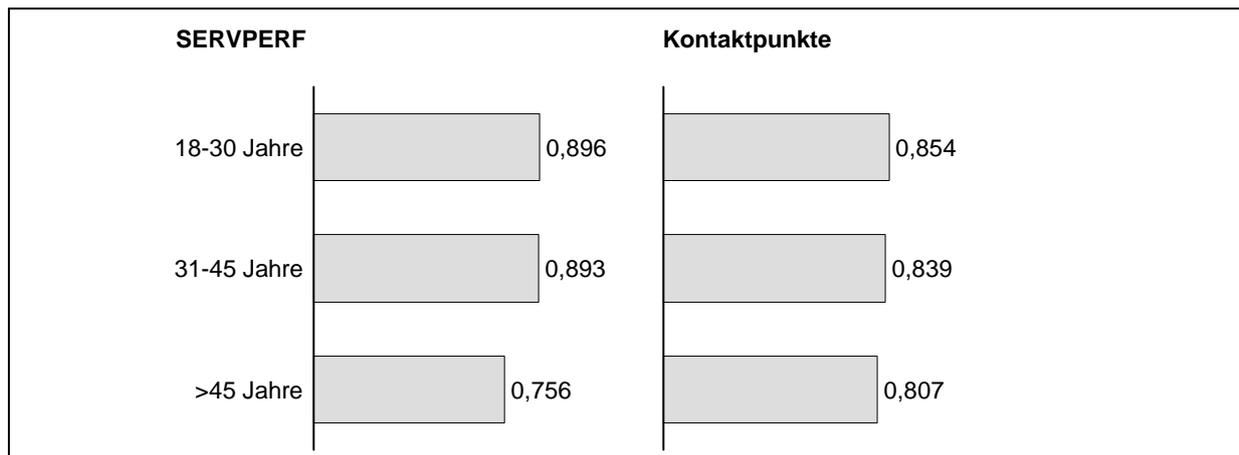
⁴²⁵ Vgl. Tabelle 3-2-3

⁴²⁶ Vgl. Tabelle 3-2-4

4.2.2.2 Streuungstendenz bezüglich des Alters der Mystery Shopper

Basierend auf den Altersangaben der Mystery Shopper wurden drei Altersklassen für die Auswertung der Komponente Streuungstendenz gebildet.⁴²⁷ Die Ergebnisse der Gegenüberstellung der durchschnittlichen Standardabweichungen der unterschiedlichen Altersklassen sind in Abbildung 4-2-2-2 dargestellt.

Abbildung 4-2-2-2: Durchschnittliche Standardabweichung (Altersklassen)



Quelle: Eigene Darstellung

Grafisch gesehen weisen die unterschiedlichen Altersklassen sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte ein unterschiedliches Streuungsprofil auf. Die Signifikanz des Unterschieds wird für einen Teil der Items durch den H-Test bestätigt, wie in den Tabellen 4-2-1-2-1 und 4-2-1-2-2 im vorangegangenen Abschnitt zum Thema Mittelwertstendenz gezeigt wurde. Sofern eine Signifikanz nach dem H-Test festgestellt wurde, wurden die Klassen mit dem U-Test paarweise auf signifikante Unterschiede geprüft. Bei 14 von 22 SERVPERF-Operationalisierungen wurde mittels H-Test ein signifikanter Unterschied zwischen den Altersklassen nachgewiesen, wobei 13 Items eine Irrtumswahrscheinlichkeit von $p < 0,01$ aufweisen. Vergleicht man die Altersklassen untereinander mit dem U-Test wird die Signifikanz des Unterschieds zwischen den einzelnen Klassen deutlicher: Zwischen der Altersklasse der 18-30jährigen Mystery Shopper und der nächst höheren Altersklasse ist nur in einem Fall ein signifikanter Unterschied mit $p < 0,05$ nachzuweisen. Die 18-30jährigen und die Mystery Shopper älter als 45 Jahre weisen bis auf drei Items sowie die 31-45jährigen und die >45 Jahre alten Mystery Shopper bei allen 14 Items signifikante bis hoch signifikante Unterschiede auf. Da die Standardabweichung der Tester >45 Jahre bei 13 von 14 Items den jeweils geringsten Werte aufweist, ist dieser Altersklasse im Vergleich zu den anderen beiden für diese Items eine höhere Einheitlichkeit im Urteil zu unterstellen.

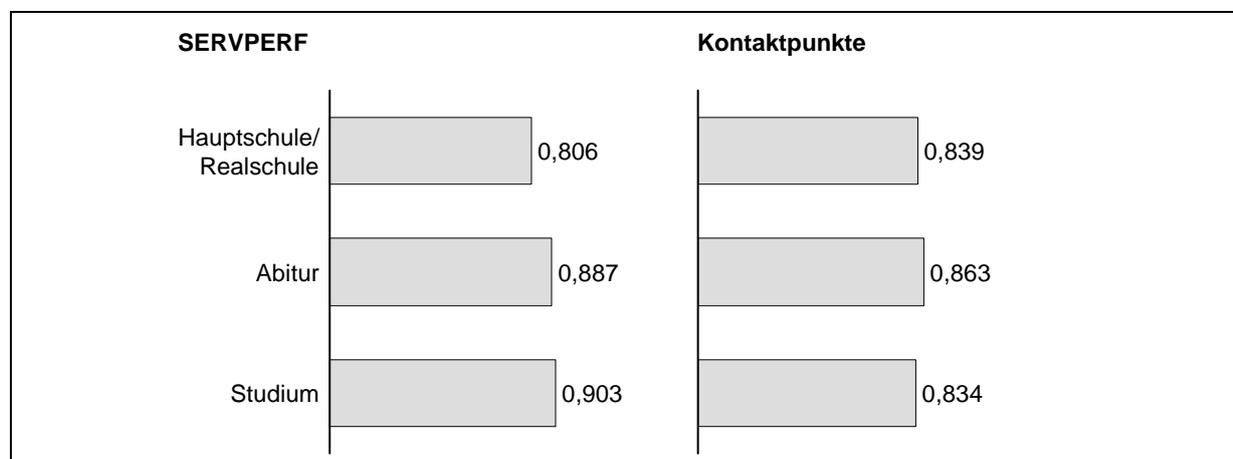
⁴²⁷ Die ursprüngliche Einteilung in vier Altersklassen erschien bei näherer Betrachtung der Fallzahlen nicht sinnvoll. Siehe Tabelle 3-3-4 für die Anzahl der Fälle bei einer Aufteilung in vier Altersklassen.

Bei der Hälfte der 22 Kontaktpunkt-Operationalisierungen wurde mittels H-Test ein signifikanter oder hoch signifikanter Unterschied der Altersklassen nachgewiesen. Vergleicht man die Signifikanz des Unterschieds bei den Altersklassen mit dem U-Test lässt sich zwischen der Altersklasse der 18-30jährigen und der 31-45jährigen bei zwei Items ein signifikanter Unterschied mit $p < 0,05$ nachweisen. Die 18-30jährigen und die Mystery Shopper älter als 45 Jahre weisen bei vier Items einen signifikanten Unterschied auf. Die beiden hohen Altersklassen unterscheiden sich bis auf eines der elf Items signifikant bis hoch signifikant. Allerdings ist kein Muster bei der Streuung zu erkennen. Keine Altersklasse überwiegt gegenüber einer anderen Klasse mit höheren oder niedrigeren Standardabweichungen bei den signifikanten Items. Insofern ist bei den Kontaktpunkt-Operationalisierungen zwar zum Teil ein Unterschied in der Streuungstendenz zu unterstellen, gleichwohl ist im Gegensatz zu den SERVPERF-Operationalisierungen keine Rangfolge zu unterstellen.

4.2.2.3 Streuungstendenz bezüglich des höchsten Bildungsabschlusses der Mystery Shopper

Die durchschnittliche Standardabweichung für die drei Bildungsklassen ist in Abbildung 4-2-2-3 dargestellt.

Abbildung 4-2-2-3: Durchschnittliche Standardabweichung (höchster Bildungsabschluss)



Quelle: Eigene Darstellung

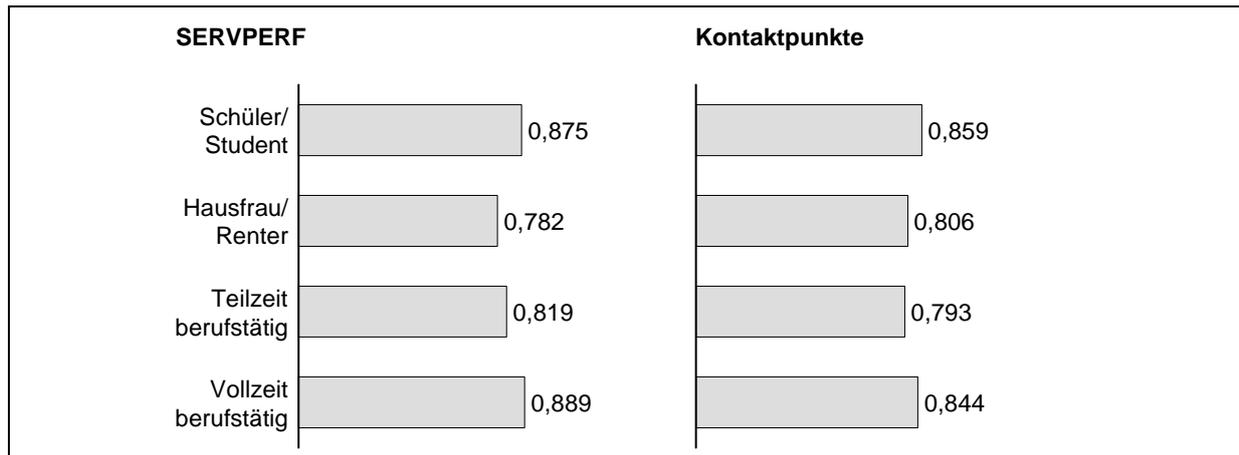
Die drei Bildungsklassen weisen sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte eine unterschiedliche Streuungstendenz auf. Die Signifikanz des Unterschieds wird für ein fünf SERVPERF-Items und für ein Kontaktpunkt-Item durch den H-Test bestätigt. Nach Durchführung von paarweisen U-Tests zum Vergleich der Bildungsklassen zeigt sich bei vier Items zwischen Hauptschule/Realschule und Abitur sowie bei drei Items zwischen Abitur und Studium ein hoch signifikanter Unterschied. Auf eine detaillierte Darstellung wird an dieser Stelle verzichtet, da es sich um lediglich vier von 44 Items handelt und diese Anzahl als unbedeutend angesehen wird. Es ist festzustellen, dass

bis auf diese Ausnahme kein Unterschied in der Streuungstendenz bei unterschiedlich hohem Bildungsabschluss der Mystery Shopper besteht.

4.2.2.4 Streuungstendenz bezüglich der Tätigkeit der Mystery Shopper

Nachfolgende Abbildung 4-2-2-4 stellt die durchschnittliche Standardabweichung der unterschiedlichen Tätigkeitsgruppen gegenüber.

Abbildung 4-2-2-4: Durchschnittliche Standardabweichung (Tätigkeit)



Quelle: Eigene Darstellung

Die vier Tätigkeitsgruppen weisen sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte unterschiedliche durchschnittliche Standardabweichungen auf. Die Signifikanz des Unterschieds wird bei 18 der 22 SERVPERF-Items und bei acht der 22 Kontaktpunkt-Items durch den H-Test bestätigt.⁴²⁸ Sofern eine Signifikanz nach dem H-Test festgestellt wurde, wurde sukzessive paarweise ein U-Test durchgeführt.

Signifikante Unterschiede sind bei der SERVPERF-Operationalisierung mit 15 Items insbesondere zwischen den Gruppen Hausfrau/Rentner und Vollzeit berufstätig auszumachen. Die Gruppen Schüler/Student und Vollzeit berufstätig (Hausfrau/Rentner) unterscheiden sich bei neun (acht) Items signifikant voneinander. Für die Gruppe Hausfrau/Rentner ist im Vergleich zur Gruppe Vollzeit berufstätig bei 13 der 15 signifikant unterschiedlichen Items eine niedrigere Streuung festzustellen, und ihr kann daher eine vergleichsweise einheitlichere Bewertung unterstellt werden. Dies tritt bei drei von vier Items aus dem Bereich Responsiveness, bei drei von vier Items aus dem Bereich Assurance und bei vier von fünf Items aus dem Bereich Empathy auf. Bei sechs von acht signifikant unterschiedlichen Items zwischen den Gruppen Hausfrau/Rentner und Schüler/Student ist der Gruppe Hausfrau/Rentner aufgrund der vergleichsweise niedrigeren Streuung ebenfalls eine einheitlichere Bewertung für diese Items zu unterstellen. Herauszustellen ist der Bereich

⁴²⁸ Siehe Tabellen 4-2-1-4-1 und 4-2-1-4-2 in Abschnitt 4.2.1.4

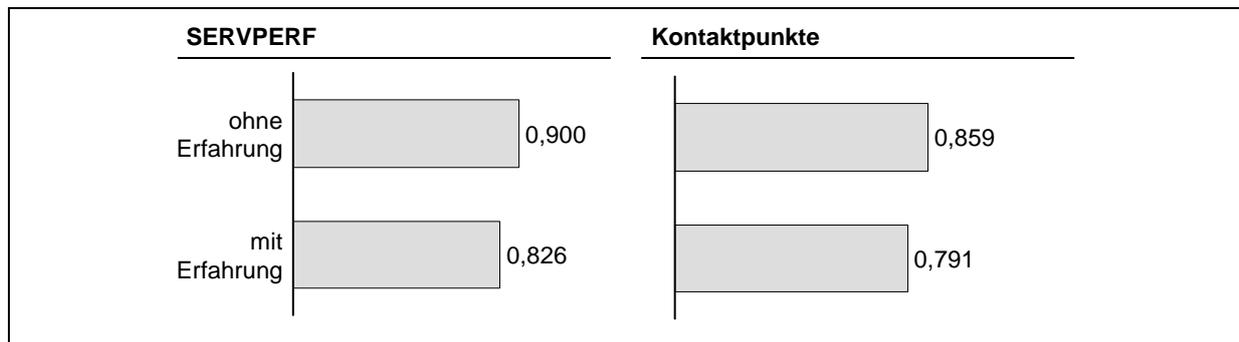
Responsiveness, der mit drei von vier Items vertreten ist. Im direkten Vergleich der Gruppen Schüler/Student und Vollzeit berufstätig ist die Streuung bei der Gruppe Schüler/Student in vier (fünf) von neun Items höher (niedriger). Es ist keine einheitliche Häufung bei bereichsspezifischen Items auszumachen. Insofern ist eine Aussage über einen Vergleich dieser beiden Gruppen hinsichtlich ihrer Streuungstendenz nicht eindeutig.

Die Signifikanz des Unterschieds zwischen den Gruppen Vollzeit berufstätig und Hausfrau/Rentner spiegeln auch die Ergebnisse der Kontaktpunkt-Analyse wieder (sieben Items mit $p < 0,05$). Allerdings sind die Unterschiede eher bei tangiblen Bewertungsobjekten, wie dem Sauberkeitszustand zu finden. Bei sechs Items ist der Gruppe Hausfrau/Rentner eine vergleichsweise einheitlichere Bewertung zu unterstellen.

4.2.2.5 Streuungstendenz bezüglich der Erfahrung als Mystery Shopper

Zur Beurteilung der Streuungstendenz bezüglich der Erfahrung als Mystery Shopper wurden zwei Gruppen gebildet: Solche mit Einsatzerfahrung als Tester und solche ohne. Die durchschnittliche Streuungstendenz der beiden Gruppen wird in der Abbildung 4-2-2-5 grafisch veranschaulicht.

Abbildung 4-2-2-5: Durchschnittliche Standardabweichung (Erfahrung als Mystery Shopper)



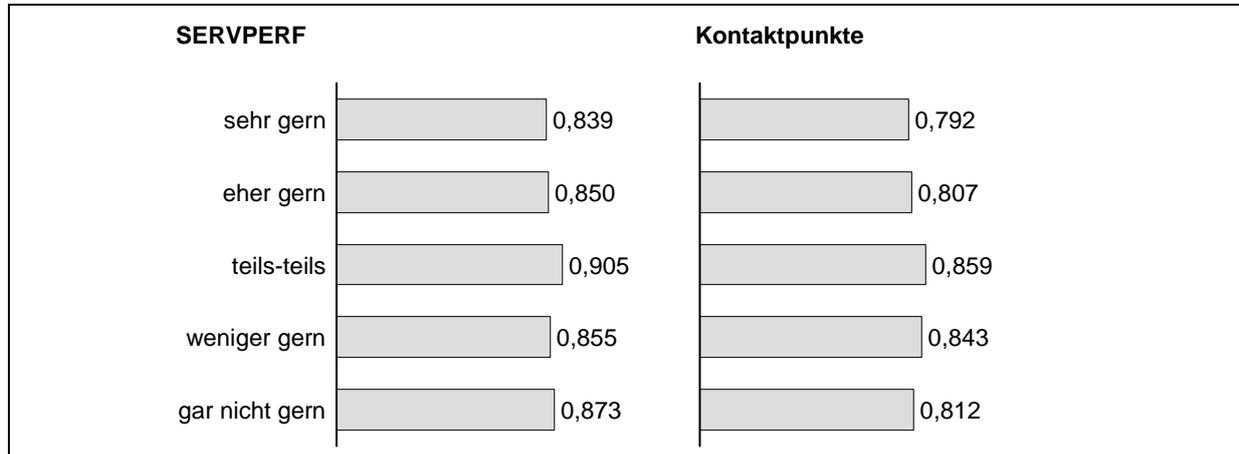
Quelle: Eigene Darstellung

Wie der Abbildung zu entnehmen ist, weisen Tester mit und ohne Mystery Shopping-Erfahrung sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte eine unterschiedliche durchschnittliche Standardabweichung auf. Allerdings wird die Signifikanz des Unterschieds durch den H-Test nur für sechs SERVPERF-Items und vier Kontaktpunktitems mit einer Irrtumswahrscheinlichkeit von $p < 0,05$ nachgewiesen. Die Items stammen alle aus unterschiedlichen Bereichen. Daher sowie aufgrund der geringen Anzahl wird auf eine nähere Darstellung verzichtet. Es ist festzustellen, dass bis auf die beschriebene Ausnahme kein Unterschied in der Streuungstendenz bei unterschiedlichem Erfahrungsstand der Mystery Shopper besteht.

4.2.2.6 Streuungstendenz bezüglich der Fast Food-Affinität der Mystery Shopper

Zur Ermittlung der Streuungstendenz bezüglich der Fast Food-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen für die einzelnen Items gegenübergestellt.⁴²⁹ Abbildung 4-2-2-6 gibt die durchschnittliche Standardabweichung der Gruppen wieder.

Abbildung 4-2-2-6: Durchschnittliche Standardabweichung (Fast Food-Affinität)



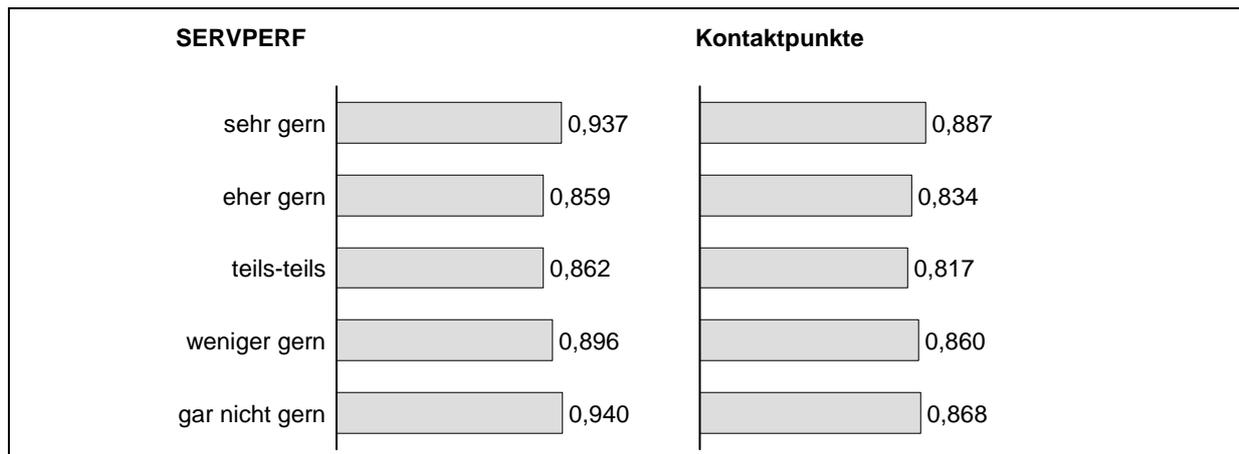
Quelle: Eigene Darstellung

Die Abbildung zeigt, dass sich die durchschnittliche Standardabweichung bei unterschiedlicher Fast Food-Affinität der Mystery Shopper unterscheidet. Allerdings erweist sich der Unterschied nach Durchführung des H-Tests als nicht signifikant. Selbst bei Reduzierung auf drei Klassen in sehr gern/ eher gern, teils-teils sowie weniger gern/ gar nicht gern ist nur bei drei SERVPERF-Items und bei vier Kontaktpunkt-Items ein signifikanter Unterschied auszumachen. Aufgrund der geringen Anzahl soll auf eine detaillierte Darstellung verzichtet werden. Genannt werden soll jedoch das Item, das die Gesamtzufriedenheit abfragt. Bei Einteilung in die beschriebenen drei Gruppen sind die Unterschiede bei diesem Item zwischen allen drei Gruppen signifikant ($p < 0,05$). Die Gruppe mit der niedrigsten Fast Food-Affinität weist dabei die höchste Standardabweichung und die Gruppe mit der höchsten Affinität die niedrigste Standardabweichung auf. Es ist festzustellen, dass bis auf die Ausnahme einer zunehmenden Streuungstendenz bei abnehmender Affinität beim Gesamtzufriedenheitsurteil kein Unterschied in der Streuungstendenz bei unterschiedlich hoher Fast Food-Affinität der Mystery Shopper besteht.

4.2.2.7 Streuungstendenz bezüglich der Restaurant-Affinität der Mystery Shopper

Zur Ermittlung der Streuungstendenz bezüglich der Restaurant-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen für die einzelnen Items gegenübergestellt.⁴³⁰ Die durchschnittliche Standardabweichung der Gruppen ist in Abbildung 4-2-2-7 dargestellt.

⁴²⁹ Die Fast Food-Affinität wurde mit dem Item „Losgelöst von Ihrer heutigen Aufgabe als Testbesucher, wie gern essen Sie in einem Fast Food Restaurant?“ erfragt.

Abbildung 4-2-2-7: Durchschnittliche Standardabweichung (Restaurant-Affinität)

Quelle: Eigene Darstellung

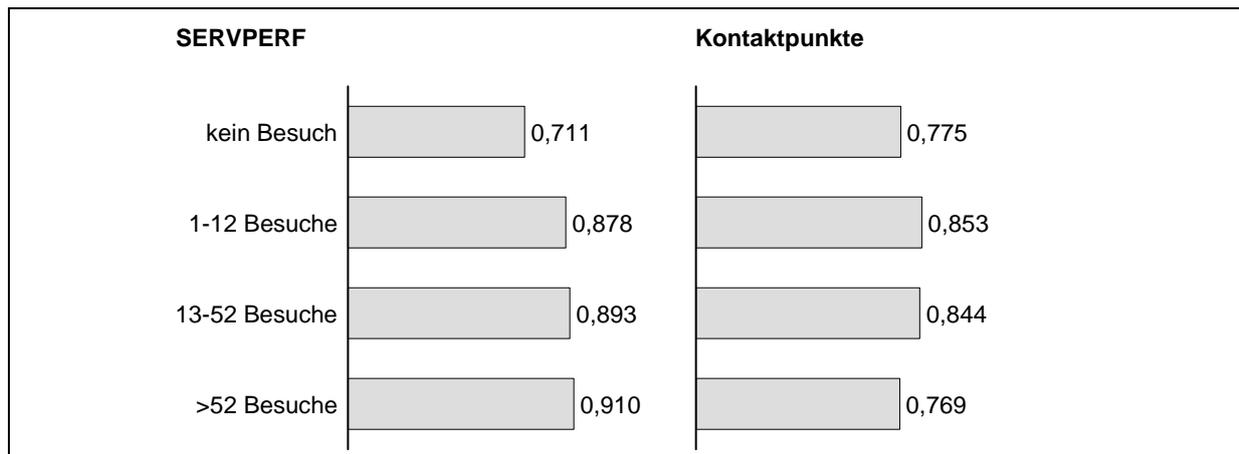
Die Abbildung zeigt, dass sich die durchschnittliche Standardabweichung bei unterschiedlicher Restaurant-Affinität der Mystery Shopper unterscheidet. Allerdings erweist sich der Unterschied nach Durchführung des H-Tests als nicht signifikant. Selbst bei Reduzierung auf drei Klassen in sehr gern/ eher gern, teils-teils sowie weniger gern/ gar nicht gern ist bei keinem SERVPERF-Item und bei nur einem Kontaktpunkt-Item ein signifikanter Unterschied auszumachen. Dieser signifikante Unterschied ($p < 0,01$) ist wie bei der Fast Food-Affinität bei der Abfrage der Gesamtzufriedenheit zu finden. Auch hier weist die Gruppe mit der höchsten Affinitätsausprägung die niedrigste Streuungstendenz und die Gruppe mit der höchsten Ausprägung die niedrigste auf. Es ist festzustellen, dass bis auf diese Ausnahme kein Unterschied in der Streuungstendenz bei unterschiedlich hoher Restaurant-Affinität der Mystery Shopper besteht.

4.2.2.8 Streuungstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs

Zur Ermittlung der Komponente Streuungstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs der Mystery Shopper wurden unterschiedliche Häufigkeitssausprägungen bei den Items miteinander verglichen.⁴³¹ Die durchschnittliche Standardabweichung der einzelnen Gruppen ist in Abbildung 4-2-2-8 dargestellt.

⁴³⁰ Die Restaurant-Affinität wurde mit dem Item „Und wie gern essen Sie in diesem Fast Food-Restaurant?“ erfragt. Dabei war nicht die spezielle Filiale sondern irgendeine Filiale dieser Kette gemeint.

⁴³¹ Die Häufigkeit eines Fast Food-Restaurantbesuchs wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit unterschiedlicher Fast Food-Restaurants addiert und in unterschiedlichen Klassen zusammengefasst wurden.

Abbildung 4-2-2-8: Durchschnittliche Standardabweichung (Häufigkeit eines Fast Food-Restaurantbesuchs)

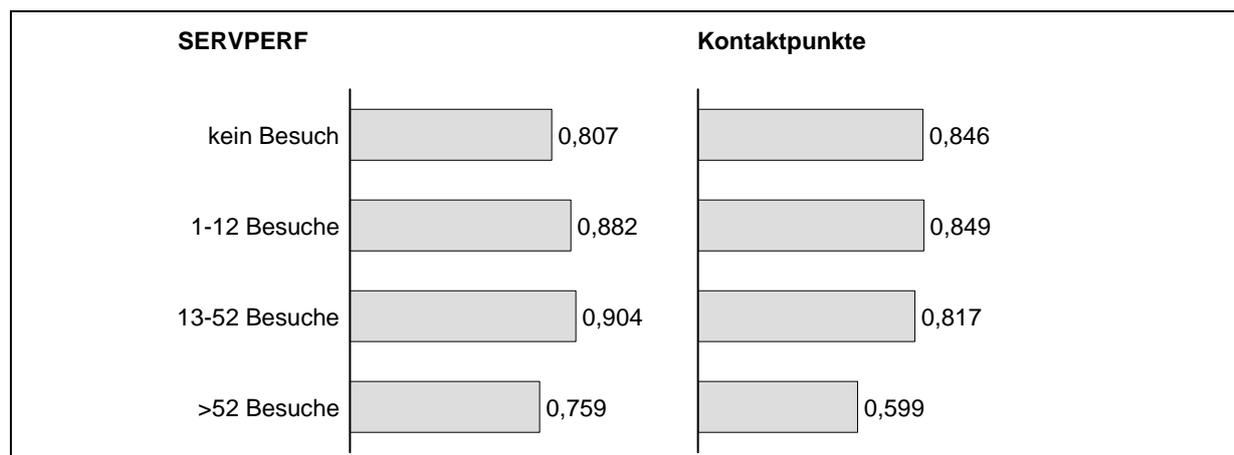
Quelle: Eigene Darstellung

Die Unterschiedlichkeit, die anhand der grafischen Abbildung vermutet wird kann durch den H-Test nicht als signifikant bestätigt werden. Einzig das Item, welches die Gesamtzufriedenheit abfragt, weist einen Wert von $p < 0,01$ auf. Allerdings wird mittels U-Test lediglich ein signifikanter Unterschied zwischen der Besuchshäufigkeit 1-12 mal pro Jahr und 13-52 mal im Jahr von $p = 0,002$ ermittelt. Insofern besteht bis auf diese Ausnahme kein Unterschied in der Streuungstendenz bei unterschiedlicher Besuchshäufigkeit in einem Fast Food-Restaurant.

4.2.2.9 Streuungstendenz bezüglich der spezifischen Restaurant-Besuchshäufigkeit

Schließlich wurde die Komponente Streuungstendenz bezüglich der Häufigkeit eines Besuchs in einer Filiale der getesteten Fast Food-Kette ermittelt.⁴³² Die durchschnittliche Standardabweichung der einzelnen Gruppen ist in Abbildung 4-2-2-9 dargestellt.

⁴³² Die Häufigkeit eines Besuchs in dem getesteten Restaurant wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit in dem spezifischen Fast Food-Restaurant addiert und in unterschiedlichen Klassen zusammengefasst wurden.

Abbildung 4-2-2-9: Durchschnittliche Standardabweichung (Restaurant-Besuchshäufigkeit)

Quelle: Eigene Darstellung

Wie der Abbildung zu entnehmen ist, weisen Tester mit unterschiedlicher Besuchshäufigkeit in einer Filiale der getesteten Restaurantkette sowohl bei der SERVPERF-Operationalisierung als auch bei der Operationalisierung der Kontaktpunkte eine unterschiedliche durchschnittliche Standardabweichung auf. Allerdings wird die Signifikanz des Unterschieds durch den H-Test nur für sieben SERVPERF-Items und einem Kontaktpunkt-Item mit einer Irrtumswahrscheinlichkeit von $p < 0,05$ nachgewiesen. Trotz der geringen Anzahl erscheint der Unterschied interessant, denn bei den SERVPERF-Items sind drei von vier Items aus dem Bereich Assurance und alle Items aus dem Bereich Responsiveness vertreten.

Wie Tabelle 4-2-1-9⁴³³ verdeutlicht, besteht bei den SERVPERF-Operationalisierungen ein signifikanter Unterschied zwischen der Bewertung der Gruppe der Mystery Shopper, die einer Filiale im vergangenen Jahr keine Besuche abgestattet hat und den Gruppen, die 1-12 mal (sieben signifikante Unterschiede) sowie 13-52 mal (vier signifikante Unterschiede) in einer Filiale zu Gast waren. Da die Gruppe, die keine Besuche durchgeführt hatte, bei diesen Items die geringste Streuungstendenz aufweist, kann ihr für die aufgeführten Items der Bereiche Assurance und Responsiveness eine Tendenz zur Einheitlichkeit unterstellt werden. Bei der Kontaktpunkt-Operationalisierung erweist sich erneut das Item, welches die Gesamtzufriedenheit abfragt, als signifikant unterschiedlich zwischen den Gruppen. Allerdings kann mittels U-Test lediglich ein signifikanter Unterschied zwischen der Besuchshäufigkeit 1-12 mal pro Jahr und 13-52 mal im Jahr von $p = 0,001$ nachgewiesen werden. Insofern bestehen zum Teil Unterschiede bei der Komponente Streuungstendenz in Abhängigkeit von der Besuchshäufigkeit in einer Filiale der getesteten Restaurantkette.

⁴³³ siehe Abschnitt 4.2.1.9

4.2.2.10 Zusammenfassung der Ergebnisse der Streuungstendenz

Für die Komponente Streuungstendenz ist zu konstatieren, dass kein signifikanter Unterschied zwischen Männern und Frauen, zwischen Mystery Shoppern mit unterschiedlich hohem Bildungsabschluss, zwischen Mystery Shoppern mit unterschiedlichem Erfahrungsstand sowie auch nicht zwischen Mystery Shoppern, die in einem Zeitraum von einem Jahr vor Testdurchführung ein Fast Food Restaurant besucht haben, festzustellen ist. Signifikante Unterschiede wurden in fünf Bereichen festgestellt, die folgende Aussagen zulassen: Erstens weisen Mystery Shopper aus der höchsten Altersklasse insbesondere in den Bereichen Responsiveness und Empathy eine Tendenz zur Einheitlichkeit auf. Zweitens verfügen Mystery Shopper aus der Gruppe Hausfrau/Rentner im Vergleich zur Gruppe Vollzeit beschäftigt über eine einheitlichere Bewertung in den Bereichen Responsiveness, Assurance, Empathy und Zustand der Innenräume sowie im Vergleich zur Gruppe der Schüler und Studenten im Bereich Responsiveness. Drittens weisen Mystery Shopper mit einer hohen Fast Food-Affinität im Gesamtzufriedenheitsurteil eine Tendenz zur Einheitlichkeit und mit einer niedrigen Affinität eine Tendenz zur Uneinheitlichkeit auf. Diese Feststellung konnte viertens analog für die spezifische Restaurant-Affinität festgestellt werden. Fünftens verfügen Mystery Shopper, die einer Filiale der getesteten Restaurantkette im vergangenen Jahr keinen Besuch abgestattet haben, über eine Tendenz zur Einheitlichkeit in den Bereichen Assurance und Responsiveness.

4.2.3 Korrelationstendenz von Mystery Shoppern

Als dritte Urteilstendenz und dritte Komponente zur Ermittlung der Ergebnisqualität soll in diesem Abschnitt die Korrelationstendenz näher betrachtet werden. Dafür werden die Pfadkoeffizienten des Strukturmodells für die einzelnen Testkundengruppen analog der Abschnitte zur Mittelwerts- und Streuungstendenz gegenübergestellt und getrennt für die Operationalisierung der SERVPERF-Dimensionen⁴³⁴ und die Operationalisierung der Kontaktpunkte wiedergegeben⁴³⁵.

4.2.3.1 Korrelationstendenz bezüglich des Geschlechtes der Mystery Shopper

Zur Ermittlung der Komponente Korrelationstendenz bezüglich des Geschlechtes der Mystery Shopper wurden die Pfadkoeffizienten von Männern und Frauen jeweils für die SERVPERF- (siehe Tabelle 4-2-3-1-1) und die Kontaktpunkt-Operationalisierung (siehe Tabelle 4-2-3-1-2) verglichen.

⁴³⁴ Vgl. Tabelle 3-2-3

⁴³⁵ Vgl. Tabelle 3-2-4

Tabelle 4-2-3-1-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Geschlecht)

	Männer			Frauen			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Tangibles	0,392	4,683**	0,071	0,320	3,548**	0,084	0,072	0,661	0,255	nein
Reliability	0,305	3,279**	0,073	0,056	0,548	0,077	0,249			ja
Responsiveness	-0,088	0,961	0,083	0,149	1,574	0,095	-0,237			nein
Assurance	0,300	3,229**	0,088	0,201	1,926	0,096	0,099			ja
Empathy	0,042	0,775	0,059	0,165	2,000*	0,080	-0,123			ja
R ²	0,476			0,468						
Fälle	193			183						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der SERVPERF-Operationalisierung lässt sich die Gesamtzufriedenheit sowohl bei den Männern als auch bei den Frauen zu knapp 50% erklären und ist zwischen „durchschnittlich“ und „substanziell“ einzustufen. Während drei der fünf Beziehungen im Strukturmodell bei den Männern signifikant sind, sind es bei den Frauen nur zwei. Bei beiden Gruppen ist Tangibles die stärkste exogene latente Variable. Der Unterschied bei den Pfadkoeffizienten resultiert bei den Variablen Reliability, Assurance und Empathy aus einer signifikanten Beziehung der einen Gruppe und einer insignifikanten Beziehung der anderen Gruppe. Die Variablen Reliability und Assurance haben bei den Männern einen signifikant positiven Einfluss auf die Gesamtzufriedenheit und die Variable Assurance bei den Frauen. Bei der Variable Tangibles ist kein signifikanter Unterschied auszumachen. Demnach besteht ein Unterschied bei der Korrelationstendenz zwischen Männern und Frauen bei der SERVPERF-Operationalisierung.

Tabelle 4-2-3-1-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Geschlecht)

	Männer			Frauen			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Außenzustand	0,187	3,595**	0,052	0,191	4,071**	0,047	-0,004	0,054	0,478	nein
Innenzustand	0,176	2,689**	0,066	0,218	4,378**	0,050	-0,042	0,505	0,307	nein
Kundengespräch	0,292	4,418**	0,066	0,330	4,484**	0,074	-0,039	0,393	0,347	nein
Mitarbeiter	0,427	7,533**	0,057	0,351	5,056**	0,069	0,076	0,857	0,196	nein
R ²	0,743			0,763						
Fälle	193			183						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Gesamtzufriedenheit sowohl bei den Männern als auch bei den Frauen zu rund 75% erklären und ist als „substanziell“ einzustufen. Es ist kein signifikanter Unterschied zwischen den Pfadkoeffizienten zwischen Männern und

Frauen auszumachen. Dies spiegelt auch die geringe Differenz zwischen den Pfadkoeffizienten der Geschlechter wieder.

4.2.3.2 Korrelationstendenz bezüglich des Alters der Mystery Shopper

Basierend auf den Altersangaben der Mystery Shopper wurden drei Altersklassen für die Auswertung der Komponente Korrelationstendenz gebildet. Die Tabellen 4-2-3-2-1 und 4-2-3-2-2 geben die Pfadkoeffizienten und ihre Unterschiede für die SERVPERF- und Kontaktpunkt-Operationalisierungen wieder.

Tabelle 4-2-3-2-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Altersklassen)

	18-30 Jahre			31-45 Jahre			>45 Jahre		
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD
Tangibles	0,346	3,148**	0,110	0,248	3,456**	0,072	0,378	3,179**	0,119
Reliability	-0,040	0,429	0,093	0,296	3,968**	0,075	-0,021	0,170	0,124
Responsiveness	0,083	0,758	0,110	-0,041	0,475	0,087	0,041	0,329	0,126
Assurance	0,327	2,669**	0,122	0,283	3,550**	0,080	0,236	1,613	0,146
Empathy	0,119	1,036	0,115	0,063	0,901	0,070	0,240	1,941	0,124
R ²	0,477			0,499			0,560		
Fälle	102			206			68		

Unterschiede:

	18-30 vs. 31-45 Jahre			18-30 vs. >45 Jahre			31-45 vs. >45 Jahre		
	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p
Tangibles	0,098	0,7670	0,222	-0,032	0,1908	0,425	-0,129	0,9119	0,181
Reliability									
Responsiveness									
Assurance	0,044	0,3099	0,378						
Empathy									

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Die Erklärung der Gesamtzufriedenheit ist bei der SERVPERF-Operationalisierung in allen drei Altersklassen zwischen „durchschnittlich“ und „substanziell“ einzustufen. Während bei der höchsten Altersklasse nur eine der fünf Beziehungen im Strukturmodell signifikant ist, sind es bei der niedrigsten zwei und bei der mittleren drei. Die mittlere Altersklasse hat bei der Variable Reliability einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Die niedrigste und die mittlere Altersklasse haben bei der Variable Assurance einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, den die höchste Altersklasse nicht hat. Allerdings ist der Unterschied zwischen niedrigster und mittlerer Altersklasse nicht signifikant verschieden. Insofern besteht bei der SERVPERF-Operationalisierung lediglich ein geringer Unterschied hinsichtlich der Korrelationstendenz zwischen den Altersklassen.

Tabelle 4-2-3-2-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Altersklassen)

	18-30 Jahre			31-45 Jahre			>45 Jahre		
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD
Außenzustand	0,167	2,854**	0,059	0,222	4,404**	0,050	0,142	1,847	0,077
Innenzustand	0,282	3,738**	0,075	0,179	3,793**	0,047	0,182	2,294**	0,080
Kundengespräch	0,329	3,497**	0,094	0,321	4,876**	0,066	0,272	2,915**	0,093
Mitarbeiter	0,291	3,792**	0,077	0,393	6,067**	0,065	0,426	3,911**	0,109
R ²	0,747			0,758			0,761		
Fälle	102			206			68		

Unterschiede:

	18-30 vs. 31-45 Jahre			18-30 vs. >45 Jahre			31-45 vs. >45 Jahre		
	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p
Außenzustand	-0,055	0,6637	0,254						
Innenzustand	0,103	1,205	0,115	0,100	0,8864	0,188	-0,003	0,0342	0,486
Kundengespräch	0,008	0,0718	0,471	0,057	0,4172	0,339	0,049	0,3895	0,349
Mitarbeiter	-0,102	0,9554	0,170	-0,135	1,0499	0,148	-0,034	0,2609	0,397

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Gesamtzufriedenheit in allen drei Altersklassen zu rund 75% erklären und ist als „substanziell“ einzustufen. Bis auf einen Pfadkoeffizienten (Außenzustand) in der höchsten Altersklasse sind alle signifikant. Allerdings ist bis auf die Ausnahme, dass sich der Außenzustand bei den niedrigen beiden Altersklassen im Einzelvergleich mit der höchsten Altersklasse jeweils positiv auf die Gesamtzufriedenheit auswirkt und in der höchsten Altersklasse kein Zusammenhang besteht, kein signifikanter Unterschied zwischen den Pfadkoeffizienten der Altersklassen auszumachen.

4.2.3.3 Korrelationstendenz bezüglich des höchsten Bildungsabschlusses der Mystery Shopper

Die Tabellen 4-2-3-3-1 und 4-2-3-3-2 geben die Pfadkoeffizienten und ihre Unterschiede für die SERVPERF- und Kontaktpunkt-Operationalisierungen der unterschiedlichen Bildungsklassen wieder.

Tabelle 4-2-3-3-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Bildungsabschluss)

	Hauptschule/ Realschule			Abitur			Studium		
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD
Tangibles	0,364	3,087**	0,118	0,354	4,410**	0,080	0,272	2,971**	0,092
Reliability	0,056	0,503	0,111	0,035	0,441	0,080	0,192	2,336*	0,082
Responsiveness	-0,145	0,995	0,146	0,029	0,266	0,109	0,050	0,538	0,092
Assurance	0,362	2,413*	0,150	0,244	2,475*	0,099	0,243	2,945**	0,083
Empathy	0,186	1,549	0,120	0,251	3,007**	0,084	0,052	0,676	0,077
R ²	0,469			0,568			0,448		
Fälle	76			103			197		

Unterschiede:

	HS/RS vs. Abitur			HS/RS vs. Studium			Abitur vs. Studium		
	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p
Tangibles	0,010	0,0708	0,472	0,092	0,5604	0,288	0,082	0,5920	0,277
Reliability									
Responsiveness									
Assurance	0,118	0,6865	0,247	0,119	0,7340	0,232	0,001	0,0067	0,497
Empathy									

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Die Erklärung der Gesamtzufriedenheit ist bei der SERVPERF-Operationalisierung in allen drei Altersklassen zwischen „durchschnittlich“ und „substanziell“ einzustufen, wobei die Bildungsklasse Abitur den höchsten Wert aufweist. Tangibles ist in allen drei Gruppen die stärkste latente exogene Variable, gefolgt von der Variable Assurance. Nur diese beiden Variablen weisen bei allen drei Gruppen einen signifikant positiven Einfluss auf die Gesamtzufriedenheit auf. Allerdings ist der Unterschied nicht signifikant. In der Bildungsklasse Abitur weist die Variable Empathy und in der Bildungsklasse Studium die Variable Reliability ist ein positiver Einfluss auf die Gesamtzufriedenheit festzustellen, der in den anderen Klassen jeweils nicht vorliegt. Insofern bestehen geringe Unterschiede in den unterschiedlichen Bildungsklassen bezüglich ihrer Korrelationstendenz.

Tabelle 4-2-3-3-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Bildungsabschluss)

	Hauptschule/ Realschule			Abitur			Studium		
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD
Außenzustand	0,141	2,116*	0,067	0,149	1,844	0,081	0,235	5,617**	0,042
Innenzustand	0,228	2,966**	0,077	0,281	2,908**	0,097	0,144	3,211**	0,045
Kundengespräch	0,255	2,646**	0,096	0,246	2,840**	0,087	0,368	5,307**	0,069
Mitarbeiter	0,441	4,648**	0,095	0,372	4,414**	0,084	0,357	6,055**	0,059
R ²	0,808			0,717			0,770		
Fälle	76			103			197		

Unterschiede:

	HS/RS vs. Abitur			HS/RS vs. Studium			Abitur vs. Studium		
	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p
Außenzustand				-0,094	1,1943	0,117			
Innenzustand	-0,053	0,411	0,341	0,084	0,9690	0,167	0,137	1,4712	0,071
Kundengespräch	0,008	0,0655	0,474	-0,113	0,8971	0,185	-0,122	1,0660	0,144
Mitarbeiter	0,069	0,5403	0,295	0,084	0,7587	0,216	0,016	0,1558	0,483

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Gesamtzufriedenheit in allen drei Bildungsklassen zu über 70% erklären und ist als „substanziell“ einzustufen. Bis auf einen Pfadkoeffizienten (Außenzustand) in der Bildungsklasse Abitur sind alle signifikant. Allerdings ist bis auf die Ausnahme, dass sich der Außenzustand bei den Bildungsklassen Hauptschule/Realschule und Studium im Einzelvergleich mit der Bildungsklasse Abitur positiv jeweils positiv auf die Gesamtzufriedenheit auswirkt und in der Bildungsklasse Abitur kein Zusammenhang besteht, kein signifikanter Unterschied zwischen den Pfadkoeffizienten der Bildungsklassen auszumachen. Insofern besteht lediglich ein sehr geringer Unterschied bezüglich der Korrelationstendenz zwischen unterschiedlichen Bildungsklassen.

4.2.3.4 Korrelationstendenz bezüglich der Tätigkeit der Mystery Shopper

Zur Ermittlung der Korrelationstendenz wurden unterschiedliche Tätigkeitsgruppen gegenübergestellt. Aufgrund geringer Fallzahlen schieden zwei Gruppen aus, so dass lediglich die Gruppen Schüler/Student und voll berufstätig in der Auswertung miteinander verglichen werden konnten. Die Ergebnisse sind in den Tabellen 4-2-3-4-1 und 4-2-3-4-2 dargestellt.

Tabelle 4-2-3-4-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Tätigkeit)

	Schüler/Student			voll berufstätig			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Tangibles	0,279	1,666	0,167	0,320	5,075**	0,063	-0,041			ja
Reliability	0,071	0,591	0,121	0,191	2,609**	0,073	-0,119			ja
Responsiveness	-0,037	0,279	0,131	0,051	0,685	0,075	-0,088			nein
Assurance	0,480	3,617**	0,133	0,234	2,858**	0,082	0,247	1,285	0,100	nein
Empathy	0,169	1,139	0,149	0,050	0,789	0,063	0,119			nein
R ²	0,584			0,498						
Fälle	51			255						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Die Erklärung der Gesamtzufriedenheit ist bei der SERVPERF-Operationalisierung in beiden Tätigkeitsgruppen zwischen „durchschnittlich“ und „substanziell“ einzustufen, wobei die Gruppe der Schüler und Studenten den höchsten Wert aufweisen. Während drei der fünf Beziehungen im Strukturmodell bei der Gruppe der voll Berufstätigen signifikant sind, ist es bei der Gruppe der Schüler und Studenten eine. Bei der Variable Assurance, die bei beiden Gruppen einen positiven Einfluss auf die Gesamtzufriedenheit hat, ist kein signifikanter Unterschied zwischen den Gruppen auszumachen. Der Unterschied bei weiteren Pfadkoeffizienten resultiert bei den Variablen Tangibles und Reliability aus einer signifikanten Beziehung der einen Gruppe und einer insignifikanten Beziehung der anderen Gruppe. Die beiden Variablen haben bei der Gruppe „voll berufstätig“ einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Insofern besteht ein geringer Unterschied bei der Korrelationstendenz zwischen Schülern/Studenten und voll Berufstätigen bei der SERVPERF-Operationalisierung.

Tabelle 4-2-3-4-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Tätigkeit)

	Schüler/Student			voll berufstätig			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Außenzustand	0,101	1,294	0,078	0,209	4,679**	0,045	-0,108			ja
Innenzustand	0,218	2,295*	0,095	0,186	3,615**	0,052	0,032	0,261	0,397	nein
Kundengespräch	0,483	3,079**	0,157	0,297	5,095**	0,058	0,186	1,263	0,104	nein
Mitarbeiter	0,220	1,821	0,121	0,406	7,928**	0,051	-0,186			ja
R ²	0,811			0,751						
Fälle	51			255						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Erklärung der Gesamtzufriedenheit bei beiden Gruppen als „substanziell“ einstufen. Während alle vier Beziehungen im Strukturmodell bei der Gruppe der voll Berufstätigen signifikant sind, sind es bei der Gruppe der Schüler und Studenten zwei. Die Variablen Innenzustand und Kundengespräch, die bei

beiden Gruppen signifikant sind, unterscheiden sich zwischen den Gruppen nicht signifikant. Die Variablen Außenzustand und Mitarbeiter haben bei der Gruppe der voll Berufstätigen einen signifikant positiven Einfluss auf die Gesamtzufriedenheit und bei der Gruppe der Schüler und Studenten keinen signifikant positiven Einfluss. Insofern unterscheiden sich die beiden Gruppen in ihrer Korrelationstendenz bei der Kontaktpunkt-Operationalisierung.

4.2.3.5 Korrelationstendenz bezüglich der Erfahrung als Mystery Shopper

Die Tabellen 4-2-3-5-1 und 4-2-3-5-2 geben Pfadkoeffizienten für die SERVPERF- und Kontaktpunkt-Operationalisierungen für Mystery Shopper mit und ohne Testerfahrung wieder.

Tabelle 4-2-3-5-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Testerfahrung)

	keine Erfahrung			Erfahrung			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Tangibles	0,253	3,783**	0,067	0,402	4,803**	0,084	-0,149	1,221	0,111	nein
Reliability	0,112	1,655	0,068	0,198	2,242*	0,088	-0,086			ja
Responsiveness	0,014	0,179	0,080	-0,034	0,353	0,096	0,048			nein
Assurance	0,291	3,941**	0,074	0,163	1,639	0,100	0,128			ja
Empathy	0,100	1,611	0,062	0,230	2,338*	0,099	-0,130			ja
R ²	0,406			0,643						
Fälle	277			99						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Während die Erklärung der Gesamtzufriedenheit bei der Gruppe ohne Testerfahrung als „durchschnittlich“ einzustufen ist, ist sie bei der Gruppe mit Testerfahrung nahezu auf dem Niveau „substanziell“. Während drei der fünf Beziehungen im Strukturmodell bei der Gruppe mit Erfahrung signifikant sind, sind es bei der Gruppe ohne Erfahrung nur zwei. Der Unterschied bei den Pfadkoeffizienten resultiert bei den Variablen Reliability, Assurance und Empathy aus einer signifikanten Beziehung der einen Gruppe und einer insignifikanten Beziehung der anderen Gruppe. Die Variable Assurance hat bei der Gruppe ohne Erfahrung einen signifikant positiven Einfluss auf die Gesamtzufriedenheit und die Variablen Reliability und Empathy bei der Gruppe mit Erfahrung. Bei der Variable Tangibles ist kein signifikanter Unterschied auszumachen. Demnach besteht ein Unterschied bei der Korrelationstendenz zwischen Personen mit und ohne Mystery Shopping-Erfahrung bei der SERVPERF-Operationalisierung.

Tabelle 4-2-3-5-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Testerfahrung)

	keine Erfahrung			Erfahrung			Unterschied			
	Pfad- koeff- fizient	t-Wert	SD	Pfad- koeff- fizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Außenzustand	0,189	4,808**	0,039	0,260	3,431**	0,076	-0,071	0,897	0,185	nein
Innenzustand	0,222	4,866**	0,046	0,096	1,610	0,059	0,126			ja
Kundengespräch	0,379	7,424**	0,051	0,187	2,309*	0,081	0,192	1,963	0,025*	ja
Mitarbeiter	0,299	6,238**	0,048	0,584	7,724**	0,076	-0,285	3,105	0,001**	ja
R ²	0,748			0,796						
Fälle	277			99						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Erklärung der Gesamtzufriedenheit als „substanziell“ einstufen. Signifikante Unterschiede bestehen bei den Variablen Innenzustand, Kundengespräch und Mitarbeiter. Der Innenzustand wirkt sich bei der Gruppe ohne Mystery Shopping-Erfahrung positiv auf die Gesamtzufriedenheit aus, während er bei der Gruppe mit Erfahrung keinen Einfluss hat. Bei der Gruppe ohne Erfahrung hatte das Kundengespräch den größten Einfluss auf die Gesamtzufriedenheit. Dabei besteht ein signifikanter Unterschied zur Gruppe mit Erfahrung. Umgekehrt ist der Zusammenhang bei der Variablen Mitarbeiter, die bei der Gruppe mit Erfahrung den stärksten Einfluss hat, wobei der Unterschied zur Gruppe ohne Erfahrung signifikant ist. Demnach hat einerseits die Interaktion zwischen Mystery Shopper und Mitarbeiter für einen Tester ohne Erfahrung und andererseits die Beurteilung der Mitarbeiter für einen Tester mit Erfahrung jeweils einen höheren Stellenwert. Insofern unterscheiden sich Mystery Shopper mit und ohne Erfahrung in ihrer Korrelationstendenz bei der Kontaktpunkt-Operationalisierung.

4.2.3.6 Korrelationstendenz bezüglich der Fast Food-Affinität der Mystery Shopper

Zur Ermittlung der Korrelationstendenz bezüglich der Fast Food-Affinität der Mystery Shopper wurden drei Gruppen gebildet.⁴³⁶ Die Tabellen 4-2-3-6-1 und 4-2-3-6-2 geben die Pfadkoeffizienten und Unterschiede für die SERVPERF- und Kontaktpunkt-Operationalisierungen wieder.

⁴³⁶ Die Fast Food-Affinität wurde mit dem Item „Losgelöst von Ihrer heutigen Aufgabe als Testbesucher, wie gern essen Sie in einem Fast Food Restaurant?“ erfragt. Die Einstellungen „sehr gern“ und „eher gern“ wurden in der Ausprägung „hoch“, die Einstellungen „weniger gern“ und „gar nicht gern“ wurden in der Ausprägung „niedrig“ zusammengefasst, während die Einstellung „teils-teils“ die Ausprägung „mittel“ bildet.

Tabelle 4-2-3-6-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Fast Food-Affinität)

	niedrig			mittel			hoch		
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD
Tangibles	0,368	3,883**	0,095	0,256	3,118**	0,082	0,303	3,247**	0,093
Reliability	0,111	1,159	0,096	0,255	3,433**	0,074	0,049	0,405	0,121
Responsiveness	-0,143	1,341	0,107	0,122	1,192	0,103	0,126	1,165	0,109
Assurance	0,246	2,624**	0,094	0,182	1,664	0,109	0,397	3,385**	0,117
Empathy	0,242	2,267*	0,107	0,057	0,685	0,083	0,014	0,160	0,090
R ²	0,489			0,526			0,509		
Fälle	120			163			92		

Unterschiede:

	niedrig vs. mittel			niedrig vs. hoch			mittel vs. hoch		
	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p
Tangibles	0,112	0,8980	0,185	0,065	0,4814	0,315	-0,047	0,3660	0,357
Reliability									
Responsiveness									
Assurance				-0,151	1,0194	0,155			
Empathy									

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Die Erklärung der Gesamtzufriedenheit ist bei der SERVPERF-Operationalisierung in allen drei Affinitätsklassen zwischen „durchschnittlich“ und „substanziell“ einzustufen, wobei die mittlere Klasse den höchsten Wert aufweist. Bei der Gruppe mit der niedrigen Fast Food-Affinität sind drei signifikante Pfadkoeffizienten festzustellen, während in den anderen beiden nur je zwei auszumachen sind. Die exogene latente Variable Tangibles weist bei allen drei Gruppen einen signifikant positiven Einfluss auf die Gesamtzufriedenheit auf. Allerdings ist der Unterschied zwischen den Gruppen nicht signifikant. Bei der Variable Assurance besteht zwischen der Gruppe niedriger und hoher Affinität kein signifikanter Unterschied. Im jeweiligen Vergleich zur Gruppe mit einer mittleren Ausprägung ist für die beiden Gruppen jeweils ein signifikant positiver Einfluss auf die Gesamtzufriedenheit festzustellen. Die Gruppe mit der mittleren Ausprägung hat bei der Variable Reliability einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, der bei den anderen beiden Gruppen nicht vorliegt. Gleiches ist für die Variable Empathy für die Gruppe mit der niedrigen Fast Food-Affinität festzustellen. Insofern besteht ein Unterschied bei der Korrelationstendenz zwischen Personen mit unterschiedlicher Fast Food-Affinität bei der SERVPERF-Operationalisierung.

Tabelle 4-2-3-6-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Fast Food-Affinität)

	niedrig			mittel			hoch		
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD
Außenzustand	0,191	2,799**	0,068	0,135	2,792**	0,048	0,240	4,116**	0,058
Innenzustand	0,134	2,403*	0,056	0,272	4,185**	0,065	0,095	1,348	0,070
Kundengespräch	0,273	3,624**	0,075	0,375	5,317**	0,071	0,348	3,602**	0,097
Mitarbeiter	0,468	6,413**	0,073	0,304	4,974**	0,061	0,391	4,696**	0,083
R ²	0,762			0,762			0,761		
Fälle	120			163			92		

Unterschiede:

	niedrig vs. mittel			niedrig vs. hoch			mittel vs. hoch		
	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p
Außenzustand	0,056	0,6872	0,246	-0,049	0,5290	0,299	-0,105	1,3477	0,090
Innenzustand	-0,138	1,543	0,062	0,040			0,177		
Kundengespräch	-0,102	0,9759	0,165	-0,075	0,6219	0,267	0,027	0,2292	0,409
Mitarbeiter	0,163	1,7285	0,043	0,076	0,6917	0,245	-0,087	0,8523	0,197

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Erklärung der Gesamtzufriedenheit mit über 76% in allen drei Gruppen als „substanziell“ einstufen. Die Pfadkoeffizienten sind bei allen Gruppen bis auf eine Ausnahme (Innenzustand bei der Gruppe mit der höchsten Fast Food-Affinität) signifikant. Allerdings sind bei den signifikanten Pfadkoeffizienten keine signifikanten Unterschiede zwischen den Gruppen auszumachen. Für die Gruppen mit der niedrigen und der mittleren Ausprägung ist im Vergleich zur Gruppe mit der hohen Affinitätsausprägung ein signifikant positiver Einfluss der Variablen Innenzustand auf die Gesamtzufriedenheit festzustellen. Demnach hat der Innenzustand für die Gruppe mit einer hohen Fast Food-Ausprägung keinen signifikanten Einfluss auf die Gesamtzufriedenheit. Insofern unterscheiden sich die Gruppen mit unterschiedlicher Fast Food-Affinität bei der Kontaktpunkt-Operationalisierung hinsichtlich ihrer Korrelationstendenz lediglich sehr gering. Die in Abschnitt 4.2.1.6 geäußerte Vermutung, die Fast Food-Affinität der Tester könnte auf ihr Gesamtzufriedenheitsurteil ausstrahlen, kann nicht erklärt werden.

4.2.3.7 Korrelationstendenz bezüglich der Restaurant-Affinität der Mystery Shopper

Zur Ermittlung der Korrelationstendenz bezüglich der Restaurant-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen für die einzelnen Items gegenübergestellt.⁴³⁷ Die Tabellen 4-2-3-7-1 und 4-2-3-7-2 geben die Pfadkoeffizienten und Unterschiede für die SERVPERF- und Kontaktpunkt-Operationalisierungen wieder.

⁴³⁷ Die Restaurant-Affinität wurde mit dem Item „Und wie gern essen Sie in diesem Fast Food-Restaurant?“ erfragt. Dabei war nicht die spezielle Filiale sondern irgendeine Filiale dieser Kette gemeint. Die Einstellungen „sehr gern“ und „eher gern“ wurden in der Ausprägung „hoch“, die Einstellungen „weniger

Tabelle 4-2-3-7-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Restaurant-Affinität)

	niedrig			mittel			hoch		
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD
Tangibles	0,363	3,589**	0,101	0,288	3,080**	0,094	0,213	2,369*	0,090
Reliability	0,143	1,385	0,103	0,117	1,554	0,076	0,319	3,159**	0,101
Responsiveness	-0,026	0,190	0,135	0,176	1,758	0,100	-0,073	0,700	0,104
Assurance	0,196	1,911	0,103	0,163	1,677	0,097	0,315	2,687**	0,117
Empathy	0,150	1,424	0,105	0,100	1,027	0,097	0,091	1,067	0,085
R2	0,505			0,455			0,506		
Fälle	123			143			105		

Unterschiede:

	niedrig vs. mittel			niedrig vs. hoch			mittel vs. hoch		
	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p
Tangibles	0,075	0,5462	0,293	0,150	1,0981	0,137	0,075	0,5655	0,286
Reliability									
Responsiveness									
Assurance									
Empathy									

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Die Erklärung der Gesamtzufriedenheit ist bei der SERVPERF-Operationalisierung in allen drei Affinitätsklassen zwischen „durchschnittlich“ und „substanziell“ einzustufen, wobei die mittlere Klasse den niedrigsten Wert aufweist. Bei der Gruppe mit der höchsten Restaurant-Affinität sind drei signifikante Pfadkoeffizienten festzustellen, während in den anderen beiden nur je einer auszumachen ist. Die exogene latente Variable Tangibles, die bei allen drei Gruppen einen signifikant positiven Einfluss auf die Gesamtzufriedenheit aufweist, unterscheidet sich zwischen den Gruppen nicht signifikant. Die Variablen Reliability und Assurance haben in der Gruppe mit der höchsten Restaurant-Affinität einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, der bei den anderen beiden Gruppen nicht vorliegt. Insofern besteht ein Unterschied bei der Korrelationstendenz zwischen Personen mit unterschiedlicher Restaurant-Affinität bei der SERVPERF-Operationalisierung.

gern“ und „gar nicht gern“ wurden in der Ausprägung „niedrig“ zusammengefasst, während die Einstellung „teils-teils“ die Ausprägung „mittel“ bildet.

Tabelle 4-2-3-7-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Restaurant-Affinität)

	niedrig			mittel			hoch		
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	t-Wert	SD
Außenzustand	0,274	4,619**	0,059	0,140	2,509*	0,056	0,132	2,403*	0,055
Innenzustand	0,143	2,348*	0,061	0,269	5,306**	0,051	0,264	3,159**	0,084
Kundengespräch	0,299	3,234**	0,092	0,406	6,513**	0,062	0,312	3,431**	0,091
Mitarbeiter	0,373	4,459**	0,084	0,302	4,849**	0,062	0,373	5,262**	0,071
R2	0,771			0,756			0,766		
Fälle	123			143			105		

Unterschiede:

	niedrig vs. mittel			niedrig vs. hoch			mittel vs. hoch		
	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p	Pfad- differenz	t-Wert	p
Außenzustand	0,133	1,6399	0,051	0,142	1,7402	0,042*	0,008	0,1048	0,458
Innenzustand	-0,126	1,614	0,054	-0,121	1,2006	0,116	0,005	0,0552	0,478
Kundengespräch	-0,107	0,9884	0,162	-0,013	0,1015	0,460	0,094	0,8855	0,188
Mitarbeiter	0,071	0,6938	0,244	0,000	0,0027	0,499	-0,071	0,7545	0,226

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Erklärung der Gesamtzufriedenheit mit über 75% in allen drei Gruppen als „substanziell“ einstufen. Alle Pfadkoeffizienten sind bei allen drei Gruppen signifikant. Allerdings ist beim Vergleich der Gruppen lediglich ein signifikanter Unterschied auszumachen: Die Variable Außenzustand hat für die Gruppe mit der niedrigen Affinitätsausprägung einen signifikant höheren Einfluss auf die Gesamtzufriedenheit als die Gruppe mit der hohen Affinitätsausprägung. Insofern unterscheiden sich die Gruppen mit unterschiedlicher Restaurant-Affinität bei der Kontaktpunkt-Operationalisierung hinsichtlich ihrer Korrelationstendenz lediglich sehr gering.

4.2.3.8 Korrelationstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs der Mystery Shopper

Zur Ermittlung der Komponente Korrelationstendenz bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs der Mystery Shopper wurden unterschiedliche Häufigkeitssausprägungen bei den Items miteinander verglichen.⁴³⁸ Die Ergebnisse sind in den Tabellen 4-2-3-8-1 und 4-2-3-8-2 dargestellt.

⁴³⁸ Die Häufigkeit eines Fast Food-Restaurantbesuchs wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit unterschiedlicher Fast Food-Restaurants addiert und in zwei Klassen zusammengefasst wurden.

Tabelle 4-2-3-8-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Fast Food-Häufigkeit)

	0-12 mal im Jahr			>12 mal im Jahr			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Tangibles	0,282	3,742**	0,075	0,350	4,446**	0,079	-0,068	0,626	0,266	nein
Reliability	0,079	1,080	0,074	0,154	2,277*	0,068	-0,075			ja
Responsiveness	0,152	1,687	0,090	-0,063	0,702	0,089				nein
Assurance	0,154	1,705	0,090	0,365	4,047**	0,090	-0,212			ja
Empathy	0,157	1,815	0,086	0,038	0,592	0,064				nein
R ²	0,468			0,514						
Fälle	200			176						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Die Erklärung der Gesamtzufriedenheit ist bei der SERVPERF-Operationalisierung in beiden Häufigkeitsklassen zwischen „durchschnittlich“ und „substanziell“ einzustufen, wobei die höhere Klasse den höheren Wert aufweist. Bei der Gruppe mit der höheren Besuchshäufigkeit sind drei signifikante Pfadkoeffizienten festzustellen, während in der niedrigeren Klasse nur einer auszumachen ist. Die exogene latente Variable Tangibles, die bei beiden Gruppen einen signifikant positiven Einfluss auf die Gesamtzufriedenheit aufweist, unterscheidet sich zwischen den Gruppen nicht signifikant. Die Variablen Reliability und Assurance haben bei der Gruppe mit der höheren Besuchshäufigkeit einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, der bei der anderen Gruppe nicht vorliegt. Insofern besteht ein Unterschied bei der Korrelationstendenz zwischen Personen mit unterschiedlich hoher Besuchshäufigkeit in einem Fast Food-Restaurant bei der SERVPERF-Operationalisierung.

Tabelle 4-2-3-8-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Fast Food-Häufigkeit)

	0-12 mal im Jahr			>12 mal im Jahr			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Außenzustand	0,196	4,332**	0,045	0,182	3,455**	0,053	0,014	0,203	0,420	nein
Innenzustand	0,160	3,955**	0,040	0,228	3,014**	0,076	-0,069	0,832	0,203	nein
Kundengespräch	0,325	5,183**	0,063	0,314	3,956**	0,079	0,011	0,112	0,455	nein
Mitarbeiter	0,400	6,699**	0,060	0,361	5,476**	0,066	0,039	0,437	0,331	nein
R ²	0,795			0,711						
Fälle	200			176						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Erklärung der Gesamtzufriedenheit mit über 70% in beiden Gruppen als „substanziell“ einstufen, wobei die Gruppe mit der höheren Besuchshäufigkeit einen niedrigeren Wert aufweist. Alle Pfadkoeffizienten sind signifikant. Allerdings erweist sich der Unterschied zwischen den Gruppen als nicht signifikant. Insofern besteht bei der Korrelationstendenz bei Personen mit unterschiedlich

hoher Besuchshäufigkeit in einem Fast Food-Restaurant bei der Kontaktpunkt-Operationalisierung kein Unterschied.

4.2.3.9 Korrelationstendenz bezüglich der spezifischen Restaurant-Besuchshäufigkeit

Schließlich wurde die Komponente Korrelationstendenz bezüglich der Häufigkeit eines Besuchs in einer Filiale der getesteten Fast Food-Kette ermittelt.⁴³⁹ Die Ergebnisse sind in den Tabellen 4-2-3-8-1 und 4-2-3-8-2 dargestellt.

Tabelle 4-2-3-9-1: Pfadkoeffizienten des SERVPERF-Strukturmodells (Restaurant-Besuchshäufigkeit)

	0-12 mal im Jahr			>12 mal im Jahr			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Tangibles	0,291	4,557**	0,064	0,260	2,214*	0,117	0,031	0,231	0,409	nein
Reliability	0,134	2,069*	0,065	0,095	0,833	0,114	0,039			ja
Responsiveness	0,105	1,451	0,072	0,050	0,402	0,125	0,054			nein
Assurance	0,215	3,091**	0,070	0,382	4,044**	0,094	-0,166	1,201	0,115	nein
Empathy	0,084	1,417	0,059	0,080	0,861	0,093	0,004			nein
R ²	0,475			0,478						
Fälle	291			85						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Die Erklärung der Gesamtzufriedenheit ist bei der SERVPERF-Operationalisierung in beiden Häufigkeitsklassen zwischen „durchschnittlich“ und „substanziell“ einzustufen. Bei der Gruppe mit der niedrigeren Besuchshäufigkeit sind drei signifikante Pfadkoeffizienten festzustellen, während in der Klasse mit der höheren Ausprägung zwei auszumachen sind. Die exogenen latenten Variablen Tangibles und Assurance, die bei beiden Gruppen einen signifikant positiven Einfluss auf die Gesamtzufriedenheit aufweisen, unterscheiden sich zwischen den Gruppen nicht signifikant. Die Variable Reliability hat bei der Gruppe mit der niedrigeren Besuchshäufigkeit einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, der bei der anderen Gruppe nicht vorliegt. Insofern besteht bei der SERVPERF-Operationalisierung ein geringer Unterschied bezüglich der Korrelationstendenz zwischen Personen mit unterschiedlich hoher Besuchshäufigkeit in einer Filiale der getesteten Fast Food-Kette.

⁴³⁹ Die Häufigkeit eines Besuchs in dem getesteten Restaurant wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit in einem Restaurant der Fast Food-Kette addiert und in unterschiedlichen Klassen zusammengefasst wurden.

Tabelle 4-2-3-9-2: Pfadkoeffizienten des Kontaktpunkt-Strukturmodells (Restaurant-Besuchshäufigkeit)

	0-12 mal im Jahr			>12 mal im Jahr			Unterschied			
	Pfad- koeffizient	t-Wert	SD	Pfad- koeffizient	T-Wert	SD	Pfad- differenz	t- Wert	p	Sig.
Außenzustand	0,214	5,188**	0,041	0,133	1,804	0,074	0,081	0,946		ja
Innenzustand	0,134	3,427**	0,039	0,378	3,269**	0,116	-0,244	2,566	0,005**	ja
Kundengespräch	0,297	5,499**	0,054	0,374	3,438**	0,109	-0,077	0,668	0,252	nein
Mitarbeiter	0,424	8,421**	0,050	0,260	2,856**	0,091	0,164	1,561	0,056	nein
R ²	0,767			0,703						
Fälle	291			85						

* Signifikant auf dem Niveau 5%; ** signifikant auf dem Niveau 1%

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung lässt sich die Erklärung der Gesamtzufriedenheit mit über 70% in beiden Gruppen als „substanziell“ einstufen, wobei die Gruppe mit der höheren Besuchshäufigkeit einen vergleichsweise niedrigeren Wert aufweist. Bis auf eine Ausnahme (Außenzustand bei der Gruppe mit hoher Besuchshäufigkeit) sind alle Pfadkoeffizienten signifikant. Signifikante Unterschiede zwischen den Gruppen bestehen bei den Variablen Außenzustand und Innenzustand. Während die Variable Außenzustand bei der Gruppe mit der niedrigen Besuchshäufigkeit einen signifikant positiven Einfluss auf die Gesamtzufriedenheit hat, hat sie bei der anderen Gruppe keinen Einfluss. Der Innenzustand hat bei der Gruppe mit der höheren Besuchshäufigkeit einen höheren Stellenwert. Insofern besteht bei der Kontaktpunkt-Operationalisierung bei Personen mit unterschiedlich hoher Besuchshäufigkeit in einem Restaurant der getesteten Fast Food-Kette ein Unterschied bezüglich der Korrelationstendenz.

4.2.3.10 Zusammenfassung der Ergebnisse der Korrelationstendenz

Für die Komponente Korrelationstendenz ist zu konstatieren, dass in allen neun untersuchten Gruppen Unterschiede zwischen den jeweiligen Merkmalsträgern festzustellen sind. Empathy hat bei den Frauen und Assurance sowie Reliability bei den Männern einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, der in der anderen Gruppe jeweils nicht vorliegt. Assurance und Außenzustand haben bei der niedrigsten und mittleren Altersklasse und Reliability bei der mittleren Altersklasse einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, der in den Vergleichsgruppen jeweils nicht vorliegt. Bei der Bildungsklasse Abitur (Studium) hat Empathy (Reliability) einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, der in den anderen Klassen nicht vorliegt. Der Außenzustand hat bei der Bildungsklasse Abitur keinen Einfluss. Tangibles, Reliability, Außenzustand und Mitarbeiter haben bei Vollzeit Berufstätigen einen signifikant positiven Einfluss auf die Gesamtzufriedenheit und bei Schülern/Studenten keinen signifikanten

Einfluss⁴⁴⁰. Assurance und Innenzustand (Reliability und Empathy) haben bei der Gruppe ohne Erfahrung (mit Erfahrung) einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Kundengespräch (Mitarbeiter) hat bei der Gruppe ohne Erfahrung (mit Erfahrung) einen höheren Stellenwert. Assurance hat bei niedriger und hoher Fast Food-Affinität einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Reliability (Empathy) hat bei mittlerer (niedriger) Fast Food-Affinität einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Der Innenzustand hat bei hoher (niedriger und mittlerer) keinen (einen positiven) signifikanten Einfluss auf die Gesamtzufriedenheit. Reliability und Assurance haben bei hoher Restaurant-Affinität einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, der bei den anderen beiden Affinitätsklassen nicht vorliegt. Außenzustand hat bei niedriger Restaurant-Affinität einen höheren Einfluss auf die Gesamtzufriedenheit als bei hoher Restaurant-Affinität. Reliability und Assurance haben bei hoher Besuchshäufigkeit in einem Fast Food-Restaurant einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, während in der Gruppe mit niedriger Besuchshäufigkeit kein signifikanter Zusammenhang festzustellen ist. Reliability und Außenzustand haben bei niedriger Besuchshäufigkeit eines Restaurants der getesteten Kette einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Innenzustand hat bei der Gruppe mit hoher Besuchshäufigkeit einen höheren Stellenwert.

4.2.4 Zuverlässigkeit von Mystery Shoppern

Zu Ermittlung der Zuverlässigkeit der Mystery Shopper wurden zwei Komponenten ausgewertet: Einerseits die Beurteilerübereinstimmung und andererseits die Beurteilerreliabilität. Die Beurteilerübereinstimmung ist für die Analyse lediglich interessant, sofern geringe Werte für die Reliabilität vorliegen, da im Falle von gleichzeitig vorliegenden niedrigen Werten für die Übereinstimmung eindeutig auf die Unzuverlässigkeit der Urteile geschlossen werden kann.⁴⁴¹ Analog der vorangegangenen Abschnitte werden die Werte der einzelnen Ausprägungen getrennt für die Operationalisierung der SERVPERF-Dimensionen⁴⁴² und die Operationalisierung der Kontaktpunkte wiedergegeben⁴⁴³. Aus forschungsökonomischen Gründen kann die Zuverlässigkeit nur für die die Geschlechter, Altersklassen, Bildungsgruppen, Tätigkeitsgruppen, Affinitätsgruppen des Restaurantbesuchs und die Häufigkeitsgruppen in einem Fast Food-Restaurant berichtet werden.

⁴⁴⁰ Einschränkung ist anzumerken, dass mangels ausreichender Fallzahlen bei den Tätigkeitsgruppen Teilzeit berufstätig und Hausfrau/Rentner lediglich die Tätigkeitsgruppen Schüler/Student und Vollzeit berufstätig miteinander verglichen werden konnten.

⁴⁴¹ Vgl. die Ausführungen in Kapitel 3.1.2

⁴⁴² Vgl. Tabelle 3-2-3

⁴⁴³ Vgl. Tabelle 3-2-4

4.2.4.1 Zuverlässigkeit bezüglich des Geschlechts

Zur Ermittlung der Zuverlässigkeit bezüglich des Geschlechts der Mystery Shopper wurde die Interraterreliabilität der einzelnen Items für Frauen und Männer berechnet. Die Werte sind in den Tabellen 4-2-4-1-1 und 4-2-4-1-2 dargestellt.

Tabelle 4-2-4-1-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Geschlecht

	Männer				Frauen			
	IRR	p	κ	p	IRR	p	κ	p
<i>Tang_1</i>	0,371	0,006	0,104	0,226	0,477*	0,000		
<i>Tang_2</i>	0,457*	0,001			0,471*	0,001		
<i>Tang_3</i>	0,491*	0,000			0,446*	0,003		
<i>Tang_4</i>	0,194	0,102	0,117	0,193	0,488*	0,001		
<i>Rel_1</i>	neg	-	-	-	0,192	0,125	k.s.	-
<i>Rel_2</i>	neg	-	-	-	0,336	0,021	k.s.	-
<i>Rel_3</i>	0,192	0,104	-0,010	0,923	0,283	0,043	k.s.	-
<i>Rel_4</i>	0,104	0,254	k.s.	-	0,240	0,075	k.s.	-
<i>Rel_5</i>	0,152	0,170	0,097	0,413	0,610*	0,000		
<i>Resp_1</i>	0,138	0,190	0,015	0,866	0,084	0,312	k.s.	-
<i>Resp_2</i>	0,369	0,007	k.s.	-	0,542*	0,000		
<i>Resp_3</i>	0,386	0,005	0,070	0,494	0,412*	0,005		
<i>Resp_4</i>	0,220	0,076	k.s.	-	0,482*	0,002		
<i>Ass_1</i>	0,609*	0,000			0,548*	0,000		
<i>Ass_2</i>	0,033	0,416	k.s.	-	0,376	0,010	k.s.	-
<i>Ass_3</i>	0,271	0,038	0,072	0,468	0,435*	0,003		
<i>Ass_4</i>	0,234	0,066	k.s.	-	0,147	0,190	k.s.	-
<i>Emp_1</i>	0,127	0,203	-0,052	0,550	0,384	0,009	0,032	0,711
<i>Emp_2</i>	0,116	0,225	k.s.	-	0,012	0,471	0,009	0,949
<i>Emp_3</i>	0,356	0,008	k.s.	-	0,442*	0,003		
<i>Emp_4</i>	0,354	0,009	0,167	0,077	0,293	0,042	0,041	0,658
<i>Emp_5</i>	0,354	0,009	k.s.	-	0,255	0,062	0,051	0,591

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Den Werten für die SERVPERF-Indikatoren ist zu entnehmen, dass bei den Männerpaaren drei und bei den Frauenpaaren die elf der 22 Items den geforderten Wert von 0,4 annehmen. Der Wert für Cohens κ ist bei den Werten mit nicht akzeptabler Interraterreliabilität durchgängig kleiner als der geforderte Wert von 0,4. Insofern kann bei diesen Items eindeutig auf eine Unzuverlässigkeit der Urteile der Mystery Shopper geschlossen werden. Gleiches gilt für die Items der Männerpaare bei denen ein negativer Wert vorliegt. Wenngleich sich die Aussage treffen lässt, dass die Frauenpaare in der Summe zuverlässigere Urteile abgeben, ist das Ergebnis dennoch wenig befriedigend, da nur die Hälfte der Indikatoren akzeptable Werte aufweisen. Eine Erklärung wird am Schluss dieses Kapitels diskutiert.

Tabelle 4-2-4-1-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Geschlecht

	Männer				Frauen			
	IRR	p	κ	p	IRR	p	κ	p
<i>Aussen_1</i>	0,395	0,004	k.s.	-	0,594*	0,000		
<i>Aussen_2</i>	0,387	0,005	k.s.	-	0,365	0,012	k.s.	-
<i>Aussen_3</i>	0,363	0,007	k.s.	-	0,300	0,031	k.s.	-
<i>Innen_1</i>	0,368	0,007	k.s.	-	0,480*	0,001		
<i>Innen_2</i>	0,439*	0,001			0,311	0,026	k.s.	-
<i>Innen_3</i>	0,330	0,015	0,099	0,398	0,223	0,087	k.s.	-
<i>Innen_4</i>	0,181	0,118	k.s.	-	0,117	0,239	k.s.	-
<i>Innen_5</i>	0,629*	0,000			0,310	0,026	0,042	0,679
<i>Kunde_1</i>	0,420*	0,002			0,753*	0,000		
<i>Kunde_2</i>	0,501*	0,000			0,720*	0,000		
<i>Kunde_3</i>	0,230	0,067	k.s.	-	0,600*	0,000		
<i>Kunde_4</i>	0,520*	0,000			0,688*	0,000		
<i>Kunde_5</i>	0,025	0,436	k.s.	-	0,817*	0,000		
<i>Kunde_6</i>	0,344	0,012	0,081	0,383	0,769*	0,000		
<i>Kunde_7</i>	0,456*	0,001			0,745*	0,000		
<i>Kunde_8</i>	0,880*	0,000	0,877	0,000	0,939*	0,000		
<i>Kunde_9</i>	0,493*	0,000			0,753*	0,000		
<i>Mit_1</i>	0,260	0,043	k.s.	-	0,629*	0,000		
<i>Mit_2</i>	0,122	0,216	0,081	0,408	0,380	0,008	k.s.	-
<i>Mit_3</i>	0,545*	0,000			0,743*	0,000		
<i>Mit_4</i>	0,235	0,680	-0,199	0,078	0,666*	0,000		
<i>Mit_5</i>	0,581*	0,000			0,628*	0,000		
<i>GesZufr</i>	0,461*	0,001			0,587*	0,000		

* akzeptabler Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Bei der Kontaktpunkt-Operationalisierung haben elf Werte der Männerpaare und 16 Werte der Frauenpaare den erforderlichen Wert von mindestens 0,4. Interessant ist, dass bei den Frauenpaaren bei allen Items des Kundengesprächs und bei nahezu allen Items der Mitarbeiterbeurteilung der erforderliche Wert vorliegt und die Urteile somit als zuverlässig eingestuft werden können. Dahingegen erfolgte die Beurteilung von Außen- und Innenzustand vergleichsweise weniger zuverlässig. Im Vergleich zu den Männerpaaren ist das Ergebnis der Frauenpaare auch bei der Kontaktpunkt-Operationalisierung in der Summe als zuverlässiger einzustufen. Allerdings ist die Zuverlässigkeit des Urteils bei Außen- und Innenzustand in beiden Gruppen als wenig befriedigend einzustufen.

4.2.4.2 Zuverlässigkeit bezüglich der Altersklasse

Zur Ermittlung der Zuverlässigkeit bezüglich der Altersklasse der Mystery Shopper wurde die Interraterreliabilität der einzelnen Items für die drei Bildungsklassen Hauptschule/Realschule, Abitur und Studium berechnet. Die Werte sind in den Tabellen 4-2-4-2-1 und 4-2-4-2-2 dargestellt.

Tabelle 4-2-4-2-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Altersklasse

	18-30 Jahre				31-45 Jahre				>45 Jahre			
	IRR	p	κ	p	IRR	p	κ	p	IRR	p	κ	p
<i>Tang_1</i>	0,590*	0,000			0,490*	0,000			0,630*	0,000		
<i>Tang_2</i>	0,779*	0,000			0,552*	0,000			0,433*	0,012		
<i>Tang_3</i>	0,404*	0,014			0,400*	0,000			0,497*	0,004		
<i>Tang_4</i>	0,329	0,039	k.s.	-	0,357	0,000	k.s.	-	0,207	0,152	k.s.	-
<i>Rel_1</i>	0,343	0,033	0,271	0,015	0,465*	0,000			neg	-	-	-
<i>Rel_2</i>	0,128	0,251	k.s.	-	0,292	0,007	k.s.	-	0,183	0,805	-0,127	0,354
<i>Rel_3</i>	0,406*	0,014			0,161	0,080	k.s.	-	0,088	0,335	k.s.	-
<i>Rel_4</i>	0,556*	0,001			0,230	0,024	k.s.	-	0,393	0,027	0,230	0,181
<i>Rel_5</i>	0,638*	0,000			0,254	0,016	k.s.	-	0,421*	0,017		
<i>Resp_1</i>	0,022	0,454	0,292	0,008	0,171	0,070	0,103	0,098	0,722*	0,000		
<i>Resp_2</i>	0,566*	0,001			0,406*	0,000			0,418*	0,017		
<i>Resp_3</i>	0,526*	0,002			0,218	0,029	0,051	0,498	0,173	0,205	k.s.	-
<i>Resp_4</i>	0,381	0,023	k.s.	-	0,402*	0,000			0,405*	0,029		
<i>Ass_1</i>	0,338	0,038	k.s.	-	0,568*	0,000			0,330	0,048	k.s.	-
<i>Ass_2</i>	0,631*	0,000			0,192	0,047	0,117	0,091	0,017	0,465	k.s.	-
<i>Ass_3</i>	0,547*	0,001			0,399	0,000	0,135	0,052	0,271	0,087	0,249	0,106
<i>Ass_4</i>	0,512*	0,002			0,402*	0,000			neg	-	-	-
<i>Emp_1</i>	0,293	0,062	0,043	0,693	0,459*	0,000			0,378	0,029	0,120	0,377
<i>Emp_2</i>	neg	-	-	-	0,275	0,008	k.s.	-	0,239	0,117	0,155	0,348
<i>Emp_3</i>	0,526*	0,002			0,365	0,001	0,206	0,002	0,347	0,042	k.s.	-
<i>Emp_4</i>	0,398	0,017	0,185	0,079	0,249	0,016	k.s.	-	0,141	0,252	k.s.	-
<i>Emp_5</i>	0,237	0,109	0,055	0,578	0,298	0,005	0,066	0,322	0,326	0,053	k.s.	-

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Den Werten der SERVPERF-Operationalisierung ist zu entnehmen, dass zwölf Items aus der niedrigsten Altersklasse, neun Items aus der mittleren und sieben aus der höchsten Altersklasse den erforderlichen Wert von mindestens 0,4 aufweisen. Die nicht als zuverlässig einzustufenden Beurteilungen erweisen sich durch Überprüfung mit dem Übereinstimmungsmaß sämtlich als eindeutig unzuverlässig. Es zeichnet sich ein Trend ab, demnach die niedrigste Altersklasse bei den SERVPERF-Items am häufigsten zuverlässige Beurteilungen abgibt. Allerdings wird nur gut die Hälfte der Items von den Paaren aus der niedrigsten Altersklasse zuverlässig beurteilt. Insofern ist das Ergebnis wenig befriedigend.

Tabelle 4-2-4-2-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Altersklasse

	18-30 Jahre				31-45 Jahre				>45 Jahre			
	IRR	p	κ	p	IRR	p	κ	p	IRR	p	κ	p
<i>Aussen_1</i>	0,432*	0,010			0,475*	0,000			0,758*	0,000		
<i>Aussen_2</i>	0,377	0,021	0,133	0,359	0,572*	0,000			0,522*	0,000		
<i>Aussen_3</i>	0,339	0,034	k.s.	-	0,435*	0,000			0,400*	0,020		
<i>Innen_1</i>	0,122	0,262	k.s.	-	0,474*	0,000			0,561*	0,001		
<i>Innen_2</i>	0,593*	0,000			0,455*	0,000			0,385	0,024	k.s.	-
<i>Innen_3</i>	0,569*	0,001			0,165	0,077	k.s.	-	0,505*	0,004		
<i>Innen_4</i>	neg	-	-	-	0,310	0,003	0,266	0,013	0,840*	0,000		
<i>Innen_5</i>	0,403*	0,014			0,471*	0,000			0,645*	0,000		
<i>Kunde_1</i>	0,733*	0,000			0,534*	0,000			0,627*	0,000		
<i>Kunde_2</i>	0,626*	0,000			0,729*	0,000			0,529*	0,003		
<i>Kunde_3</i>	0,555*	0,001			0,452*	0,000			0,330	0,051	0,314	0,037
<i>Kunde_4</i>	0,481*	0,004			0,549*	0,000			0,619*	0,001		
<i>Kunde_5</i>	0,431*	0,009			0,598*	0,000			0,455*	0,010		
<i>Kunde_6</i>	0,494*	0,003			0,544*	0,000			0,240	0,120	k.s.	-
<i>Kunde_7</i>	0,606*	0,000			0,555*	0,000			0,492*	0,006		
<i>Kunde_8</i>	0,926*	0,000			0,906*	0,000			1,000*	0,000		
<i>Kunde_9</i>	0,658*	0,000			0,627*	0,000			0,576*	0,001		
<i>Mit_1</i>	0,442*	0,008			0,488*	0,000			0,581*	0,001		
<i>Mit_2</i>	0,191	0,158	0,150	0,153	0,380	0,000	0,106	0,177	0,558*	0,002		
<i>Mit_3</i>	0,426*	0,010			0,456*	0,000			0,537*	0,002		
<i>Mit_4</i>	0,544*	0,001			0,488*	0,000			0,076	0,357	0,063	0,681
<i>Mit_5</i>	0,475*	0,004			0,528*	0,000			0,551*	0,002		
<i>GesZufr</i>	0,431*	0,010			0,464*	0,000			0,639*	0,000		

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Der Trend aus der SERVPERF-Operationalisierung wird bei den Items der Kontaktpunkt-Operationalisierung nicht widerspiegelt. In der niedrigsten und höchsten Altersklasse sind vier und in der mittleren Altersklasse drei Items als eindeutig nicht zuverlässig beurteilt einzustufen, während sämtliche andere Items den erforderlichen Wert von 0,4 überschreiten. Insofern ist die Zuverlässigkeit der Mystery Shopper unterschiedlicher Altersklassen bei der Kontaktpunkt-Operationalisierung nicht unterschiedlich.

4.2.4.3 Zuverlässigkeit bezüglich des höchsten Bildungsabschlusses

Zur Ermittlung der Zuverlässigkeit bezüglich des höchsten Bildungsabschlusses der Mystery Shopper wurde die Interraterreliabilität der einzelnen Items für die drei Bildungsklassen Hauptschule/Realschule, Abitur und Studium berechnet. Die Werte sind in den Tabellen 4-2-4-3-1 und 4-2-4-3-2 dargestellt.

Tabelle 4-2-4-3-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Bildungsabschluss

	Hauptschule/Realschule				Abitur				Studium			
	IRR	p	κ	p	IRR	p	κ	p	IRR	p	κ	p
<i>Tang_1</i>	0,587*	0,001			0,548*	0,001			0,417*	0		
<i>Tang_2</i>	0,401*	0,027			0,547*	0,001			0,539*	0		
<i>Tang_3</i>	0,411*	0,024			0,674*	0			0,355	0,001	0,121	0,158
<i>Tang_4</i>	0,224	0,148	-0,031	0,816	0,499*	0,003			0,295	0,007	k.s.	-
<i>Rel_1</i>	neg	-	-	-	0,252	0,091	0,06	0,602	0,464*	0		
<i>Rel_2</i>	0,341	0,062	k.s.	-	neg	-	-	-	0,107	0,192	0,111	0,177
<i>Rel_3</i>	neg	-	-	-	0,445*	0,007			0,071	0,276	k.s.	-
<i>Rel_4</i>	0,191	0,198	-0,111	0,531	0,375	0,021	k.s.	-	0,258	0,016	k.s.	-
<i>Rel_5</i>	0,398	0,031	0,323	0,054	0,567*	0,001			0,417*	0		
<i>Resp_1</i>	0,556*	0,004			0,082	0,337	0,15	0,16	0,266	0,013	0,134	0,049
<i>Resp_2</i>	0,420*	0,024			0,625*	0			0,367	0,001	k.s.	-
<i>Resp_3</i>	0,422*	0,026			0,398	0,017	0,25	0,05	0,203	0,044	0,004	0,956
<i>Resp_4</i>	0,252	0,131	0,398	0,006	0,183	0,177	0,367	0,003	0,359	0,001	k.s.	-
<i>Ass_1</i>	0,370	0,039	k.s.	-	0,518*	0,003			0,457*	0		
<i>Ass_2</i>	0,022	0,459	k.s.	-	0,454*	0,008			0,192	0,054	0,173	0,017
<i>Ass_3</i>	0,559*	0,002			0,285	0,072	0,041	0,74	0,318	0,003	0,04	0,58
<i>Ass_4</i>	0,327	0,061	0,411	0,008	0,096	0,317	k.s.	-	0,247	0,02	k.s.	-
<i>Emp_1</i>	0,303	0,082	0,019	0,887	0,346	0,034	0,011	0,916	0,433*	0		
<i>Emp_2</i>	0,400*	0,027			neg	-	-	-	0,282	0,008	k.s.	-
<i>Emp_3</i>	0,439*	0,019			0,351	0,032	k.s.	-	0,424*	0		
<i>Emp_4</i>	0,296	0,087	k.s.	-	0,540	0,001			0,372	0,001	0,132	0,055
<i>Emp_5</i>	0,276	0,103	k.s.	-	0,355	0,03	0,357	0,003	0,263	0,013	0,055	0,431

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Die Werte für die SERVPERF-Operationalisierung geben ein diffuses Bild wieder. Die Gruppe mit dem höchsten Bildungsabschluss hat bei sieben, die Gruppe mit dem niedrigsten Bildungsabschluss bei neun und die Gruppe mit Abitur bei zehn Items den geforderten Wert von mindestens 0,4 für die Interraterreliabilität. Bei den Items, die den Wert nicht aufweisen, deuten die Werte für die Beurteilerübereinstimmung mit unter 0,4 auf eine eindeutig unzuverlässige Beurteilung hin. Die Gruppe der Abiturientenpaare beurteilt die Items aus dem Konstrukt Tangibles durchgängig zuverlässig, während die Paare aus der niedrigsten Bildungsklasse je drei von vier Items aus den Konstrukten Tangibles und Responsiveness zuverlässig beurteilt. Die Werte sind in der Summe wenig befriedigend. Daher soll für die SERVPERF-Indikatoren keine Aussage getroffen werden, ob eine bestimmte Bildungsklasse vergleichsweise zuverlässigere Urteile trifft.

Tabelle 4-2-4-3-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Bildungsabschluss

	Hauptschule/Realschule				Abitur				Studium			
	IRR	p	κ	p	IRR	p	κ	p	IRR	p	κ	p
<i>Aussen_1</i>	0,529*	0,005			0,586*	0			0,375	0,001	0,287	0,002
<i>Aussen_2</i>	0,837*	0			0,451*	0,006			0,358	0,001	0,232	0,022
<i>Aussen_3</i>	0,628*	0,001			0,489*	0,003			0,374	0,001	k.s.	-
<i>Innen_1</i>	0,759*	0			0,239	0,107	k.s.	-	0,214	0,036	k.s.	-
<i>Innen_2</i>	0,684*	0			0,527*	0,001			0,292	0,006	k.s.	-
<i>Innen_3</i>	0,251	0,125	k.s.	-	0,591	0			0,191	0,056	k.s.	-
<i>Innen_4</i>	neg	-	-	-	neg	-	-	-	0,36	0,001	0,316	0,005
<i>Innen_5</i>	0,637*	0,001			0,544*	0,001			0,362	0,001	0,13	0,092
<i>Kunde_1</i>	0,416*	0,025			0,750*	0			0,417*	0		
<i>Kunde_2</i>	0,407*	0,028			0,635*	0			0,595*	0		
<i>Kunde_3</i>	0,316	0,073	k.s.	-	0,444*	0,007			0,223*	0,03	0,16	0,034
<i>Kunde_4</i>	0,181	0,206	0,25	0,027	0,522*	0,002			0,449*	0		
<i>Kunde_5</i>	neg	-	-	-	0,488*	0,003			0,489*	0		
<i>Kunde_6</i>	0,314	0,074	0,182	0,13	0,714*	0			0,358	0,001	0,253	0
<i>Kunde_7</i>	0,470*	0,012			0,561*	0,001			0,518*	0		
<i>Kunde_8</i>	0,882*	0			1,000*	0			0,909*	0		
<i>Kunde_9</i>	0,407*	0,028			0,632*	0			0,505*	0		
<i>Mit_1</i>	0,405*	0,026			0,638*	0			0,367	0,001	k.s.	-
<i>Mit_2</i>	0,287	0,088	0,087	0,535	0,430*	0,009			0,161	0,089	0,141	0,061
<i>Mit_3</i>	0,671*	0			0,338	0,035	0,221	0,043	0,402*	0		
<i>Mit_4</i>	0,213	0,166	k.s.	-	0,488*	0,004			0,358	0,001	0,036	0,651
<i>Mit_5</i>	0,394	0,029	k.s.	-	0,681*	0			0,377	0,001	k.s.	-
<i>GesZufr</i>	0,691*	0			0,356	0,03	k.s.	-	0,488*	0		

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Bei den Items der Kontaktpunkt-Operationalisierung stellt sich ein differenzierteres Bild dar. Von den Paaren mit Studienabschluss wurden neun von 23 Items zuverlässig beurteilt, während es bei den Paaren mit dem niedrigsten Bildungsabschluss 14 und bei den Abiturientenpaaren 19 Items waren. Die Items mit geringer Interraterreliabilität weisen auch beim Übereinstimmungsmaß einen geringen Wert auf, so dass diese Items als eindeutig unzuverlässig beurteilt bewertet werden können. Die Gruppe mit Studienabschluss hat die Items aus dem Bereich Innenzustand und Außenzustand vollständig unzuverlässig beurteilt. In der Summe sind die Beurteilungen bei der Kontaktpunkt-Operationalisierung der Gruppe der Paare mit dem höchsten Bildungsabschluss Abitur als am zuverlässigsten zu bewerten.

4.2.4.4 Zuverlässigkeit bezüglich der Tätigkeit

Zur Ermittlung der Zuverlässigkeit unterschiedlicher Tätigkeitsgruppen von Mystery Shoppern wurde die Interraterreliabilität von Testpaaren ermittelt. Aufgrund geringer Fallzahlen schieden zwei Gruppen aus, so dass lediglich die Gruppen Schüler/Student und voll berufstätig in der Auswertung miteinander verglichen werden konnten. Die Ergebnisse sind in den Tabellen 4-2-4-4-1 und 4-2-4-4-2 dargestellt.

Tabelle 4-2-4-4-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Tätigkeit

	Schüler/Student				voll berufstätig			
	IRR	p	κ	p	IRR	p	κ	p
<i>Tang_1</i>	0,554*	0,003			0,434*	0,000		
<i>Tang_2</i>	0,760*	0,000			0,473*	0,000		
<i>Tang_3</i>	0,310	0,076	0,103	0,502	0,421*	0,000		
<i>Tang_4</i>	0,346	0,054	k.s.	-	0,248	0,005	0,041	0,476
<i>Rel_1</i>	0,287	0,094	0,268	0,058	0,315	0,000	0,184	0,004
<i>Rel_2</i>	0,116	0,301	k.s.	-	0,347	0,000	k.s.	-
<i>Rel_3</i>	0,321	0,069	k.s.	-	0,196	0,020	k.s.	-
<i>Rel_4</i>	0,313	0,075	k.s.	-	0,294	0,001	k.s.	-
<i>Rel_5</i>	0,514*	0,006			0,391	0,000	0,216	0,001
<i>Resp_1</i>	neg	-	-	-	0,157	0,053	0,100	0,070
<i>Resp_2</i>	0,498*	0,010			0,562*	0,000		
<i>Resp_3</i>	0,157	0,244	0,099	0,576	0,317	0,000	0,126	0,049
<i>Resp_4</i>	0,084	0,355	k.s.	-	0,363	0,000	0,363	0,000
<i>Ass_1</i>	0,377	0,047	k.s.	-	0,467*	0,000		
<i>Ass_2</i>	0,496*	0,012			0,282	0,001	0,158	0,007
<i>Ass_3</i>	0,451*	0,021			0,351	0,000	0,180	0,003
<i>Ass_4</i>	0,690*	0,000			0,300	0,001	0,029	0,655
<i>Emp_1</i>	0,252	0,131	k.s.	-	0,319	0,000	0,096	0,070
<i>Emp_2</i>	neg	-	-	-	0,247	0,005	k.s.	-
<i>Emp_3</i>	0,242	0,140	0,315	0,010	0,415*	0,000		
<i>Emp_4</i>	0,578*	0,003			0,379	0,000	0,175	0,002
<i>Emp_5</i>	0,367	0,048	k.s.	-	0,287	0,001	0,038	0,492

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Die Werte der SERVPERF-Operationalisierung sind wenig befriedigend. Nur acht Items der Paare bestehend aus Schülern und Studenten und nur sechs Items aus den Paaren der voll Berufstätigen weisen den erforderlichen Wert von 0,4 für die Interraterreliabilität auf. Diejenigen Items, die den Wert nicht aufweisen, haben auch beim Übereinstimmungsmaß einen geringen Wert, so dass diese Items als eindeutig unzuverlässig beurteilt eingestuft werden müssen.

Tabelle 4-2-4-4-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Tätigkeit

	Schüler/Student				voll berufstätig			
	IRR	p	κ	p	IRR	p	κ	p
<i>Aussen_1</i>	0,550*	0,004			0,469*	0,000		
<i>Aussen_2</i>	0,497*	0,008			0,366	0,000	k.s.	-
<i>Aussen_3</i>	0,325	0,066	k.s.	-	0,329	0,000	k.s.	-
<i>Innen_1</i>	0,340	0,058	-0,026	0,873	0,403*	0,000		
<i>Innen_2</i>	0,649*	0,000			0,406*	0,000		
<i>Innen_3</i>	0,467*	0,013			0,322	0,000	k.s.	-
<i>Innen_4</i>	0,088	0,349	0,077	0,718	0,170	0,039	0,117	0,192
<i>Innen_5</i>	0,287	0,094	0,233	0,054	0,449*	0,000		
<i>Kunde_1</i>	0,765*	0,000			0,553*	0,000		
<i>Kunde_2</i>	0,723*	0,000			0,626*	0,000		
<i>Kunde_3</i>	0,651*	0,000			0,356	0,000	0,197	0,001
<i>Kunde_4</i>	0,551*	0,003			0,556*	0,000		
<i>Kunde_5</i>	0,630*	0,001			0,413*	0,000		
<i>Kunde_6</i>	0,666*	0,000			0,487*	0,000		
<i>Kunde_7</i>	0,555*	0,003			0,575*	0,000		
<i>Kunde_8</i>	0,901*	0,000			0,918*	0,000		
<i>Kunde_9</i>	0,696*	0,000			0,556*	0,000		
<i>Mit_1</i>	0,384	0,036	k.s.	-	0,460*	0,000		
<i>Mit_2</i>	0,278	0,101	k.s.	-	0,353	0,000	0,175	0,003
<i>Mit_3</i>	0,353	0,051	0,058	0,647	0,584*	0,000		
<i>Mit_4</i>	0,468*	0,015			0,494*	0,000		
<i>Mit_5</i>	0,635*	0,001			0,511*	0,000		
<i>GesZufr</i>	0,509*	0,008			0,457*	0,000		

* akzeptabler Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Bei den Items der Kontaktpunkt-Operationalisierung wurden in beiden Gruppen deutlich mehr Items zuverlässig beurteilt. Eine Unterscheidung der Gruppen auf Basis von mehr zuverlässig beurteilten Items ist allerdings nicht auszumachen. Lediglich ein Item wurde bei den voll berufstätigen Paaren zuverlässiger beurteilt. Insofern ist eine Unterscheidung der Gruppen bezüglich der Zuverlässigkeit nicht zu treffen.

4.2.4.5 Zuverlässigkeit bezüglich der spezifischen Restaurant-Affinität

Zur Ermittlung der Zuverlässigkeit bezüglich der Restaurant-Affinität der Mystery Shopper wurde die Interraterreliabilität für unterschiedliche Affinitätsausprägungen für die einzelnen Items berechnet.⁴⁴⁴ Die Tabellen 4-2-4-5-1 und 4-2-4-5-2 geben die jeweiligen Werte für die SERVPERF- und Kontaktpunkt-Operationalisierungen wieder.

⁴⁴⁴ Die Restaurant-Affinität wurde mit dem Item „Und wie gern essen Sie in diesem Fast Food-Restaurant?“ erfragt. Dabei war nicht die spezielle Filiale sondern irgendeine Filiale dieser Kette gemeint. Die Einstellungen „sehr gern“ und „eher gern“ wurden in der Ausprägung „hoch“, die Einstellungen „weniger gern“ und „gar nicht gern“ wurden in der Ausprägung „niedrig“ zusammengefasst, während die Einstellung „teils-teils“ die Ausprägung „mittel“ bildet.

Tabelle 4-2-4-5-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Restaurant-Affinität

	niedrig				mittel				hoch			
	IRR	p	κ	p	IRR	p	κ	p	IRR	p	κ	p
<i>Tang_1</i>	0,796*	0,000			0,574*	0,000			0,434*	0,008		
<i>Tang_2</i>	0,666*	0,000			0,419*	0,010			0,344	0,030	0,032	0,734
<i>Tang_3</i>	0,568*	0,003			0,438*	0,007			0,414*	0,011		
<i>Tang_4</i>	0,567*	0,003			0,235	0,103	k.s.	-	0,324	0,041	k.s.	-
<i>Rel_1</i>	0,490*	0,013			neg	-	-	-	0,489*	0,003		
<i>Rel_2</i>	0,539*	0,006			0,024	0,449	k.s.	-	0,247	0,096	0,262	0,057
<i>Rel_3</i>	0,375	0,048	k.s.	-	neg	-	-	-	0,581*	0,000		
<i>Rel_4</i>	0,021	0,464	-0,100	0,555	0,375	0,019	0,204	0,126	0,475*	0,004		
<i>Rel_5</i>	0,518*	0,008			0,130	0,257	k.s.	-	0,648*	0,000		
<i>Resp_1</i>	0,488*	0,013			0,212	0,131	0,137	0,182	0,219	0,128	0,172	0,103
<i>Resp_2</i>	0,897*	0,000			0,574*	0,000			0,592*	0,000		
<i>Resp_3</i>	0,633*	0,001			0,490*	0,003			0,317	0,042	k.s.	-
<i>Resp_4</i>	0,628*	0,001			0,459*	0,005			0,214	0,129	0,455*	0,000
<i>Ass_1</i>	0,609*	0,001			0,608*	0,000			0,219	0,123	k.s.	-
<i>Ass_2</i>	0,282	0,104	k.s.	-	0,100	0,297	k.s.	-	0,541*	0,001		
<i>Ass_3</i>	0,633*	0,001			0,636*	0,000			0,015	0,468	-0,028	0,841
<i>Ass_4</i>	0,485*	0,012			0,390	0,017	0,155	0,224	0,134	0,241	k.s.	-
<i>Emp_1</i>	0,433*	0,023			neg	-	-	-	0,253	0,086	-0,011	0,911
<i>Emp_2</i>	0,309	0,082	k.s.	-	0,108	0,283	k.s.	-	Neg	-	-	-
<i>Emp_3</i>	0,365	0,049	k.s.	-	0,344	0,030	k.s.	-	0,516*	0,002		
<i>Emp_4</i>	0,644*	0,001			0,001	0,496	k.s.	-	0,426*	0,010		
<i>Emp_5</i>	0,668*	0,000			0,062	0,370	k.s.	-	0,289	0,062	k.s.	-

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Den Werten der SERVPERF-Operationalisierung ist zu entnehmen, dass 17 der 22 Items bei den Paaren mit einer niedrigen spezifischen Restaurant-Affinität den erforderlichen Wert von mindestens 0,4 aufweisen. Die Paare mit mittlerer Affinität erreichen bei acht und die Paare mit hoher Affinität bei zehn Items den erforderlichen Wert. Items, bei denen der Wert unterschritten wird, sind bis auf eine Ausnahme (*Resp_4* bei hoher Affinität) aufgrund des niedrigen Wertes für das Übereinstimmungsmaß als eindeutig unzuverlässig beurteilt einzustufen. Insofern ist für die SERVPERF-Indikatoren festzustellen, dass Mystery Shopper-Paare mit niedriger Affinität für die getestete Restaurantkette häufiger eine zuverlässige Beurteilung abgeben als Paare mit mittlerer oder hoher Affinität.

Tabelle 4-2-4-5-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Restaurant-Affinität

	niedrig				mittel				hoch			
	IRR	p	κ	p	IRR	p	κ	p	IRR	p	κ	p
<i>Aussen_1</i>	0,644*	0,001			0,577*	0,000			0,316	0,045	k.s.	-
<i>Aussen_2</i>	0,639*	0,001			0,522*	0,001			0,320	0,041	k.s.	-
<i>Aussen_3</i>	0,585*	0,002			0,318	0,042	k.s.	-	0,458*	0,005		
<i>Innen_1</i>	0,286	0,106	k.s.	-	0,333	0,034	0,193	0,215	0,238	0,100	k.s.	-
<i>Innen_2</i>	neg	-	-	-	0,404*	0,012			0,444*	0,006		
<i>Innen_3</i>	0,181	0,212	k.s.	-	0,362	0,023	0,000	1,000	0,663*	0,000		
<i>Innen_4</i>	0,303	0,087	k.s.	-	0,404*	0,012			0,060	0,375	0,189	0,242
<i>Innen_5</i>	0,615*	0,001			0,210	0,130	k.s.	-	0,435*	0,007		
<i>Kunde_1</i>	0,612*	0,001			0,677*	0,000			0,156	0,202	0,017	0,882
<i>Kunde_2</i>	0,519*	0,007			0,423*	0,010			0,691*	0,000		
<i>Kunde_3</i>	0,477*	0,013			0,580*	0,000			0,261	0,079	0,046	0,667
<i>Kunde_4</i>	0,476*	0,015			0,334	0,037	k.s.	-	0,278	0,066	0,273	0,027
<i>Kunde_5</i>	0,543*	0,005			0,334	0,037	k.s.	-	0,365	0,022	0,204	0,134
<i>Kunde_6</i>	0,389	0,038	k.s.	-	0,471*	0,004			0,370	0,021	0,159	0,122
<i>Kunde_7</i>	0,504*	0,009			0,732*	0,000			0,373	0,020	0,009	0,941
<i>Kunde_8</i>	0,655*	0,001			0,912*	0,000			0,924*	0,000		
<i>Kunde_9</i>	0,760*	0,000			0,521*	0,002			0,265	0,076	k.s.	-
<i>Mit_1</i>	0,732*	0,000			0,622*	0,000			0,323	0,039	0,137	0,302
<i>Mit_2</i>	0,625*	0,001			0,414*	0,012			0,114	0,271	k.s.	-
<i>Mit_3</i>	0,463*	0,018			0,668*	0,000			0,378	0,019	k.s.	-
<i>Mit_4</i>	0,797*	0,000			0,455*	0,005			0,481*	0,004		
<i>Mit_5</i>	0,737*	0,000			0,681*	0,000			0,495*	0,002		
<i>GesZufr</i>	0,913*	0,000			0,369	0,021	k.s.	-	0,358	0,025	k.s.	-

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Die Feststellung bei den SERVPERF-Indikatoren spiegeln zum Teil auch die Werte der Kontaktpunkt-Operationalisierung wider. 18 Items wurden von den Paaren mit niedriger Affinität zuverlässig beurteilt, während es bei den Paaren mit mittlerer Affinität 16 und bei den Paaren mit hoher Affinität acht Items waren. Insofern ist für die Kontaktpunkt-Operationalisierung festzustellen, dass Mystery Shopper-Paare mit hoher Affinität zu der getesteten Restaurantkette seltener zuverlässig beurteilen als Paare mit niedriger oder mittlerer Affinität.

4.2.4.6 Zuverlässigkeit bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs

Zur Ermittlung der Komponente Zuverlässigkeit bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs der Mystery Shopper wurde die Interraterreliabilität für unterschiedliche Häufigkeitssausprägungen von Besuchen in einem Fast Food-Restaurant miteinander verglichen.⁴⁴⁵ Die Ergebnisse sind in den Tabellen 4-2-4-6-1 und 4-2-4-6-2 dargestellt.

⁴⁴⁵ Die Häufigkeit eines Fast Food-Restaurantbesuchs wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit unterschiedlicher Fast Food-Restaurants addiert und in zwei Klassen zusammengefasst wurden.

Tabelle 4-2-4-6-1: SERVPERF-Operationalisierung: Zuverlässigkeit bezüglich Fast Food-Häufigkeit

	1-12 mal pro Jahr				13-52 mal pro Jahr			
	IRR	p	κ	p	IRR	p	κ	p
<i>Tang_1</i>	0,596*	0,000			0,454*	0,001		
<i>Tang_2</i>	0,599*	0,000			0,513*	0,000		
<i>Tang_3</i>	0,523	0,000			0,458*	0,001		
<i>Tang_4</i>	0,567*	0,000			0,149	0,174	-0,009	0,920
<i>Rel_1</i>	0,155	0,122	k.s.	-	0,509*	0,000		
<i>Rel_2</i>	0,121	0,188	0,163	0,061	0,144	0,180	0,101	0,308
<i>Rel_3</i>	0,293	0,013	0,169	0,078	0,236	0,064	k.s.	-
<i>Rel_4</i>	0,168	0,106	k.s.	-	0,409*	0,003		
<i>Rel_5</i>	0,300	0,013	k.s.	-	0,687*	0,000		
<i>Resp_1</i>	0,462*	0,000			0,427*	0,002		
<i>Resp_2</i>	0,662*	0,000			0,475*	0,001		
<i>Resp_3</i>	0,504*	0,000			0,334	0,015	0,071	0,495
<i>Resp_4</i>	0,484*	0,000			0,557*	0,000		
<i>Ass_1</i>	0,521*	0,000			0,669*	0,000		
<i>Ass_2</i>	0,123	0,178	k.s.	-	0,450*	0,001		
<i>Ass_3</i>	0,408*	0,001			0,413*	0,003		
<i>Ass_4</i>	0,459*	0,000			0,335	0,015	k.s.	-
<i>Emp_1</i>	0,478*	0,000			0,353	0,010	0,218	0,006
<i>Emp_2</i>	0,179	0,088	k.s.	-	neg	-	-	-
<i>Emp_3</i>	0,419*	0,001			0,390	0,005	0,262	0,002
<i>Emp_4</i>	0,454*	0,000			0,372	0,007	0,081	0,350
<i>Emp_5</i>	0,586*	0,000			0,285	0,033	-0,041	0,640

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Die Werte aus der SERVPERF-Operationalisierung geben ein interessantes Bild wieder: Während 15 Items von den Testerpaaren mit einer Besuchshäufigkeit von 1-12 mal pro Jahr zuverlässig beurteilt wurden, sind es bei den Paaren, die 13-52 mal im Jahr in ein Fast Food-Restaurant gehen, zwölf Items. Interessant ist, dass die Items aus dem Bereich Reliability von den Paaren mit der niedrigen Besuchshäufigkeit vollständig unzuverlässig beurteilt wurden. Würde man diese aus der Betrachtung herausnehmen, ließe sich die Aussage treffen, dass diese Gruppe bei den SERVPERF-Indikatoren in der Summe häufiger zuverlässige Beurteilungen getätigt hat.

Tabelle 4-2-4-6-2: Kontaktpunkt-Operationalisierung: Zuverlässigkeit bezüglich Fast Food-Häufigkeit

	1-12 mal pro Jahr				13-52 mal pro Jahr			
	IRR	p	κ	p	IRR	p	κ	p
<i>Aussen_1</i>	0,593*	0,000			0,530	0,000		
<i>Aussen_2</i>	0,615*	0,000			0,254	0,051	k.s.	-
<i>Aussen_3</i>	0,484*	0,000			0,404*	0,004		
<i>Innen_1</i>	0,453*	0,000			0,486*	0,000		
<i>Innen_2</i>	0,529*	0,000			0,363	0,009	k.s.	-
<i>Innen_3</i>	0,270	0,020	k.s.	-	0,280	0,037	k.s.	-
<i>Innen_4</i>	0,453*	0,000			0,116	0,234	0,055	0,706
<i>Innen_5</i>	0,633*	0,000			0,404*	0,004		
<i>Kunde_1</i>	0,632*	0,000			0,616*	0,000		
<i>Kunde_2</i>	0,709*	0,000			0,622*	0,000		
<i>Kunde_3</i>	0,411*	0,001			0,462*	0,001		
<i>Kunde_4</i>	0,700*	0,000			0,377	0,007	0,226	0,010
<i>Kunde_5</i>	0,315	0,008	k.s.	-	0,343	0,014	k.s.	-
<i>Kunde_6</i>	0,548*	0,000			0,411*	0,003		
<i>Kunde_7</i>	0,647*	0,000			0,483*	0,001		
<i>Kunde_8</i>	0,834*	0,000			1,000*	0,000		
<i>Kunde_9</i>	0,757*	0,000			0,603*	0,000		
<i>Mit_1</i>	0,585*	0,000			0,591*	0,000		
<i>Mit_2</i>	0,499*	0,000			0,187	0,116	k.s.	-
<i>Mit_3</i>	0,790*	0,000			0,333	0,015	k.s.	-
<i>Mit_4</i>	0,561*	0,000			0,465*	0,001		
<i>Mit_5</i>	0,653*	0,000			0,535*	0,000		
<i>GesZufr</i>	0,695*	0,000			0,517*	0,000		

* akzeptabler Wert

neg. = negativer Wert

k.s. = keine symmetrische Kreuztabelle

Quelle: Eigene Darstellung

Bei Betrachtung der Kontaktpunkt-Items bestätigt sich dieser Trend. Während nur zwei Items von der Gruppe mit der niedrigeren Besuchshäufigkeit unzuverlässig beurteilt wurden, sind es bei der Gruppe mit der höheren Besuchshäufigkeit neun Items. Insofern ist für die Kontaktpunkt-Operationalisierung festzustellen, dass Mystery Shopper-Paare mit einer niedrigeren Besuchshäufigkeit in einem Fast Food-Restaurant häufiger zuverlässig urteilen.

4.2.4.7 Zusammenfassung der Ergebnisse der Komponente Zuverlässigkeit

Für die Komponente Zuverlässigkeit ist zu konstatieren, dass in allen sechs untersuchten Gruppen Unterschiede zwischen den jeweiligen Merkmalsträgern festzustellen sind. Abbildung 4-2-4-1 gibt den relativen Anteil zuverlässig beurteilter Items für SERVPERF- und Kontaktpunkt-Operationalisierung sowie für die Gesamtbetrachtung der untersuchten Merkmalsausprägungen wieder.

Abbildung 4-2-4-1: Relativer Anteil zuverlässig beurteilter Items



Quelle: Eigene Darstellung

In der Summe sind die Urteile von Frauen häufiger zuverlässig als die Urteile von Männern. Allerdings ist die Zuverlässigkeit bei der SERVPERF-Operationalisierung mit nur der Hälfte als zuverlässig einzustufenden Indikator-Beurteilungen als wenig befriedigend zu bewerten. Die Männerpaare bilden mit weniger als ein Drittel zuverlässig beurteilter Items das Schlusslicht der Komponente Zuverlässigkeit dieser Untersuchung. Bezogen auf das Alter, beurteilen Mystery Shopper aus der niedrigsten Altersklasse in der Summe am häufigsten zuverlässig, während Mystery Shopper aus der höchsten Altersklasse am wenigsten häufig zuverlässig beurteilen. Allerdings sind die Ergebnisse der SERVPERF-Indikatoren, die von Paaren der niedrigsten Altersklasse mit 54,5% der Indikatoren am häufigsten zuverlässig

beurteilt wurden, nur wenig befriedigend. Bezogen auf den höchsten Bildungsabschluss haben Testpaare mit Abitur in der Summe am häufigsten als zuverlässig einzustufende Urteile abgegeben, während Testerpaare mit Abitur und Paare mit Haupt- oder Realschulabschluss weniger zuverlässig beurteilt haben. Mit in der Summe nur 35,1% zuverlässig beurteilter Items sind die Paare mit abgeschlossenem Studium die zweitschlechteste Merkmalsausprägung dieser Untersuchung. Bezogen auf die Tätigkeit ist die Zuverlässigkeit von Paaren bestehend aus Schülern/Studenten und Paaren, die sich aus voll Berufstätigen zusammensetzen, auf dem gleichen niedrigen Niveau mit in der Summe nur knapp über die Hälfte an zuverlässig beurteilten Items.⁴⁴⁶ Bezogen auf die Affinität zur untersuchten Fast Food-Kette haben sich diejenigen Testerpaare als am zuverlässigsten in ihrer Beurteilung erwiesen, die die niedrigste Affinität aufweisen, während bei Paaren mit hoher Affinität eine unbefriedigende Zuverlässigkeit festgestellt wurde. Letztlich wurde bei Paaren mit niedriger Besuchshäufigkeit in einem Fast Food-Restaurant eine höhere Zuverlässigkeit ausgemacht als bei solchen Paaren mit einer hohen Besuchshäufigkeit.

4.2.5 Sorgfaltstendenz bei Mystery Shoppern

Zur Ermittlung der Sorgfaltstendenz wurden zwei Komponenten ausgewertet: Zum einen die Gewissenhaftigkeit beim Ausfüllen des Fragebogens, zum anderen die Vollständigkeit des ausgefüllten Fragebogens.

4.2.5.1 Gewissenhaftigkeit bei Mystery Shoppern

Die Gewissenhaftigkeit des Ausfüllens wurde ermittelt, indem überprüft wurde, ob ein inhaltlich gleiches Item, das an unterschiedlichen Stellen im Fragebogen positioniert wurde, auch zu gleichen Antworten führte.⁴⁴⁷ Die Komponente Gewissenhaftigkeit von Merkmalsträgern ist umso höher einzustufen, je weniger (mehr) ungleiche (gleiche) Antworten gegeben wurden. Es wurde über alle Fälle festgestellt, dass 27,7% der Mystery Shopper für die inhaltlich gleiche Fragen unterschiedliche Antworten geben.⁴⁴⁸

Nachfolgend werden für die einzelnen Gruppen die jeweiligen Merkmalsausprägungen bezüglich der Gewissenhaftigkeit gegenüber gestellt.

⁴⁴⁶ Einschränkung ist anzumerken, dass mangels ausreichender Fallzahlen bei den Tätigkeitsgruppen Teilzeit berufstätig und Hausfrau/Rentner lediglich die Tätigkeitsgruppen Schüler/Student und Vollzeit berufstätig miteinander verglichen werden konnten.

⁴⁴⁷ Dies waren die Items C.IV.1. „Die Mitarbeiter wirken ordentlich“ und D.I.3. „Die Angestellten machen einen ordentlichen Eindruck“

⁴⁴⁸ Der (abgestufte) Grad der Übereinstimmung im Sinne der oben ausgeführten Beurteilerübereinstimmung wurde nicht ermittelt. Wie oben ausgeführt, ist zur Bestimmung der Gewissenhaftigkeit eine exakte Übereinstimmung bei inhaltlich gleichen Items erforderlich.

4.2.5.1.1 Gewissenhaftigkeit bezüglich des Geschlechtes der Mystery Shopper

Zur Ermittlung der Komponente Gewissenhaftigkeit bezüglich des Geschlechts der Mystery Shopper wurden Frauen und Männer hinsichtlich gleicher und ungleicher Beantwortung der inhaltlich identischen Items verglichen. Das Ergebnis ist in Tabelle 4-2-5-1-1 dargestellt.

Tabelle 4-2-5-1-1: Komponente Gewissenhaftigkeit bezüglich Geschlecht

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Geschlecht	Männlich	Anzahl	49	144	193
		% von Geschlecht	25,4%	74,6%	100,0%
	Weiblich	Anzahl	55	128	183
		% von Geschlecht	30,1%	69,9%	100,0%
Gesamt		Anzahl	104	272	376
		% von Geschlecht	27,7%	72,3%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Gewissenhaftigkeit bezogen auf das Geschlecht zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Signifikanztest von Fisher mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei Männern und Frauen nicht signifikant verschieden.

4.2.5.1.2 Gewissenhaftigkeit bezüglich des Alters der Mystery Shopper

Basierend auf den Altersangaben der Mystery Shopper wurden drei Altersklassen für die Auswertung der Komponente Gewissenhaftigkeit gebildet. Die ursprüngliche Einteilung in vier Altersklassen erschien bei näherer Betrachtung der Fallzahlen nicht sinnvoll.⁴⁴⁹ Die Ergebnisse sind in Tabelle 4-2-5-1-2-1 dargestellt.

⁴⁴⁹ Siehe Tabelle 3-3-4 für die Anzahl der Fälle bei einer Aufteilung in vier Altersklassen.

Tabelle 4-2-5-1-2-1: Komponente Gewissenhaftigkeit bezüglich Altersklasse

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Altersklasse	18-30 Jahre	Anzahl	33	69	102
		% von Altersklasse	32,4%	67,6%	100,0%
	31-45 Jahre	Anzahl	57	149	206
		% von Altersklasse	27,7%	72,3%	100,0%
	älter als 45 Jahre	Anzahl	14	54	68
		% von Altersklasse	20,6%	79,4%	100,0%
Gesamt		Anzahl	104	272	376
		% von Altersklasse	27,7%	72,3%	100,0%

Quelle: Eigene Darstellung

Die dargestellten Ergebnisse veranlassen zu der Annahme, dass ein Zusammenhang zwischen Alter der Mystery Shopper und der identischen Beantwortung inhaltlich gleiche Items, die an unterschiedlichen Stellen im Beobachtungsbogen positioniert sind, besteht. Der Zusammenhang zwischen Alter und gleich/ungleich beantworteten Items soll mit Spearmans r_s überprüft werden. Dafür werden die von den Mystery Shoppern gemachten Altersangaben verwendet und nicht in drei Klassen eingeteilt. Das Ergebnis ist in Tabelle 4-2-5-1-2-2 dargestellt.

Tabelle 4-2-5-1-2-2: Korrelation Alter/Item gleich oder ungleich beantwortet

			Alter	Item gleich oder ungleich beantwortet
Spearman-Rho	Alter	Korrelationskoeffizient	1,000	,114(*)
		Sig. (1-seitig)	.	,014
		N	376	376
	Item gleich oder ungleich beantwortet	Korrelationskoeffizient	,114(*)	1,000
		Sig. (1-seitig)	,014	.
		N	376	376

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).

Quelle: Eigene Darstellung

Die Korrelation zwischen Alter und gleich oder ungleich beantwortetem Item ist mit 0,114 bei $p < 0,05$ als schwach einzustufen. Daher kann nur ein geringer Zusammenhang nachgewiesen werden.

4.2.5.1.3 Gewissenhaftigkeit bezüglich des höchsten Bildungsabschlusses der Mystery Shopper

Analog der Überprüfung der Gewissenhaftigkeit bezüglich des Alters der Mystery Shopper wurden bei der Ermittlung der Gewissenhaftigkeit bezüglich des höchsten Bildungsabschlusses in Ermanglung ausreichender Fallzahlen Hauptschul- und Realschulabschluss zu einer Gruppe zusammengeführt. Die Ergebnisse gibt Tabelle 4-2-5-1-3 wieder.

Tabelle 4-2-5-1-3: Komponente Gewissenhaftigkeit bezüglich höchstem Bildungsabschluss

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Höchster Bildungsabschluss	Hauptschule/Realschule	Anzahl	18	58	76
		% von Höchster Bildungsabschluss	23,7%	76,3%	100,0%
	Abitur	Anzahl	29	74	103
		% von Höchster Bildungsabschluss	28,2%	71,8%	100,0%
	Studium	Anzahl	57	140	197
		% von Höchster Bildungsabschluss	28,9%	71,1%	100,0%
Gesamt		Anzahl	104	272	376
		% von Höchster Bildungsabschluss	27,7%	72,3%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Gewissenhaftigkeit bezogen auf den höchsten Bildungsabschluss zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei unterschiedlich hohem Bildungsabschluss nicht signifikant verschieden.

4.2.5.1.4 Gewissenhaftigkeit bezüglich der Tätigkeit der Mystery Shopper

Die nachfolgende Tabelle 4-2-5-1-4 stellt die Ausprägung der Komponente Gewissenhaftigkeit für unterschiedliche Tätigkeiten der Mystery Shopper gegenüber.

Tabelle 4-2-5-1-4: Komponente Gewissenhaftigkeit bezüglich Tätigkeit

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Tätigkeit	Schüler/Student	Anzahl	16	35	51
		% von Tätigkeit	31,4%	68,6%	100,0%
	Hausfrau/Rentner	Anzahl	8	23	31
		% von Tätigkeit	25,8%	74,2%	100,0%
	Voll berufstätig	Anzahl	74	181	255
		% von Tätigkeit	29,0%	71,0%	100,0%
	Teilzeit berufstätig	Anzahl	6	33	39
		% von Tätigkeit	15,4%	84,6%	100,0%
Gesamt		Anzahl	104	272	376
		% von Tätigkeit	27,7%	72,3%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Gewissenhaftigkeit bezogen auf die Tätigkeit zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei unterschiedlicher Tätigkeit nicht signifikant verschieden.

4.2.5.1.5 Gewissenhaftigkeit bezüglich der Erfahrung als Mystery Shopper

Analog der Überprüfung der Gewissenhaftigkeit bezüglich des Alters sowie des höchsten Bildungsabschlusses der Mystery Shopper wurden bei der Ermittlung der Gewissenhaftigkeit bezüglich der Testerfahrung in Ermanglung ausreichender Fallzahlen Gruppen zusammengeführt. Die Fälle wurden in Mystery Shopper mit und ohne Erfahrung gruppiert. Die Ergebnisse gibt Tabelle 4-2-5-1-5 wieder.

Tabelle 4-2-5-1-5: Komponente Gewissenhaftigkeit bezüglich Mystery Shopping-Erfahrung

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Mystery Shopping-Erfahrung	keine Erfahrung vorhanden	Anzahl	81	196	277
		% von Mystery Shopping-Erfahrung	29,2%	70,8%	100,0%
	Erfahrung vorhanden	Anzahl	23	76	99
		% von Mystery Shopping-Erfahrung	23,2%	76,8%	100,0%
Gesamt		Anzahl	104	272	376
		% von Mystery Shopping-Erfahrung	27,7%	72,3%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Gewissenhaftigkeit bezogen auf die Erfahrung als Mystery Shopper zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Signifikanztest von Fisher mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei Mystery Shoppern mit oder ohne Testerfahrung nicht signifikant verschieden. Ermittelt man zudem eine Korrelation zwischen der exakten Anzahl vergangener Testeinsätze, die von den Mystery Shoppern angegeben wurden, mit gleich und ungleich beantworteten Items, so ist keine signifikante Korrelation nach Spearman festzustellen.

4.2.5.1.6 Gewissenhaftigkeit bezüglich der Fast Food-Affinität der Mystery Shopper

Zur Ermittlung der Komponente Gewissenhaftigkeit bezüglich der Fast Food-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen hinsichtlich gleicher und ungleicher Beantwortung der inhaltlich identischen Items verglichen.⁴⁵⁰ Das Ergebnis ist in Tabelle 4-2-5-1-6 dargestellt.

Tabelle 4-2-5-1-6: Komponente Gewissenhaftigkeit bezüglich Fast Food-Affinität

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Fast Food-Affinität	Sehr gern	Anzahl	4	15	19
		% von Fast Food-Affinität	21,1%	78,9%	100,0%
	Eher gern	Anzahl	24	49	73
		% von Fast Food-Affinität	32,9%	67,1%	100,0%
	Teils-teils	Anzahl	48	115	163
		% von Fast Food-Affinität	29,4%	70,6%	100,0%
	Weniger gern	Anzahl	16	66	82
		% von Fast Food-Affinität	19,5%	80,5%	100,0%
	Gar nicht gern	Anzahl	12	26	38
		% von Fast Food-Affinität	31,6%	68,4%	100,0%
Gesamt		Anzahl	104	271	375
		% von Fast Food-Affinität	27,7%	72,3%	100,0%

Quelle: Eigene Darstellung

Der bei Betrachtung der Werte zu vermutende Zusammenhang zwischen Fast Food-Affinität der eingesetzten Mystery Shopper und Gewissenhaftigkeit erweist sich nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei unterschiedlicher Fast Food-Affinität nicht signifikant verschieden.

⁴⁵⁰ Die Fast Food-Affinität wurde mit dem Item „Losgelöst von Ihrer heutigen Aufgabe als Testbesucher, wie gern essen Sie in einem Fast Food Restaurant?“ erfragt.

4.2.5.1.7 Gewissenhaftigkeit bezüglich der Restaurant-Affinität der Mystery Shopper

Zur Ermittlung der Komponente Gewissenhaftigkeit bezüglich der Restaurant-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen hinsichtlich gleicher und ungleicher Beantwortung der inhaltlich identischen Items verglichen.⁴⁵¹ Das Ergebnis ist in Tabelle 4-2-5-1-7 dargestellt.

Tabelle 4-2-5-1-7: Komponente Gewissenhaftigkeit bezüglich Restaurant-Affinität

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Restaurant-Affinität	Sehr gern	Anzahl	4	20	24
		% von Restaurant-Affinität	16,7%	83,3%	100,0%
	Eher gern	Anzahl	33	66	99
		% von Restaurant - Affinität	33,3%	66,7%	100,0%
	Teils-teils	Anzahl	40	103	143
		% von Restaurant - Affinität	28,0%	72,0%	100,0%
	Weniger gern	Anzahl	15	59	74
		% von Restaurant - Affinität	20,3%	79,7%	100,0%
	Gar nicht gern	Anzahl	10	21	31
		% von Restaurant - Affinität	32,3%	67,7%	100,0%
Gesamt		Anzahl	102	269	371
		% von Restaurant - Affinität	27,5%	72,5%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Gewissenhaftigkeit bezogen auf die Restaurant-Affinität des Testobjekts zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei unterschiedlicher Affinität zu dem getesteten Restaurant nicht signifikant verschieden.

4.2.5.1.8 Gewissenhaftigkeit bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs

Zur Ermittlung der Komponente Gewissenhaftigkeit bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs der Mystery Shopper wurden unterschiedliche

⁴⁵¹ Die Restaurant-Affinität wurde mit dem Item „Und wie gern essen Sie in diesem Fast Food-Restaurant?“ erfragt. Dabei war nicht die spezielle Filiale sondern irgendeine Filiale dieser Kette gemeint.

Häufigkeitssausprägungen hinsichtlich gleicher und ungleicher Beantwortung der inhaltlich identischen Items verglichen.⁴⁵² Das Ergebnis ist in Tabelle 4-2-5-1-8 dargestellt.

Tabelle 4-2-5-1-8: Komponente Gewissenhaftigkeit bezüglich Häufigkeit eines Fast Food-Restaurantbesuchs

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Fast Food-Restaurantbesuch - Häufigkeitslevel	keine Besuche	Anzahl	3	11	14
		% von FastFood Häufigkeitslevel	21,4%	78,6%	100,0%
	1-12 mal im Jahr	Anzahl	54	132	186
		% von FastFood Häufigkeitslevel	29,0%	71,0%	100,0%
	13-52 mal im Jahr	Anzahl	41	109	150
		% von FastFood Häufigkeitslevel	27,3%	72,7%	100,0%
	mehr als 52 mal im Jahr	Anzahl	6	20	26
		% von FastFood Häufigkeitslevel	23,1%	76,9%	100,0%
Gesamt		Anzahl	104	272	376
		% von FastFood Häufigkeitslevel	27,7%	72,3%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Gewissenhaftigkeit bezogen auf Häufigkeit eines Fast Food-Restaurantbesuchs zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei unterschiedlicher Häufigkeit eines Fast Food-Restaurantbesuchs nicht signifikant verschieden. Ermittelt man zudem eine Korrelation zwischen der exakten Anzahl von Fast Food-Restaurantbesuchen mit gleich und ungleich beantworteten Items, ist keine signifikante Korrelation nach Spearman festzustellen.

4.2.5.1.9 Gewissenhaftigkeit bezüglich der spezifischen Restaurant-Besuchshäufigkeit

Zur Ermittlung der Komponente Gewissenhaftigkeit bezüglich der Häufigkeit eines Besuchs in einer Filiale der getesteten Fast Food-Kette wurden unterschiedliche Häufigkeitssausprägungen hinsichtlich gleicher und ungleicher Beantwortung der inhaltlich identischen Items verglichen.⁴⁵³ Das Ergebnis ist in Tabelle 4-2-5-1-9 dargestellt.

⁴⁵² Die Häufigkeit eines Fast Food-Restaurantbesuchs wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit unterschiedlicher Fast Food-Restaurants addiert und in unterschiedlichen Klassen zusammengefasst wurden.

⁴⁵³ Die Häufigkeit eines Besuchs in dem getesteten Restaurant wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit in dem spezifischen Fast Food-Restaurant addiert und in unterschiedlichen Klassen zusammengefasst wurden.

Tabelle 4-2-5-1-9: Komponente Gewissenhaftigkeit bezüglich der spezifischen Restaurant-Besuchshäufigkeit

			Item gleich oder ungleich beantwortet		Gesamt
			ungleich beantwortet	gleich beantwortet	
Restaurantbesuch in einem Restaurant der Kette - Häufigkeitslevel	keine Besuche	Anzahl	7	29	36
		% von Häufigkeitslevel	19,4%	80,6%	100,0%
	1-12 mal im Jahr	Anzahl	73	182	255
		% von Häufigkeitslevel	28,6%	71,4%	100,0%
	13-52 mal im Jahr	Anzahl	23	57	80
		% von Häufigkeitslevel	28,8%	71,3%	100,0%
	mehr als 52 mal im Jahr	Anzahl	1	4	5
		% von Häufigkeitslevel	20,0%	80,0%	100,0%
Gesamt		Anzahl	104	272	376
		% von Häufigkeitslevel	27,7%	72,3%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Gewissenhaftigkeit bezogen auf Häufigkeit eines Besuchs in einer Filiale der getesteten Fast Food-Kette zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei unterschiedlicher Besuchshäufigkeit in der getesteten Restaurantkette nicht signifikant verschieden. Ermittelt man zudem eine Korrelation zwischen der exakten Besuchsanzahl in einer Filiale der getesteten Fast Food-Kette mit gleich und ungleich beantworteten Items so ist keine signifikante Korrelation nach Spearman festzustellen.

4.2.5.1.10 Zusammenfassung der Ergebnisse der Gewissenhaftigkeit

Es ist zu konstatieren, dass die in diesem Abschnitt aufgeführten Eigenschaften der Tester keinen Rückschluss auf die Gewissenhaftigkeit des Ausfüllens, welche durch die identische Beantwortung von zwei inhaltlich gleichen Items ermittelt wurde, ergeben. Es bleibt daher offen, ob spezifische Mystery Shopper-Eigenschaften existieren, die ein gewissenhaftes Ausfüllen des Fragebogens determinieren.

4.2.5.2 Vollständigkeit des Ausfüllens bei Mystery Shoppern

Die Vollständigkeit des Ausfüllens wurde ermittelt, indem überprüft wurde, ob Mystery Shopper beim Ausfüllen des Beobachtungsbogens Items ausgelassen haben. Die Komponente Vollständigkeit von Merkmalsträgern ist umso höher (niedriger) einzustufen, je weniger (mehr) Items von einer Gruppe ausgelassen wurden. Über alle Fälle wurde festgestellt, dass 18,1% der Mystery Shopper Items ausgelassen haben.

Nachfolgend werden für die einzelnen Gruppen die jeweiligen Merkmalsausprägungen bezüglich der Komponente Vollständigkeit gegenüber gestellt.

4.2.5.2.1 Vollständigkeit bezüglich des Geschlechtes der Mystery Shopper

Zur Ermittlung der Komponente Vollständigkeit bezüglich des Geschlechts der Mystery Shopper wurden Frauen und Männer hinsichtlich ausgelassener Items verglichen. Das Ergebnis ist in Tabelle 4-2-5-2-1 dargestellt.

Tabelle 4-2-5-2-1: Komponente Vollständigkeit bezüglich Geschlecht

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
Geschlecht	Männlich	Anzahl	31	162	193
		% von Geschlecht	16,1%	83,9%	100,0%
	Weiblich	Anzahl	37	146	183
		% von Geschlecht	20,2%	79,8%	100,0%
Gesamt		Anzahl	68	308	376
		% von Geschlecht	18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Vollständigkeit bezogen auf das Geschlecht zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Signifikanztest von Fisher mit $p > 0,05$ als statistisch insignifikant. Insofern ist das vollständige Ausfüllen bei Männern und Frauen nicht signifikant verschieden.

4.2.5.2.2 Vollständigkeit bezüglich des Alters der Mystery Shopper

Basierend auf Altersangaben der Mystery Shopper wurden drei Altersklassen für die Auswertung der Komponente Vollständigkeit gebildet. Die ursprüngliche Einteilung in vier Altersklassen erschien bei näherer Betrachtung der Fallzahlen nicht sinnvoll.⁴⁵⁴ Die Ergebnisse sind in Tabelle 4-2-5-2-2-1 dargestellt.

⁴⁵⁴ Siehe Tabelle 3-3-4 für die Anzahl der Fälle bei einer Aufteilung in vier Altersklassen.

Tabelle 4-2-5-2-2-1: Komponente Vollständigkeit bezüglich Altersklasse

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
Altersklasse	18-30 Jahre	Anzahl	13	89	102
		% von Altersklasse	12,7%	87,3%	100,0%
	31-45 Jahre	Anzahl	40	166	206
		% von Altersklasse	19,4%	80,6%	100,0%
	älter als 45 Jahre	Anzahl	15	53	68
		% von Altersklasse	22,1%	77,9%	100,0%
Gesamt		Anzahl	68	308	376
		% von Altersklasse	18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Die dargestellten Ergebnisse veranlassen zu der Annahme, dass ein Zusammenhang zwischen Alter der Mystery Shopper und der Komponente Vollständigkeit besteht. Der Zusammenhang zwischen Alter und vollständigen Ausfüllen des Fragebogens soll mit Spearmans r_s überprüft werden. Dafür werden die von den Mystery Shoppern gemachten Altersangaben verwendet und nicht in drei Klassen eingeteilt. Ferner wird die exakte Anzahl der ausgelassenen Items zur Berechnung verwendet. Das Ergebnis ist in Tabelle 4-2-5-2-2-2 dargestellt.

Tabelle 4-2-5-2-2-2: Korrelation Alter/Items ausgelassen oder vollständig ausgefüllt

			Alter	Anzahl ausgelassener Items
Spearman-Rho	Alter	Korrelationskoeffizient	1,000	,089(*)
		Sig. (1-seitig)	.	,042
		N	376	376
	Anzahl ausgelassener Items	Korrelationskoeffizient	,089(*)	1,000
		Sig. (1-seitig)	,042	.
		N	376	376

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).

Quelle: Eigene Darstellung

Die Korrelation zwischen Alter und Anzahl ausgelassener Items ist mit 0,089 bei $p < 0,05$ als schwach einzustufen. Daher kann nur ein geringer Zusammenhang nachgewiesen werden.

4.2.5.2.3 Vollständigkeit bezüglich des höchsten Bildungsabschlusses der Mystery Shopper

Analog der Überprüfung der Gewissenhaftigkeit bezüglich des Bildungsstands der Mystery Shopper wurden bei der Ermittlung der Komponente Vollständigkeit bezüglich des höchsten Bildungsabschlusses in Ermanglung ausreichender Fallzahlen Hauptschul- und Realschulabschluss zu einer Gruppe zusammengeführt. Die Ergebnisse gibt Tabelle 4-2-5-2-3 wieder.

Tabelle 4-2-5-2-3: Komponente Vollständigkeit bezüglich höchstem Bildungsabschluss

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
Höchster Bildungsabschluss	Hauptschule/Realschule	Anzahl	14	62	76
		% von Höchster Bildungsabschluss	18,4%	81,6%	100,0%
	Abitur	Anzahl	19	84	103
		% von Höchster Bildungsabschluss	18,4%	81,6%	100,0%
	Studium	Anzahl	35	162	197
		% von Höchster Bildungsabschluss	17,8%	82,2%	100,0%
Gesamt		Anzahl	68	308	376
		% von Höchster Bildungsabschluss	18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Die ermittelten Werte bei der Komponente Vollständigkeit bezogen auf den höchsten Bildungsabschluss sind sehr ähnlich. Untersucht man die Werte auf Unterschiedlichkeit erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die Vollständigkeit des Ausfüllens bei unterschiedlich hohem Bildungsabschluss nicht signifikant verschieden.

4.2.5.2.4 Vollständigkeit bezüglich der Tätigkeit der Mystery Shopper

Die nachfolgende Tabelle 4-2-5-2-4 stellt die Ausprägung der Komponente Gewissenhaftigkeit für unterschiedliche Tätigkeiten der Mystery Shopper gegenüber.

Tabelle 4-2-5-2-4: Komponente Vollständigkeit bezüglich Tätigkeit

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
Tätigkeit	Schüler/Student	Anzahl	5	46	51
		% von Tätigkeit	9,8%	90,2%	100,0%
	Hausfrau/Rentner	Anzahl	7	24	31
		% von Tätigkeit	22,6%	77,4%	100,0%
	Voll berufstätig	Anzahl	47	208	255
		% von Tätigkeit	18,4%	81,6%	100,0%
	Teilzeit berufstätig	Anzahl	9	30	39
		% von Tätigkeit	23,1%	76,9%	100,0%
Gesamt		Anzahl	68	308	376
		% von Tätigkeit	18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Vollständigkeit bezogen auf die Tätigkeit zu vermuten. Allerdings erweist sich der Zusammenhang nach dem

Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die Vollständigkeit des Ausfüllens bei unterschiedlicher Tätigkeit nicht signifikant verschieden.

4.2.5.2.5 Vollständigkeit bezüglich der Erfahrung als Mystery Shopper

Analog der Überprüfung der Gewissenhaftigkeit bezüglich der Erfahrung als Mystery Shopper wurden bei der Ermittlung der Vollständigkeit bezüglich der Testerfahrung in Ermanglung ausreichender Fallzahlen Gruppen zusammengeführt. Die Fälle wurden in Mystery Shopper mit und ohne Erfahrung gruppiert. Die Ergebnisse gibt Tabelle 4-2-5-2-5 wieder.

Tabelle 4-2-5-2-5: Komponente Vollständigkeit bezüglich Mystery Shopping-Erfahrung

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
Mystery Shopping-Erfahrung	keine Erfahrung vorhanden	Anzahl	50	227	277
		% von Mystery Shopping-Erfahrung	18,1%	81,9%	
	Erfahrung vorhanden	Anzahl	18	81	99
		% von Mystery Shopping-Erfahrung	18,2%	81,8%	100,0%
Gesamt		Anzahl	68	308	376
		% von Mystery Shopping-Erfahrung	18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Die ermittelten Werte bei der Komponente Vollständigkeit bezogen auf die Mystery Shopping-Erfahrung sind sehr ähnlich. Untersucht man die Werte auf Unterschiedlichkeit erweist sich der Zusammenhang nach dem Signifikanztest von Fisher mit $p > 0,05$ als statistisch insignifikant. Insofern ist die Vollständigkeit des Ausfüllens bei unterschiedlich hohem Bildungsabschluss nicht signifikant verschieden. Ermittelt man zudem eine Korrelation zwischen der exakten Anzahl vergangener Testeinsätze, die von den Mystery Shoppern angegeben wurden, mit der Vollständigkeit des Fragebogens, so ist keine signifikante Korrelation nach Spearman festzustellen.

4.2.5.2.6 Vollständigkeit bezüglich der Fast Food-Affinität der Mystery Shopper

Zur Ermittlung der Komponente Vollständigkeit bezüglich der Fast Food-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen hinsichtlich der Vollständigkeit verglichen.⁴⁵⁵ Das Ergebnis ist in Tabelle 4-2-5-2-6 dargestellt.

⁴⁵⁵ Die Fast Food-Affinität wurde mit dem Item „Losgelöst von Ihrer heutigen Aufgabe als Testbesucher, wie gern essen Sie in einem Fast Food Restaurant?“ erfragt.

Tabelle 4-2-5-2-6: Komponente Vollständigkeit bezüglich Fast Food-Affinität

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
Fast Food-Affinität	Sehr gern	Anzahl	0	19	19
		% von Fast Food-Affinität	,0%	100,0%	100,0%
	Eher gern	Anzahl	14	59	73
		% von Fast Food-Affinität	19,2%	80,8%	100,0%
	Teils-teils	Anzahl	25	138	163
		% von Fast Food-Affinität	15,3%	84,7%	100,0%
	Weniger gern	Anzahl	23	59	82
		% von Fast Food-Affinität	28,0%	72,0%	100,0%
	Gar nicht gern	Anzahl	6	32	38
		% von Fast Food-Affinität	15,8%	84,2%	100,0%
Gesamt	Anzahl		68	307	375
	% von Fast Food-Affinität		18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Der bei Betrachtung der Werte zu vermutende Zusammenhang zwischen Fast Food-Affinität der eingesetzten Mystery Shopper und Vollständigkeit erweist sich nach dem Chi-Quadrat-Test mit $p=0,030$ als statistisch signifikant. Insofern ist die Vollständigkeit des Ausfüllens bei unterschiedlicher Fast Food-Affinität signifikant verschieden. Allerdings ist diese Erkenntnis aus praktischer Perspektive wenig befriedigend, da aus den Ergebnissen kein Trend abzuleiten ist (etwa, je höher die Affinität, desto weniger wurden Items ausgelassen). Daher ist es nicht erstaunlich, dass sich mit einer punktserialen Korrelation der Werte kein signifikanter Zusammenhang nachweisen lässt.

4.2.5.2.7 Vollständigkeit bezüglich der Restaurant-Affinität der Mystery Shopper

Zur Ermittlung der Komponente Vollständigkeit bezüglich der Restaurant-Affinität der Mystery Shopper wurden unterschiedliche Affinitätsausprägungen hinsichtlich des Auslassverhaltens bei den Items verglichen.⁴⁵⁶ Das Ergebnis ist in Tabelle 4-2-5-2-7 dargestellt.

⁴⁵⁶ Die Restaurant-Affinität wurde mit dem Item „Und wie gern essen Sie in diesem Fast Food-Restaurant?“ erfragt. Dabei war nicht die spezielle Filiale sondern irgendeine Filiale dieser Kette gemeint.

Tabelle 4-2-5-2-7: Komponente Vollständigkeit bezüglich Restaurant-Affinität

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
Restaurant-Affinität	Sehr gern	Anzahl	2	22	24
		% von Restaurant-Affinität	8,3%	91,7%	100,0%
	Eher gern	Anzahl	16	83	99
		% von Restaurant-Affinität	16,2%	83,8%	100,0%
	Teils-teils	Anzahl	27	116	143
		% von Restaurant-Affinität	18,9%	81,1%	100,0%
	Weniger gern	Anzahl	17	57	74
		% von Restaurant-Affinität	23,0%	77,0%	100,0%
	Gar nicht gern	Anzahl	5	26	31
		% von Restaurant-Affinität	16,1%	83,9%	100,0%
Gesamt	Anzahl		67	304	371
	% von Restaurant-Affinität		18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Vollständigkeit bezogen auf die Restaurant-Affinität des Testobjekts zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die Vollständigkeit des Ausfüllens bei unterschiedlicher Affinität zu dem getesteten Restaurant nicht signifikant verschieden.

4.2.5.2.8 Vollständigkeit bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs

Zur Ermittlung der Komponente Vollständigkeit bezüglich der Häufigkeit eines Fast Food-Restaurantbesuchs der Mystery Shopper wurden unterschiedliche Häufigkeitssausprägungen hinsichtlich des Ausfüllverhaltens gegenüber gestellt.⁴⁵⁷ Das Ergebnis ist in Tabelle 4-2-5-2-8 dargestellt.

⁴⁵⁷ Die Häufigkeit eines Fast Food-Restaurantbesuchs wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit unterschiedlicher Fast Food-Restaurants addiert und in unterschiedlichen Klassen zusammengefasst wurden.

Tabelle 4-2-5-2-8: Komponente Vollständigkeit bezüglich Häufigkeit eines Fast Food-Restaurantbesuchs

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
FastFood Häufigkeitslevel	keine Besuche	Anzahl	2	12	14
		% von FastFood Häufigkeitslevel	14,3%	85,7%	100,0%
	1-12 mal im Jahr	Anzahl	42	144	186
		% von FastFood Häufigkeitslevel	22,6%	77,4%	100,0%
	13-52 mal im Jahr	Anzahl	22	128	150
		% von FastFood Häufigkeitslevel	14,7%	85,3%	100,0%
	mehr als 52 mal im Jahr	Anzahl	2	24	26
		% von FastFood Häufigkeitslevel	7,7%	92,3%	100,0%
Gesamt		Anzahl	68	308	376
		% von FastFood Häufigkeitslevel	18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Vollständigkeit bezogen auf Häufigkeit eines Fast Food-Restaurantbesuchs zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist Vollständigkeit des Ausfüllens bei unterschiedlicher Häufigkeit eines Fast Food-Restaurantbesuchs nicht signifikant verschieden. Ermittelt man zudem eine Korrelation zwischen der exakten Anzahl von Fast Food-Restaurantbesuchen mit der Vollständigkeit des Ausfüllens ist nach Spearman nur eine schwache signifikante Korrelation ($r_s = -0,125$ bei $p = 0,008$) festzustellen.

4.2.5.2.9 Vollständigkeit bezüglich der spezifischen Restaurant-Besuchshäufigkeit

Zur Ermittlung der Komponente Vollständigkeit bezüglich der Häufigkeit eines Besuchs in einer Filiale der getesteten Fast Food-Kette wurden unterschiedliche Häufigkeitssausprägungen hinsichtlich des vollständigen Ausfüllens des Beobachtungsbogens verglichen.⁴⁵⁸ Das Ergebnis ist in Tabelle 4-2-5-2-9 dargestellt.

⁴⁵⁸ Die Häufigkeit eines Besuchs in dem getesteten Restaurant wurde ermittelt, indem die Angaben der Mystery Shopper bezüglich der Besuchshäufigkeit in dem spezifischen Fast Food-Restaurant addiert und in unterschiedlichen Klassen zusammengefasst wurden.

Tabelle 4-2-5-2-9: Komponente Vollständigkeit bezüglich der spezifischen Restaurant-Besuchshäufigkeit

			Items ausgelassen		Gesamt
			Items ausgelassen	Vollständig ausgefüllt	
Restaurantbesuch in einer Filiale der Kette – Häufigkeitslevel	keine Besuche	Anzahl	11	25	36
		% von Häufigkeitslevel	30,6%	69,4%	100,0%
	1-12 mal im Jahr	Anzahl	46	209	255
		% von Häufigkeitslevel	18,0%	82,0%	100,0%
	13-52 mal im Jahr	Anzahl	11	69	80
		% von Häufigkeitslevel	13,8%	86,3%	100,0%
mehr als 52 mal im Jahr	Anzahl	0	5	5	
	% von Häufigkeitslevel	,0%	100,0%	100,0%	
Gesamt		Anzahl	68	308	376
		% von Häufigkeitslevel	18,1%	81,9%	100,0%

Quelle: Eigene Darstellung

Anhand der Werte ist zunächst eine Unterschiedlichkeit bei der Komponente Gewissenhaftigkeit bezogen auf Häufigkeit eines Besuchs in einer Filiale der getesteten Fast Food-Kette zu vermuten. Allerdings erweist sich der Zusammenhang nach dem Chi-Quadrat-Test mit $p > 0,05$ als statistisch insignifikant. Insofern ist die gleiche Beantwortung inhaltlich gleicher Items bei unterschiedlicher Besuchshäufigkeit in der getesteten Restaurantkette nicht signifikant verschieden. Ermittelt man zudem eine Korrelation zwischen der exakten Anzahl von Besuchen in einer Filiale der getesteten Fast Food-Kette mit der Vollständigkeit des Ausfüllens ist nach Spearman nur eine schwache signifikante Korrelation ($r_s = -0,140$ bei $p = 0,003$) festzustellen.

4.2.5.2.10 Zusammenfassung der Ergebnisse der Vollständigkeit

Es ist zu konstatieren, dass die in diesem Abschnitt aufgeführten Eigenschaften der Tester weitgehend keinen Rückschluss auf die Vollständigkeit des Ausfüllens ergeben. Lediglich die Fast Food-Affinität ergab einen signifikanten Unterschied, der allerdings keine Einstufung zulässt. Es bleibt daher offen, ob spezifische Mystery Shopper-Eigenschaften existieren, die ein vollständiges Ausfüllen des Fragebogens determinieren. Dieser Schluss ist für beide Komponenten der als Sorgfaltstendenz bezeichneten Komponente zu ziehen.

4.3 Kernaussagen der ermittelten Ergebnisse bezogen auf die Merkmalsausprägungen von Mystery Shoppern

Nach Darlegung der Ergebnisse der unterschiedlichen Determinanten der Ergebnisqualität im vorangegangenen Abschnitt sollen in diesem Abschnitt die Kernaussagen der Untersuchungsergebnisse bezogen auf die einzelnen Gruppen wiedergegeben werden. Dabei werden insbesondere Unterschiede zwischen den Merkmalsausprägungen adressiert.

4.3.1 Ergebnisqualität bezogen auf das Geschlecht der Mystery Shopper

Bis auf eine Komponente wurden zwischen Männern und Frauen hinsichtlich ihrer Ergebnisqualität keine signifikanten Unterschiede festgestellt. Unterschiede bestehen bei der Komponente Zuverlässigkeit. Während Männerpaare in der Summe 31,1% der Items zuverlässig beurteilt haben, waren es bei Frauen 60%. Insofern besteht auf Basis der Ergebnisse dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Männern und Frauen.

4.3.2 Ergebnisqualität bezogen auf das Alter der Mystery Shopper

Zur Ermittlung der Ergebnisqualität von Mystery Shoppern unterschiedlichen Alters wurden die drei Altersklassen 18-30 Jahre, 31-45 Jahr und >45 Jahre gebildet. Mit Ausnahme bei den Komponenten Gewissenhaftigkeit und Sorgfältigkeit wurden Unterschiede zwischen den Klassen festgestellt. Die höchste Altersklasse weist eine Tendenz zur Milde auf, die insbesondere in den Bereichen Responsiveness und Empathy bei der SERVPERF-Operationalisierung und in den Bereichen Kundengespräch und Mitarbeiter bei der Kontaktpunkt-Operationalisierung auftritt. Ferner wurde für diese Gruppe eine Tendenz zur Einheitlichkeit im Urteil festgestellt, die insbesondere in den Bereichen Responsiveness und Empathy auftritt. Bezüglich der Korrelationstendenz wurde in dieser Untersuchung festgestellt, dass die Bereiche Assurance und Außenzustand bei der niedrigsten und mittleren Altersklasse einen signifikant positiven Einfluss auf die Gesamtzufriedenheit haben. Daneben hat die Variable Reliability bei der mittleren Altersklasse einen positiven Einfluss auf die Gesamtzufriedenheit. In der Summe sind die Urteile der niedrigsten Altersklasse am häufigsten und in der höchsten Altersklasse am wenigsten häufig zuverlässig. Insofern besteht auf Basis der Ergebnisse dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Mystery Shoppern aus unterschiedlichen Altersklassen.

4.3.3 Ergebnisqualität bezogen auf den höchsten Bildungsabschluss der Mystery Shopper

Zur Ermittlung der Ergebnisqualität von Mystery Shoppern bezogen auf den höchsten Bildungsabschluss wurden die drei Klassen Hauptschule/Realschule, Abitur und Studium miteinander verglichen. Unterschiede waren nur bei den Komponenten Korrelationstendenz

und Zuverlässigkeit festzustellen. Im Rahmen der Untersuchung der Korrelationstendenz wurde festgestellt, dass die Variable Empathy bei der Bildungsklasse Abitur einen signifikant positiven Einfluss auf die Gesamtzufriedenheit hat, der in den anderen Klassen jeweils nicht vorliegt. Gleiches wurde für die Variable Reliability für die Gruppe mit abgeschlossenem Studium ermittelt. Ferner wurde ermittelt, dass die Variable Außenzustand bei der Bildungsklasse Abitur keinen signifikant positiven Einfluss auf die Gesamtzufriedenheit hat. Für die Komponente Zuverlässigkeit wurde in dieser Untersuchung festgestellt, dass die Urteile der Gruppe mit dem höchsten Bildungsabschluss Abitur im Vergleich zu den anderen beiden Gruppen in der Summe am häufigsten als zuverlässig bewertet werden können. Ferner wurden die Urteile von Mystery Shoppern mit abgeschlossenem Studium als am wenigsten zuverlässig eingestuft. Sie bilden das Schlusslicht aller betrachteten Ausprägungen bei der Komponente Zuverlässigkeit. Insofern besteht auf Basis der Ergebnisse dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Mystery Shoppern mit unterschiedlich hohem Bildungsabschluss.

4.3.4 Ergebnisqualität bezogen auf die Tätigkeit der Mystery Shopper

Die Gruppen Schüler/Student, Hausfrau/Rentner, Teilzeit berufstätig und Vollzeit berufstätig wurden zur Ermittlung der Ergebnisqualität miteinander verglichen. Mit Ausnahme bei den Komponenten Gewissenhaftigkeit und Sorgfältigkeit wurden Unterschiede zwischen den Gruppen festgestellt. Die Vollzeit Berufstätigen weisen im Vergleich zu Schülern/Studenten in den Bereichen Tangibles und Assurance eine Tendenz zur Strenge auf. Im Vergleich zur Gruppe Hausfrau/Rentner ist die Tendenz zur Strenge in den Bereichen Responsiveness, Assurance, Empathy, Außenzustand und Zustand der Innenräume auszumachen. Die Gruppe Hausfrau/Rentner bewertet im Vergleich zur Gruppe Vollzeit berufstätig in den Bereichen Responsiveness, Assurance, Empathy und Zustand der Innenräume sowie im Bereich Responsiveness im Vergleich zur Gruppe Schüler/Studenten einheitlicher. Die Variablen Tangibles, Reliability, Außenzustand und Mitarbeiter haben bei den Vollzeit Berufstätigen einen signifikant positiven Einfluss auf die Gesamtzufriedenheit und bei Schülern/Studenten keinen signifikanten Einfluss. Bei der der Betrachtung der Verlässlichkeit wurden nur die Gruppen Schüler/Student und voll berufstätig ausgewertet, wobei kein Unterschied ausgemacht werden konnte. Gleichwohl wurden in beiden Gruppen in der Summe nur knapp über die Hälfte der Items zuverlässig beurteilt. Insofern besteht auf Basis der Ergebnisse dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Mystery Shoppern aus unterschiedlichen Tätigkeitsgruppen.

4.3.5 Ergebnisqualität bezogen auf die Erfahrung als Mystery Shopper

Zur Ermittlung der Ergebnisqualität bezogen auf die Erfahrung als Mystery Shopper wurden die einzelnen Komponenten für solche Mystery Shopper mit Testerfahrung mit denen ohne

Erfahrung verglichen.⁴⁵⁹ Unterschiede wurden nur bei einer der ausgewerteten Komponenten festgestellt. Bei der Korrelationstendenz haben die Variablen Assurance, Innenzustand und Kundengespräch bei der Gruppe ohne Mystery Shopping-Erfahrung einen signifikant positiven Einfluss auf die Gesamtzufriedenheit, während dieser Einfluss bei der Gruppe mit Erfahrung bei den Variablen Reliability, Empathy und Mitarbeiter vorliegt. Insofern besteht auf Basis der Ergebnisse dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Mystery Shoppern mit und ohne Testerfahrung.

4.3.6 Ergebnisqualität bezogen auf die Fast Food-Affinität der Mystery Shopper

Zur Ermittlung der Ergebnisqualität bezogen auf die Fast Food-Affinität der Mystery Shopper wurden bis auf die Komponente Zuverlässigkeit alle Komponenten ausgewertet. Mit Ausnahme der Komponente Sorgfaltstendenz konnten Unterschiede zwischen den Affinitätsgruppen ausgemacht werden. Mystery Shoppern mit einer hohen Fast Food-Affinität konnte in dieser Untersuchung eine Tendenz zur Strenge sowie eine Tendenz zur Einheitlichkeit im Gesamtzufriedenheitsurteil nachgewiesen werden. Entsprechend wurde Mystery Shoppern mit einer niedrigen Fast Food-Affinität eine Tendenz zur Milde sowie eine Tendenz zur Uneinheitlichkeit im Gesamtzufriedenheitsurteil nachgewiesen. Die Variable Assurance hat bei niedriger und hoher Affinität einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Die Variable Reliability hat bei mittlerer Affinität einen signifikant positiven Einfluss auf die Gesamtzufriedenheit und die Variable Empathy bei niedriger Affinität. Ein weiterer Unterschied besteht bei der Variablen Innenzustand, die bei hoher Affinität keinen Einfluss auf die Gesamtzufriedenheit und bei niedriger und mittlerer Affinität einen positiven Einfluss hat. Bezogen auf die Vollständigkeit des Ausfüllens wurden signifikante Unterschiede zwischen den Klassen festgestellt, allerdings ist keine Ableitung einer Tendenz möglich. Insofern besteht auf Basis der Ergebnisse dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Mystery Shoppern mit unterschiedlicher Fast Food-Affinität.

4.3.7 Ergebnisqualität bezogen auf die Restaurant-Affinität der Mystery Shopper

Bei der Ermittlung der Ergebnisqualität bezogen auf die Restaurant-Affinität⁴⁶⁰ der Mystery Shopper wurden alle Komponenten ausgewertet. Bis auf die Komponenten Sorgfaltstendenz und Vollständigkeit konnten Unterschiede zwischen den Affinitätsgruppen ausgemacht werden. Mystery Shoppern mit einer hohen Restaurant-Affinität konnte in dieser Untersuchung eine Tendenz zur Strenge sowie eine Tendenz zur Einheitlichkeit im Gesamtzufriedenheitsurteil nachgewiesen werden. Entsprechend wurde Mystery Shoppern

⁴⁵⁹ Die Komponente Zuverlässigkeit wurde nicht ausgewertet.

⁴⁶⁰ Die Restaurant-Affinität wurde mit dem Item „Und wie gern essen Sie in diesem Fast Food-Restaurant?“⁴⁶⁰ erfragt. Dabei war nicht die spezielle Filiale sondern irgendeine Filiale dieser Kette gemeint.

mit einer niedrigen Restaurant-Affinität eine Tendenz zur Milde sowie eine Tendenz zur Uneinheitlichkeit im Gesamtzufriedenheitsurteil nachgewiesen. Die Variablen Reliability und Assurance haben bei hoher Affinität einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Die Variable Außenzustand hat bei niedriger Affinität einen höheren Einfluss auf die Gesamtzufriedenheit als bei hoher Affinität. Bei der Untersuchung der Komponente Zuverlässigkeit konnte festgestellt werden, dass Mystery Shopper mit einer niedrigen Restaurant Affinität häufiger eine zuverlässige Beurteilung abgeben und im Gegensatz dazu Mystery Shopper mit einer hohen Restaurant Affinität weniger häufig eine zuverlässige Beurteilung abgeben. Insofern besteht auf Basis der Ergebnisse dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Mystery Shoppern mit unterschiedlicher Restaurant-Affinität.

4.3.8 Ergebnisqualität bezogen auf die Fast Food-Häufigkeit der Mystery Shopper

Bei der Ermittlung der Ergebnisqualität bezogen auf die Fast Food-Häufigkeit der Mystery Shopper wurden bei zwei Komponenten Unterschiede zwischen den Affinitätsgruppen ausgemacht. Die Variablen Reliability und Assurance haben bei hoher Besuchshäufigkeit in einem Fast Food-Restaurant einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Bei der Komponente Zuverlässigkeit konnte in dieser Untersuchung festgestellt werden, dass Mystery Shopper, die 1-12 Mal im Jahr in ein Fast Food-Restaurant gehen, häufiger eine zuverlässige Beurteilung abgeben als solche, die 13-52 Mal im Jahr ein Fast Food-Restaurant besuchen. Eine Ausnahme bilden die Reliability Items der SERVPERF-Operationalisierung. Insofern besteht auf Basis der Ergebnisse dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Mystery Shoppern mit unterschiedlicher Besuchshäufigkeit in einem Fast Food-Restaurant.

4.3.9 Ergebnisqualität bezogen auf die spezifische Restaurant-Besuchshäufigkeit

Bei der Ermittlung der Ergebnisqualität bezogen auf die Restaurant-Besuchshäufigkeit⁴⁶¹ der Mystery Shopper wurden bis auf die Komponente Zuverlässigkeit alle Komponenten ausgewertet. Bei Mystery Shoppern, die im vergangenen Jahr keine Restaurantfiliale der Kette besucht haben, wurde in dieser Untersuchung eine Tendenz zur Milde und eine Tendenz zur Einheitlichkeit in den Bereichen Assurance und Responsiveness festgestellt. Ferner wurde bei der Korrelationstendenz festgestellt, dass Reliability und Außenzustand bei niedriger Besuchshäufigkeit einen signifikant positiven Einfluss auf die Gesamtzufriedenheit haben. Daneben wurde ermittelt, dass der Innenzustand bei der Gruppe mit hoher Besuchshäufigkeit einen höheren Stellenwert hat. Insofern besteht auf Basis der Ergebnisse

⁴⁶¹ Dabei ist mit spezifischer Restaurant-Besuchshäufigkeit nicht die spezielle Filiale sondern irgendeine Filiale dieser Kette gemeint.

dieser Untersuchung ein Unterschied bei der Ergebnisqualität zwischen Mystery Shoppern mit unterschiedlicher Besuchshäufigkeit in einer Filiale der getesteten Fast Food-Kette.

In Tabelle 4-3-1 sind die Kernaussagen der Ergebnisse dieser Untersuchung zusammengefasst. Somit ist die Ergebnisdarstellung abgeschlossen und das vierte Teilziel der Arbeit wurde erreicht. Das letzte Kapitel liefert eine Zusammenfassung der Untersuchung, zeigt die Limitationen der Arbeit auf und gibt Anregungen für zukünftige Forschungsbemühungen.

Tabelle 4-3-1: Übersicht der Kernaussagen der empirischen Befunde

	Geschlecht	Altersklassen	höchster Bildungsabschluss	Tätigkeit
<i>Merkmalsausprägungen</i>	<i>weiblich männlich</i>	<i>18-30 Jahre 31-45 >45</i>	<i>Hauptschule/ Realschule Abitur Studium</i>	<i>Schüler/Student Hausfrau/Rentner Teilzeit berufstätig Vollzeit berufstätig</i>
Mittelwertstendenz	keine signifikanten Unterschiede	>45 Jahre: Tendenz zur Milde, insbesondere bei Responsiveness und Empathy sowie Kundengespräch und Mitarbeiter	keine signifikanten Unterschiede	Vollzeit berufstätig: Tendenz zur Strenge im Vergleich zu Schüler/ Student (Tangibles, Assurance) sowie zu Hausfrau/ Rentner (Responsiveness, Assurance, Empathy, Außenzustand, Zustand der Innenräume)
Streuungstendenz	keine signifikanten Unterschiede	>45 Jahre: Tendenz zur Einheitlichkeit im Urteil, insbesondere bei Responsiveness und Empathy	keine signifikanten Unterschiede	Hausfrau/Rentner: Einheitlichere Bewertung im Vergleich zu Vollzeit berufstätig (Responsiveness, Assurance, Empathy, Zustand der Innenräume) sowie zu Schüler/ Student (Responsiveness)
Korrelationstendenz	Empathy hat bei den Frauen und Assurance sowie Reliability bei den Männern einen sign. positiven Einfluss auf die Gesamtzufriedenheit, der in der anderen Gruppe jeweils nicht vorliegt	Assurance und Außenzustand haben bei der niedrigsten und mittleren Altersklasse und Reliability bei der mittleren Altersklasse einen sign. positiven Einfluss auf die Gesamtzufriedenheit	Bei der Bildungsstufe Abitur (Studium) hat Empathy (Reliability) einen sign. positiven Einfluss auf die Gesamtzufriedenheit, der in den anderen Klassen nicht vorliegt. Außenzustand hat bei Abitur keinen sign. Einfluss.	Tangibles, Reliability, Außenzustand und Mitarbeiter haben bei Vollzeit Berufstätigen einen sign. positiven Einfluss auf die Gesamtzufriedenheit und bei Schülern/Studenten keinen sign. Einfluss.
Zuverlässigkeit	In der Summe sind die Urteile von Frauen häufiger zuverlässig. Allerdings ist die Zuverlässigkeit bei der SERVPERF-Operationalisierung mit nur der Hälfte als zuverlässig einzustufenden Indikator-Beurteilungen als wenig befriedigend zu bewerten. Männer bilden das Schlusslicht bei der Komponente Zuverlässigkeit.	In der Summe sind bei der SERVPERF-Operationalisierung die Urteile von Mystery Shoppern aus der niedrigsten Altersklasse häufiger zuverlässig. Allerdings sind hier nur die Hälfte der Indikatoren als zuverlässig bewertet einzustufen, was wenig befriedigend ist.	In der Summe sind die Urteile der Gruppe mit dem höchsten Bildungsabschluss Abitur am häufigsten als zuverlässig zu bewerten. Allerdings ist die Zuverlässigkeit bei der SERVPERF-Operationalisierung mit nur der Hälfte als zuverlässig einzustufenden Indikator-Beurteilungen als wenig befriedigend zu bewerten. Die Urteile von Mystery Shoppern mit abgeschlossenem Studium sind am wenigsten zuverlässig.	Nur Auswertung Schüler/Student und voll berufstätig: keine Unterscheidung möglich. Die SERVPERF-Items sind als nur wenig zuverlässig beurteilt zu bewerten.
Gewissenhaftigkeit	keine signifikanten Unterschiede	keine signifikanten Unterschiede	keine signifikanten Unterschiede	keine signifikanten Unterschiede
Vollständigkeit	keine signifikanten Unterschiede	keine signifikanten Unterschiede	keine signifikanten Unterschiede	keine signifikanten Unterschiede

	Mystery Shopping-Erfahrung	Fast Food-Affinität	Restaurant-Affinität	Besuchshäufigkeit Fast Food-Restaurant	Besuchshäufigkeit bei getesteter Kette
<i>Merkmalsausprägungen</i>	<i>keine Erfahrung Erfahrung</i>	<i>sehr gern eher gern teils-teils weniger gern gar nicht gern</i>	<i>sehr gern eher gern teils-teils weniger gern gar nicht gern</i>	<i>kein Mal im letzten Jahr 1-12 Mal 13-52 Mal >52 Mal</i>	<i>kein Mal im letzten Jahr 1-12 Mal 13-52 Mal >52 Mal</i>
Mittelwertstendenz	keine signifikanten Unterschiede	hohe Affinität: Tendenz zur Strenge beim Gesamtzufriedenheitsurteil; niedrige Affinität: Tendenz zur Milde beim Gesamt-zufriedenheitsurteil	hohe Affinität: Tendenz zur Strenge beim Gesamt-zufriedenheitsurteil; niedrige Affinität: Tendenz zur Milde beim Gesamtzufriedenheitsurteil	keine signifikanten Unterschiede	kein Besuch: Tendenz zur Milde in den Bereichen Assurance und Responsiveness
Streuungstendenz	keine signifikanten Unterschiede	hohe Affinität: Tendenz zur Einheitlichkeit beim Gesamt-zufriedenheitsurteil; niedrige Affinität: Tendenz zur Uneinheitlichkeit beim Gesamtzufriedenheitsurteil	hohe Affinität: Tendenz zur Einheitlichkeit beim Gesamt-zufriedenheitsurteil; niedrige Affinität: Tendenz zur Uneinheitlichkeit beim Gesamtzufriedenheitsurteil	keine signifikanten Unterschiede	kein Besuch: Tendenz zur Einheitlichkeit in den Bereichen Assurance und Responsiveness
Korrelationstendenz	Assurance und Innenzustand (Reliability und Empathy) haben bei der Gruppe ohne Erfahrung (mit Erfahrung) einen signifikant positiven Einfluss auf die Gesamtzufriedenheit. Das Kundengespräch (die Mitarbeiter) hat bei der Gruppe ohne Erfahrung (mit Erfahrung) einen höheren Stellenwert.	Assurance hat bei niedriger und hoher Affinität einen sign. positiven Einfluss auf die Gesamtzufriedenheit. Reliability (Empathy) hat bei mittlerer (niedriger) Affinität einen sign. positiven Einfluss auf die Gesamtzufriedenheit. Der Innenzustand hat bei hoher (niedriger und mittlerer) keinen (einen positiven) sign. Einfluss auf die Gesamtzufriedenheit.	Reliability und Assurance haben bei hoher Affinität einen sign. positiven Einfluss auf die Gesamtzufriedenheit. Außenzustand hat bei niedriger Affinität einen höheren Einfluss auf die Gesamtzufriedenheit als bei hoher Affinität.	Reliability und Assurance haben bei hoher Besuchshäufigkeit einen sign. positiven Einfluss auf die Gesamtzufriedenheit.	Reliability und Außenzustand haben bei niedriger Besuchshäufigkeit einen sign. positiven Einfluss auf die Gesamtzufriedenheit. Innenzustand hat bei der Gruppe mit hoher Besuchshäufigkeit einen höheren Stellenwert.
Zuverlässigkeit	nicht ausgewertet	nicht ausgewertet	Zuverlässige Beurteilung bei niedriger Restaurant-Affinität. Wenig zuverlässige Beurteilung bei hoher Restaurant-Affinität.	Häufiger zuverlässige Beurteilung bei niedriger Besuchshäufigkeit in einem Fast Food-Restaurant (Ausnahme: Reliability Items der SERVPERF-Operationalisierung).	nicht ausgewertet
Gewissenhaftigkeit	keine signifikanten Unterschiede	keine signifikanten Unterschiede	keine signifikanten Unterschiede	keine signifikanten Unterschiede	keine signifikanten Unterschiede
Vollständigkeit	keine signifikanten Unterschiede	signifikante Unterschiede zwischen den Klassen vorhanden, allerdings ist keine Ableitung einer Tendenz möglich	keine signifikanten Unterschiede	keine signifikanten Unterschiede	keine signifikanten Unterschiede

Quelle: Eigene Darstellung

5 Schlussbetrachtung

Im Rahmen einer Schlussbetrachtung werden in diesem Kapitel die Ergebnisse zusammengefasst und die daraus resultierenden Implikationen für Unternehmen, die Mystery Shopper einsetzen, dargelegt. Ferner werden die Restriktionen der Untersuchung genannt und Ansatzpunkte für zukünftige Forschungsbemühungen zur Mystery Shopping-Thematik aufgezeigt.

5.1 Zusammenfassung der Ergebnisse und Implikationen für die Praxis

Der Ausgangspunkt der vorliegenden Arbeit war der wenig befriedigende Forschungsstand zur Mystery Shopping-Methodik. Während eine Vielzahl an praktischen, nicht theoriegeleiteten Publikationen zu finden ist, sind wissenschaftliche Beiträge rar. Diese Beiträge belegen allerdings, dass sich Mystery Shopping eignet, um theoretische und wahrnehmungsrelevante Aspekte der Dienstleistungsqualität zu erfassen und zu analysieren. Ferner setzen sich die Beiträge insbesondere mit Gütekriterien auseinander und konnten Reliabilität und Validität der Methode nachweisen. Vergangene Forschungsbemühungen ließen gleichwohl offen, ob sich unterschiedliche soziodemografische Merkmalsträger oder Personen mit unterschiedlichen Einstellungen zum Testobjekt in ihrer Ergebnisqualität unterscheiden. Die Kenntnis dieses Sachverhaltes würde Unternehmen, die Mystery Shopper einsetzen, einen Nutzen bei der Zusammensetzung einer Feldmannschaft für ein Mystery Shopping-Projekt stiften. Daher wurde dieser Fragestellung in der vorliegenden Arbeit nachgegangen.

Dafür wurde erstens die historische Entwicklung des Phänomens Mystery Shopping in Theorie und Praxis von den Ursprüngen bis zum Status Quo dargestellt. Dabei erfolgte eine Abgrenzung zu anderen Verfahren der Evaluierung der Dienstleistungsqualität sowie eine Einordnung des Verfahrens. Zweitens wurden Anforderungen und Voraussetzungen für die Konzeption einer Mystery Shopping-Untersuchung herausgearbeitet, die es ermöglicht, handlungsrelevante Informationen über die Dienstleistungsqualität zu gewinnen. Drittens wurden Determinanten der Ergebnisqualität von Testkunden hergeleitet, die Urteilstendenzen (Strenge/Milde, Einheitlichkeit und Korrelationstendenz), Zuverlässigkeit (Interraterreliabilität und Beurteilerübereinstimmung) sowie Sorgfaltstendenz (Vollständigkeit und Gewissenhaftigkeit des Ausfüllens) umfassen. Viertens wurden die einzelnen Komponenten der Ergebnisqualität unterschiedlicher soziodemografischer und einstellungsbezogener Gruppen im Rahmen einer empirischen Analyse untersucht. Als empirisches Beispiel diente ein Mystery Shopping in der Fast Food-Restaurantbranche, bei dem 188 Testpaare eingesetzt wurden.

Tabelle 5-1 liefert eine zusammenfassende Übersicht der Untersuchungshypothesen und der entsprechenden Befunde. Als Ergebnis der vorliegenden Arbeit ist zu konstatieren, dass auf

Basis des vorhandenen Datenmaterials in allen untersuchten Gruppen Unterschiede in der Ergebnisqualität zwischen den Merkmalsträgern bestehen.⁴⁶²

Tabelle 5-1: Zusammenfassende Übersicht der Untersuchungshypothesen

Hypothesen	Befund
H1: Männer und Frauen unterscheiden sich beim Mystery Shopping in ihrer Ergebnisqualität.	bestätigt
H2: Die Ergebnisqualität von Mystery Shoppern unterschiedlichen Alters unterscheidet sich.	bestätigt
H3: Die Ergebnisqualität von Mystery Shoppers mit unterschiedlichen Bildungsabschlüssen unterscheidet sich.	bestätigt
H4: Die Ergebnisqualität von Mystery Shoppern unterscheidet sich bei unterschiedlichen (beruflichen) Tätigkeiten.	bestätigt
H5: Die Ergebnisqualität von Mystery Shoppern mit unterschiedlicher Erfahrung hinsichtlich der Menge der Durchführung von Testkäufen unterscheidet sich.	bestätigt
H6: Die Ergebnisqualität von Mystery Shoppern mit unterschiedlicher Involvierung bzw. Einstellung zum Testobjekt unterscheidet sich.	bestätigt

Quelle: Eigene Darstellung

Während Urteilstendenzen die Urteilstendenzgewohnheiten der Merkmalsträger beschreiben und ein Indikator für erforderliche Schulungsmaßnahmen sind, lassen die Komponenten Zuverlässigkeit und Sorgfaltstendenz Aussagen über die Arbeitsqualität der eingesetzten Mystery Shopper zu. Die Eignung für die Tätigkeit als Mystery Shopper ergibt sich aus der Kombination der Ausprägungen von erforderlichem Schulungsbedarf und dem Niveau der Arbeitsqualität. Es ist zu betonen, dass diese Einstufung auf pragmatischen Überlegungen beruht und lediglich als Indikator für die Eignung im Rahmen der Ableitung von Implikationen für Unternehmen, die Mystery Shopper einsetzen, dienen soll. Für die Merkmalsträger aller Gruppen wurden in dieser Untersuchung Unterschiede bei einer oder mehreren Komponenten der Urteilstendenzen festgestellt. Folglich besteht bei allen Gruppen Schulungsbedarf. Da daneben bei dieser Untersuchung keine signifikanten Unterschiede bei der Sorgfaltstendenz zwischen den jeweiligen Merkmalsträgern der jeweiligen Gruppen bestehen, bleibt die Zuverlässigkeit als Indikator für die Entscheidung über die Vorteilhaftigkeit des Einsatzes einer spezifischen Merkmalsausprägung einer jeden Gruppe als Mystery Shopper. Tabelle 5-2 liefert eine Übersicht der Einstufungsrangfolge der

⁴⁶² Vgl. auch Tabelle 4-3-1 für eine Übersicht der Unterschiede

jeweiligen Merkmalsträger für diejenigen Gruppen, in denen die Zuverlässigkeit untersucht wurde.

Tabelle 5-2: Eignungseinstufung als Mystery Shopper

Gruppe	Rangfolge Merkmalsträger
Geschlecht	Frauen
	Männer
Altersklasse	18-30 Jahre
	31-45 Jahre
	>45 Jahre
Höchster Bildungsabschluss	Abitur
	Hauptschule/ Realschule
	Studium
Restaurant-Affinität	niedrig
	mittel
	hoch
Häufigkeit eines Fast Food-Restaurantbesuchs	1-12 Mal in den letzten 12 Monaten
	13-52 Mal in den letzten 12 Monaten

Quelle: Eigene Darstellung

Basierend auf den in dieser Erhebung untersuchten Gruppen ist für Unternehmen, die Mystery Shopper einsetzen, abzuleiten, dass eher Frauen als Testkunden eingesetzt werden sollten als Männer. Daneben ist auf Basis der vorliegenden Ergebnisse eine rückläufige Ergebnisqualität bei zunehmendem Alter zu unterstellen. Insofern sind jüngere Mystery Shopper älteren zu bevorzugen. Unternehmen, die vornehmlich eine ältere Zielgruppe bedienen, müssten allerdings ältere Testkunden zur Überprüfung einer Dienstleistungssituation einsetzen. Die Ergebnisse wären dann mit gebotener Skepsis zu verwenden. Beim Bildungsstand ist keine Ableitung eines eindeutigen Trends möglich. In dieser Studie erwiesen sich die Mystery Shopper mit Abitur als zuverlässiger im Vergleich zu denjenigen mit Haupt-/ Realschulabschluss oder einem abgeschlossenen Studium. Insofern besteht möglicherweise

ein Eignungskorridor, der durch einen erworbenen Bildungsabschluss jeweils nach unten und nach oben begrenzt sein könnte. Ferner sind solche Mystery Shopper einzusetzen, die eine geringere Affinität zu dem getesteten Restaurant aufweisen. Letztlich sind Mystery Shopper bevorzugt zu beauftragen, die in den vergangenen 12 Monaten weniger häufig ein Fast Food-Restaurant besucht haben. Die beiden letztgenannten Ergebnisse lassen für diese Untersuchung den Schluss zu, dass mit zunehmender Involvierung in die zu testende Dienstleistung eine Abnahme der Ergebnisqualität einhergeht. Insofern wären Mystery Shopper mit hoher Affinität von einer Untersuchung auszuschließen. Dies könnte sich bei bestimmten Untersuchungen als schwierig erweisen. Sofern ein Unternehmen beispielsweise eine Dienstleistungssituation mit Testkunden überprüfen möchte, die über eine echte Kundenhistorie verfügen, wären die gewonnenen Ergebnisse nur mit gebotener Skepsis als objektiv einzustufen. Hier sollten Kundenhistorien speziell für Testkunden angelegt werden, statt echte Kunden als Mystery Shopper einzusetzen, um die Ergebnisqualität sicherzustellen.

Für Unternehmen, die Mystery Shopping einsetzen, stiftet die vorliegende Arbeit neben der Ableitung der Eignungseinstufung weiteren Nutzen.

In der kritischen Auseinandersetzung mit der bis dato publizierten praktischen und wissenschaftlichen Literatur wurden insbesondere Umsetzungsaspekte berücksichtigt. Hierdurch wird Unternehmen eine fundierte und praktikable Vorgehensweise zur Konzeption und Durchführung von Mystery Shopping-Untersuchungen präsentiert, die eine Verwendung der Ergebnisse einschließt. Neben diesem Leitfaden ermöglicht das Mehrkomponenten-Modell eine Analyse der Ergebnisqualität der in einer Felduntersuchung erhobenen Daten. Somit gewinnen Unternehmen Sicherheit hinsichtlich der Qualität einer durchgeführten Untersuchung.

Als weitere Empfehlungen für die Durchführung von Mystery Shopping ist zu konstatieren, dass Schulungsmaßnahmen erforderlich sind. Wenngleich der Grad bisher unerforscht ist, liefert die vorliegende Arbeit ein Indiz dafür, dass lediglich schriftliche Anweisungen nicht für alle Merkmalsträger ausreichend sind. Ferner wird aus den Ergebnissen der Sorgfaltstendenz deutlich, dass der Einsatz datenverarbeitungsgestützter Beobachtungsbögen sinnvoll ist, um Vollständigkeit und Gewissenhaftigkeit beim Ausfüllen sicherzustellen. Vor dem Hintergrund der zum Teil niedrigen Werte bei der Rater-Übereinstimmung einiger Merkmalsträger in dieser Untersuchung ist zu empfehlen, nur solche Ergebnisse in eine Analyse einzubeziehen, die als zuverlässig einzustufen sind. Andernfalls kann nicht von einer Übereinstimmung mit der Realität ausgegangen werden.

5.2 Restriktionen der Untersuchung

In diesem Abschnitt werden die Einschränkungen der Arbeit, welche die Interpretationsfähigkeit der ermittelten Befunde und deren Generalisierbarkeit einschränken, aufgezeigt. Die Restriktionen ergeben sich aus dem gewählten Anwendungskontext sowie

dem gewählten Auswertungsschema. Die kritische Reflektion soll Anregungen für zukünftige Forschungsbemühungen geben.

Aus dem gewählten Untersuchungsdesign ergeben sich unterschiedliche Restriktionen: Der gewählte Anwendungskontext der Untersuchung begründet eine wesentliche Restriktion der vorliegenden Arbeit. Es wurde lediglich eine Fast Food-Kette betrachtet. In diesem Zusammenhang ist auch der Komplexitätsgrad des gewählten Dienstleistungsprozesses anzuführen, der z.B. bei einer Finanzdienstleistungsberatung höher ausfallen könnte. Folglich ist eine Übertragung der Ergebnisse auf eine andere Kette, eine andere Dienstleistungsbranche oder eine komplexere Dienstleistungssituation nur mit gebotener Skepsis zulässig. Insofern könnten zukünftige Forschungsbemühungen sich der Untersuchung anderer Branchen und anderer Dienstleistungssituationen widmen, um genauere Erkenntnisse zu gewinnen

Im Rahmen dieser Untersuchung wurden Mystery Shopper im Kontext der Überprüfung von Dienstleistungsqualität eingesetzt. Wie gezeigt wurde, können Mystery Shopper aber auch mit anderen Zielsetzungen eingesetzt werden. Insofern ist die Übertragung der Ergebnisse auf andere Einsatzbereiche gründlich zu überdenken. Auch diese Restriktion bietet Raum für zukünftige Forschungsbemühungen, die sich beispielsweise mit dem Einsatz von Mystery Shoppern im Kontext von Benchmarking oder Wettbewerbsbeobachtung auseinandersetzen.

Aus forschungsökonomischen Gründen konnten nicht alle Gruppen vollständig in die Auswertung einbezogen werden.⁴⁶³ Es bleibt daher offen, wie die Ergebnisse sich darstellen würden, wenn mit einer vorteilhafteren Struktur der Stichprobe alle Gruppen entsprechend ausgewertet werden hätten können. Ferner wurde in dieser Untersuchung immer nur eine Merkmalsausprägung betrachtet. Hierdurch verschließt sich diese Untersuchung wertvollen Erkenntnissen. Aus einer detaillierteren Analyse, in der Merkmalsausprägungen der eingesetzten Mystery Shopper kombiniert betrachtet würden, ergäbe sich möglicherweise ein differenzierteres Bild. Neben der Ermittlung des Einflusses konkreter Merkmalsausprägungen der eingesetzten Mystery Shopper auf die Ergebnisqualität ist es dabei auch interessant zu ermitteln, ob die Kombination von Merkmalsausprägungen von Mystery Shopper und Servicepersonal (z.B. unterschiedliches Geschlecht, Altersabstand) einen Einfluss auf die Ergebnisqualität hat. Insofern könnte diese Limitation dazu anregen, im Rahmen zukünftiger Forschungsbemühungen mit einer breiter angelegten Datenlage, genauere Kenntnis über die Ergebnisqualität von Mystery Shoppern zu erlangen.

Neben dem gewählten Untersuchungsdesign besteht eine weitere Restriktion dieser Arbeit im Rückgriff auf das Schrifttum der Personal- und Organisationspsychologie sowie der Einbringung pragmatischer Überlegungen zur Herleitung von Komponenten der Ergebnisqualität. Dies geschieht aus Mangel an vergleichbaren Beiträgen zum

⁴⁶³ Zudem ist der Vollständigkeit halber anzumerken, dass die Zusammensetzung der Tester nicht proportional zur Bevölkerung Deutschlands erfolgte.

Untersuchungsgegenstand. Insofern könnte diese Restriktion eine Anregung für zukünftige Forschungsbemühungen sein, das Auswertungsschema zu erweitern, indem weitere Komponenten identifiziert werden, die in eine Betrachtung einbezogen werden können.

5.3 Weitere Ansatzpunkte für zukünftige Forschungsbemühungen zur Mystery Shopping-Thematik

Im vorangegangenen Abschnitt wurde dargestellt, dass die vorliegende Untersuchung bestimmten Limitationen unterliegt, die gleichzeitig Anregungen für zukünftige Forschungsbemühungen markieren. Daneben zeigen die Übersicht der Theorie, die Betrachtung praktischer Umsetzungsaspekte und die Näherung an die theoretischen Fragestellungen in Kapitel 2 Schwachstellen auf und verdeutlichen, welche Aspekte im Rahmen zukünftiger Forschungsbemühung untersucht werden sollten, um einerseits einen höheren theoretischen Erkenntnisstand zu erlangen und um andererseits Unternehmen Hilfestellungen bei Umsetzungsfragen zu geben. Nachfolgend werden diese Aspekte benannt und es wird ein Vorgehen vorgeschlagen:

Im Vergleich mit der Kundenbefragung hat sich Mystery Shopping in der Ergebnisqualität als verlässlicher und beim Vergleich der Kosten als vorteilhafter erwiesen. Eine Kombination der Instrumente könnte weitere Vorteile bieten als der isolierte Einsatz. Eine Bestätigung oder Widerlegung der Vorteilhaftigkeit des kombinierten Einsatzes von Mystery Shopping und Kundenbefragung stellt eine Forschungslücke dar, deren Schließung der Praxis helfen würde, einen effektiven und effizienten Einsatz der Instrumente gemeinsam oder allein zu ermöglichen.

Hinsichtlich des Einsatzes in der Praxis ist eine fehlende Systematik des Einsatzes des Instrumentes festzustellen. Ziel eines Forschungsvorhabens könnte es sein, eine Systematik zur effektiven und effizienten Anwendung des Instrumentes zu entwickeln. Dies könnte Unternehmen eine wissenschaftlich fundierte Hilfestellung bei der Lösung von Praxisproblemen im Zusammenhang mit dem Einsatz des Instrumentes liefern.

Der Grad der Schulung der Tester ist umstritten. Ziel einer Untersuchung könnte es sein, Gewissheit darüber zu erlangen, ob und wie Testpersonen zu schulen sind. Unternehmen und Anbietern von Mystery Shopping könnte dies bei der Konzeption von Untersuchungen basierend auf wissenschaftlichen Erkenntnissen dienlich sein. Dafür könnten beispielsweise drei unterschiedliche Gruppen eingesetzt werden: Tester ohne Schulung, Tester mit schriftlicher Schulung und Tester mit persönlicher Schulung. Die Überprüfung dieser verschiedenen Schulungsformen könnte dabei auf Dienstleistungssituationen mit unterschiedlichen Komplexitätsgraden Anwendung finden. Das Ergebnis wäre ein Leitfaden für den anzuwendenden Schulungsgrad in Bezug auf den Komplexitätsgrad einer Dienstleistungssituation.

Bisher wurden lediglich qualitative Aussagen zum Nutzen von Mystery Shopping getroffen. Eine quantitative Aussage basierend auf einer Kosten-Nutzen-Betrachtung ist bisher ausgeblieben. Es erscheint sinnvoll, einen Ansatz zu entwickeln, der sich nicht nur mit Kosten-Nutzen-Aspekten der aus einem Mystery Shopping abgeleiteten Maßnahmen auseinandersetzt, sondern der vorher ansetzt. Ein solcher Return on Mystery Shopping kann Unternehmen eine Hilfestellung bei der Entscheidung hinsichtlich Durchführung und Umfang der Durchführung liefern.

Letztlich stellt sich für zukünftige Forschungen die Frage nach der internen oder externen Leistungserstellung von Mystery Shopping. In der vorliegenden Arbeit wurde sich dieser Frage genähert. Dabei wurde deutlich, dass die Verwendung der Transaktionskostentheorie als alleinige Grundlage einer Make-or-Buy-Entscheidung nicht ausreichend ist, da zu wenige Aspekte berücksichtigt werden. Eine fundiertere Untersuchung dieser Fragestellung unter Einbeziehung weiterer Theorien und idealerweise in Bezug auf die Art der Durchführung erscheint sinnvoll. Dies könnte Unternehmen eine praktikable Entscheidungsgrundlage liefern.

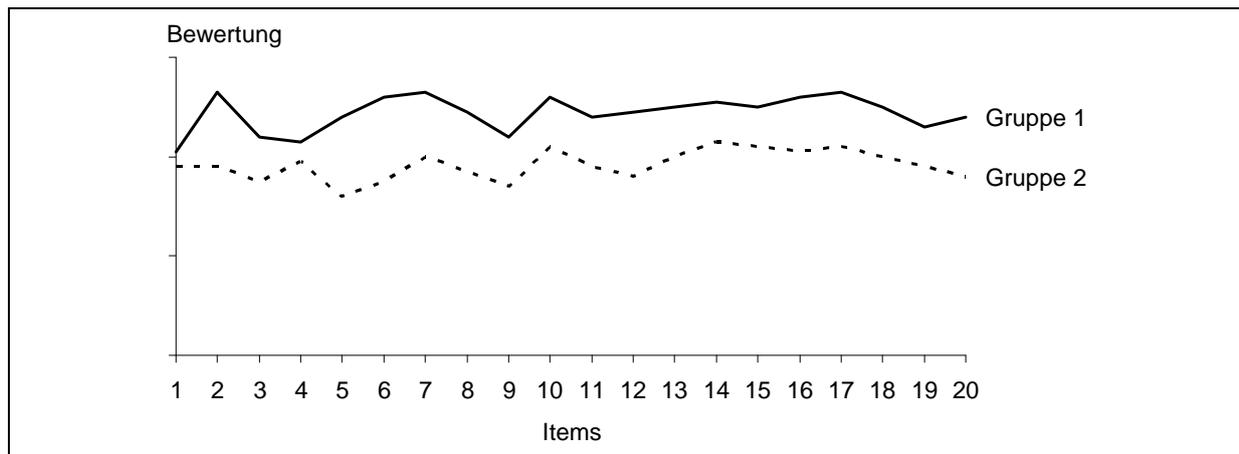
Anhang 1 – Ermittlung der Mittelwerts- und Streuungstendenz

Die *Mittelwertstendenz* kann erfasst werden, indem die Mittelwerte für alle Items innerhalb einer Gruppe (z.B. Männer) berechnet und in einem Mittelwertprofil zusammengefasst werden. Das arithmetische Mittel von n Werten x_i ist die Summe dieser Werte geteilt durch ihre Anzahl und wird wie folgt berechnet:

$$(M-1) \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Das erhaltene Mittelwertprofil kann mit einem erwarteten Profil (z.B. Normalverteilung) verglichen werden. Zudem können Mittelwertprofile von verschiedenen Gruppen (z.B. Männer vs. Frauen) gegenübergestellt werden, so dass eine Aussage über den relativen Unterschied der Gruppen getroffen werden kann. Abbildung A-1 soll den Vergleich der Mittelwertprofile zweier Gruppen veranschaulichen. Im direkten Vergleich bewertet Gruppe 1 milder und Gruppe 2 strenger. Dieser Unterschied ist rein deskriptiv und nicht als kausal zu interpretieren. Wie oben beschrieben, kann die unterschiedliche Ausprägung ein Indiz für eine notwendige „Eichung“ des Testerstabs, beispielsweise durch Schulungsmaßnahmen, sein. Der Vergleich der Mittelwertprofile lässt jedoch keine Aussage über das Verhältnis zu einem wahren Wert zu. Allerdings ließe sich ein wahrer Wert beispielsweise bestimmen, indem die Ergebnisse gegen die durch einen besonders geeigneten Tester erhobenen Werte gespiegelt würden.⁴⁶⁴ Gleichwohl können auch solche Expertenurteile falsch sein. Zudem wäre es wenig praktikabel, eine vollständige Erhebung der Dienstleistungsqualität eines Filialisten mit einer Vielzahl an Mystery Shoppern durch einen so genannten „Normcontroller“ zu überprüfen. Die Gegenüberstellung mit einem wahren Wert wäre einfacher zu realisieren, indem überprüfbare Werte abgefragt würden (etwa die Farbe des Logos). Allerdings ist die Erhebung solcher Attribute kaum mit der Erhebung der Dienstleistungsqualität zu vergleichen und daher nur wenig sinnvoll, wenn Schlüsse auf die Ergebnisqualität der Mystery Shopper gezogen werden sollen.

⁴⁶⁴ Vgl. hierzu die Ausführung in Kapitel 2.3.4.1 dieser Arbeit

Abbildung A-1: Exemplarischer grafischer Vergleich von Mittelwertprofilen

Quelle: Eigene Darstellung

Die Mittelwertsunterschiede zwischen den Gruppen müssen allerdings noch auf den Zufallseinfluss überprüft werden, damit festgestellt wird, ob die Unterschiede auf zufälligen Stichprobenschwankungen basieren oder systematischer Natur sind. Diese Überprüfung kann bei zwei Gruppen bei Normalverteilung der Werte mittels t-Test erfolgen.⁴⁶⁵ Bei mehr als zwei zu vergleichenden Gruppen kann die Überprüfung mittels einfaktorieller Varianzanalyse erfolgen, sofern die Werte der Stichproben normalverteilt sind und eine Varianzhomogenität über die Stichproben hinweg vorliegt.⁴⁶⁶ Sofern die Werte die Voraussetzungen nicht erfüllen, kann für den Vergleich von zwei Stichproben der U-Test von MANN und WHITNEY⁴⁶⁷ oder für k Stichproben der H-Test von KRUSKAL und WALLIS⁴⁶⁸ durchgeführt werden. Im Zuge der Untersuchung werden in der vorliegenden Arbeit Daten aus zwei (z.B. Geschlecht) und mehr Gruppen (z.B. Bildung) untersucht. Dabei werden in einigen Fällen normalverteilte, in anderen Fällen nicht normalverteilte Werte vorliegen. Damit ein schwer interpretierbares Durcheinander verschiedener Tests vermieden wird und der Testaufwand gering gehalten werden kann, werden die Signifikanztests für k Stichproben einheitlich mit dem H-Test durchgeführt, da dieser an keine der oben beschriebenen Voraussetzungen gebunden ist. Sofern der H-Test bedeutsame Unterschiede zwischen den mittleren Rängen aufweist wird nachfolgend mit dem U-Test von MANN und WHITNEY durch multiple Vergleiche geprüft, welche der paarweisen Differenzen statistisch signifikant sind.⁴⁶⁹

Beim H-Test besagt die Nullhypothese H_0 , dass zwischen den zu untersuchenden Gruppen kein Unterschied besteht.⁴⁷⁰ Als Prüfgröße wird ein H-Wert berechnet, der mit einem

⁴⁶⁵ Vgl. z.B. Homburg/Herrmann/Pflesser (2000: 112)

⁴⁶⁶ Vgl. Zöfel (2003: 129f.)

⁴⁶⁷ Vgl. Hartung/Elpelt/Klösener (2005: 519f)

⁴⁶⁸ Vgl. ebenda (2005: 613f.)

⁴⁶⁹ Vgl. Diehl/Staufenbiel (2007: 287)

⁴⁷⁰ Vgl. hierzu und nachfolgend Zöfel (2003: 143ff.)

theoretischen Wert aus der Chi-Quadrat-Tabelle für eine ausgewählte Irrtumswahrscheinlichkeit verglichen wird. Sofern der H -Wert größer als der Wert aus der Tabelle ist, kann die Nullhypothese verworfen und ein signifikanter Unterschied zwischen den Werten der Gruppen unterstellt werden. Bezeichnet man die Anzahl der Stichproben mit k , die Stichprobenumfänge mit n_i ($i=1, \dots, k$), und die Rangsummen in den Stichproben mit T_i ($i=1, \dots, k$), wird der H -Wert wie folgt berechnet:

$$(M-2) \quad H = \frac{12}{n \cdot (n+1)} \cdot \sum_{i=1}^k \frac{T_i^2}{n_i} - 3 \cdot (n+1)$$

Der H -Wert folgt einer Chi-Quadrat-Verteilung mit $df = k - 1$ Freiheitsgraden, so dass der H -Wert bei der gegebenen Anzahl dieser Freiheitsgrade die kritischen Werte der Chi-Quadrat-Tabelle übersteigen muss, damit eine Signifikanz auf der Stufe vorliegt.

Beim U-Test werden die gegebenen Werte der Variablen durch Rangplätze ersetzt.⁴⁷¹ Die Werte der beiden Stichproben werden dabei mit einer gemeinsamen Rangreihe versehen, wobei der kleinste Wert den Rangplatz 1 erhält, der nächst größere den Rangplatz 2. Tritt ein Wert zweimal auf, wird ein Rangplatz gemittelt, z.B. zu 6,5. Die Summen der Rangplätze R_1 und R_2 der beiden zu vergleichenden Stichproben n_1 und n_2 werden nun jeweils zu einem U-Wert (U_1 und U_2) umgerechnet. Nachfolgend ist die Berechnung des Wertes U_1 dargestellt. Die Berechnung für U_2 erfolgt analog durch Einsetzen von n_2 statt n_1 .

$$(M-3) \quad U_1 = R_1 - \frac{n_1 \cdot (n_1 + 1)}{2}$$

Die Prüfgröße des U-Tests ist nun der kleinere der beiden Werte U_1 und U_2 . Diesen Wert vergleicht man mit den zu n_1 und n_2 gehörenden kritischen Werten der U-Tabelle. Der berechnete U-Wert muss kleiner oder gleich dem kritischen U-Wert sein, sofern Signifikanz auf der betreffenden Stufe vorliegen soll. In der U-Tabelle sind die kritischen U-Werte bis $n_2=20$ aufgeführt. Wird $n_2 > 20$, nutzt man die Tatsache aus, dass sich die Verteilung von U sehr schnell einer Normalverteilung nähert. Man rechnet dann die errechnete Prüfgröße U in einen z -Wert um und vergleicht diesen mit dem zugehörigen p -Wert in der z -Tabelle. Der z -Wert wird wie folgt berechnet:⁴⁷²

$$(M-4) \quad z = \frac{\frac{n_1 \cdot n_2}{2} - U}{\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}}$$

⁴⁷¹ Vgl. hierzu und nachfolgend Zöfel (2003: 136ff.)

⁴⁷² Sofern gehäuft geteilte Rangplätze auftreten, ist die Berechnung zu modifizieren. Vgl. für die modifizierte Formel Zöfel (2003: 138)

Neben der Mittelwertstendenz soll an dieser Stelle die Erfassung der *Streuungstendenz* gezeigt werden. Während mit dem Mittelwert die Lage der Verteilung beschrieben wird, lässt sich mit Streuungsmaßen ihre Breite beschreiben. Das gebräuchlichste Maß ist die Standardabweichung. Sie wird berechnet, indem die Summe der quadratischen Abweichungen aller Messwerte x_i vom Mittelwert gebildet wird, diese durch die um eins verminderte Fallzahl n teilt und daraus die Wurzel zieht.⁴⁷³

$$(S-1) \quad s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

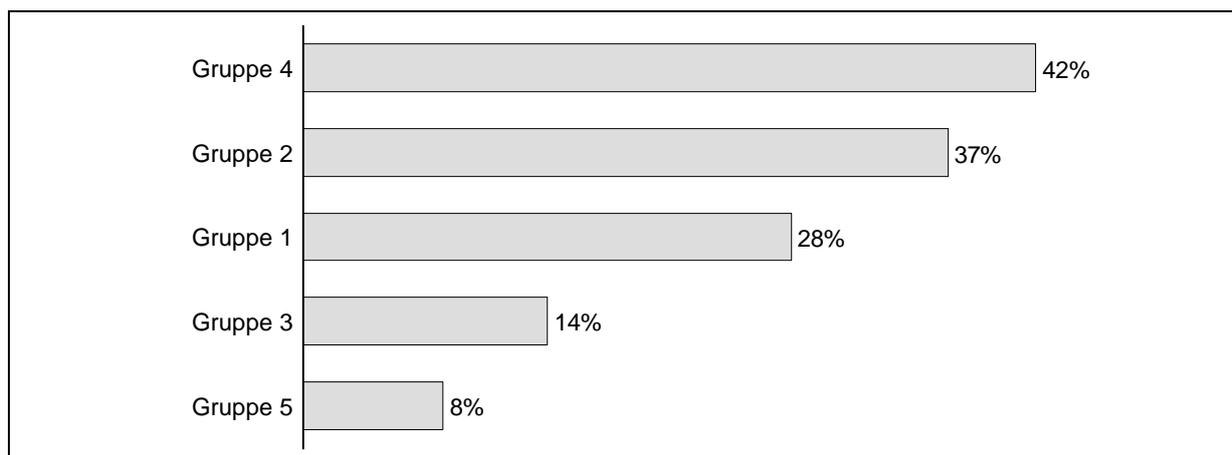
Analog zum Vergleich der Mittelwertprofile unterschiedlicher Gruppen von Mystery Shoppern können ihre Streuungsprofile verglichen werden. Sofern unterschiedliche Mittelwerte vorliegen, ist der Vergleich auf Basis eines relativen Streuungsmaßes sinnvoll. Der Variationskoeffizient V ist die Relativierung der Standardabweichung am Mittelwert und errechnet sich wie folgt:⁴⁷⁴

$$(S-2) \quad V = \frac{s}{\bar{x}}$$

Zudem können die Werte durch Durchschnittsbildung zu einem Einzelwert aggregiert werden. Während die Einzelwerte die Verteilungsbreite einzelner Items angeben, kann mit einem Durchschnittswert das Verhalten einer Gruppe bezogen auf den Gesamtbeobachtungsbogen abgebildet werden. Ein solcher einzelner Indikator erscheint überschaubarer als eine Vielzahl itemspezifischer Indikatoren. Während zwei Streuungsprofile über alle Items hinweg vergleichbar mit den in Abbildung A-1 dargestellten Mittelwertprofilen veranschaulicht werden können, soll nachstehende Abbildung A-2 den Vergleich fünf aggregierter Werte veranschaulichen. Die unterschiedlichen Gruppen haben jeweils unterschiedlich hohe Variationskoeffizienten. Während Gruppe 5 mit 8% einen vergleichsweise niedrigen Variationskoeffizienten aufweist, ist der Koeffizient bei Gruppe 4 mit 42% vergleichsweise hoch. Die durchschnittliche Streuungstendenz einer Gruppe über alle Items ist umso niedriger, je niedriger der Variationskoeffizient ist.

⁴⁷³ Vgl. z.B. Hartung/Elpelt/Klösener (2005:46f.)

⁴⁷⁴ Vgl. z.B. Reichardt/Reichardt (2000: 68)

Abbildung A-2: Exemplarischer grafischer Vergleich von aggregierten Variationskoeffizienten

Quelle: Eigene Darstellung

Als übergeordneter Indikator erscheint die Darstellung eines Einzelwertes für eine Gruppe sinnvoller und für den Zweck der Untersuchung praktikabler, da auf das globale Urteilsverhalten einer Gruppe bezüglich Urteilsstreuung geschlossen werden kann. Da einheitliche Maßeinheiten für die Mittelwerte berechnet werden, kann die durchschnittliche Standardabweichung als Maß zur Ermittlung der Streuungstendenz eingesetzt werden.

Allerdings wäre die alleinige Angabe eines aggregierten Wertes zur Beschreibung der Streuungstendenz nicht ganz eindeutig. Schließlich könnte der Unterschied insignifikant sein. Daher ist eine Signifikanzprüfung der Unterschiedlichkeit bei den einzelnen Items erforderlich. Die Überprüfung kann mittels einfaktorieller Varianzanalyse erfolgen, sofern die Werte der Stichproben normalverteilt sind und eine Varianzhomogenität über die Stichproben hinweg vorliegt.⁴⁷⁵ Sofern die Werte diese Voraussetzungen nicht erfüllen, können beispielsweise die von BROWN und FORSYTHE oder von WELCH entwickelten robusteren Verfahren eingesetzt werden.⁴⁷⁶ Gleichwohl erscheint eine erneute Überprüfung der Signifikanz des Unterschieds nicht erforderlich, da sie bei Ermittlung der Mittelwertstendenz bereits mittels H-Test und für paarweise Unterschiede mittels U-Test erfolgte.

Damit der Vorteil der Übersichtlichkeit eines durchschnittlichen Wertes zur Gegenüberstellung einzelner Gruppen als Indikator genutzt werden kann, wurde er in dieser Arbeit gemeinsam mit den signifikanten Items berichtet.

⁴⁷⁵ Vgl. Zöfel (2003: 130)

⁴⁷⁶ Vgl. Diehl/Staufenbiel (2007: 253)

Anhang 2 – Mittelwerte, Standardabweichungen der Gruppen

Ermittelte Daten für die SERVPERF-Operationalisierung

Tabelle A-1-1: Mittelwert, Standardabweichung, Varianz (Geschlecht)

	Geschlecht											
	Männlich				Weiblich				Insgesamt			
	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	2,21	193	,989	,978	2,23	183	1,029	1,060	2,22	376	1,008	1,015
Die Einrichtung ist ansprechend gestaltet.	2,44	193	1,069	1,144	2,43	183	1,050	1,103	2,43	376	1,059	1,121
Die Angestellten machen einen ordentlichen Eindruck.	2,11	192	,650	,423	2,15	183	,776	,603	2,13	375	,714	,510
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,45	193	,906	,821	2,43	181	,961	,924	2,44	374	,932	,869
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	2,17	189	,739	,546	2,07	181	,720	,518	2,12	370	,730	,533
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,33	187	,787	,619	2,33	178	,828	,686	2,33	365	,806	,650
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	1,87	191	,757	,574	1,97	183	,892	,796	1,92	374	,826	,683
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	1,91	187	,774	,599	1,95	182	,760	,578	1,93	369	,766	,587
McDonald's liefert fehlerfreie	1,85	186	,912	,831	1,87	180	,808	,653	1,86	366	,861	,741

Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,49	189	1,040	1,081	2,48	178	1,048	1,099	2,49	367	1,042	1,087
Der Gast erhält von den Angestellten einen prompten Service.	2,06	190	,746	,557	2,04	182	,894	,800	2,05	372	,821	,674
Die Mitarbeiter sind immer willens zu helfen.	2,26	191	,762	,581	2,27	181	,842	,710	2,26	372	,801	,642
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,69	190	,910	,829	2,56	179	,943	,888	2,63	369	,927	,860
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,68	191	,900	,810	2,64	182	,879	,772	2,66	373	,889	,790
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	2,19	190	,896	,803	2,20	182	,846	,715	2,20	372	,871	,758
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	2,20	191	,874	,763	2,25	182	,866	,751	2,23	373	,869	,756
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	2,34	188	,841	,707	2,36	180	,864	,746	2,35	368	,851	,724
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	2,99	192	1,008	1,016	2,90	181	1,131	1,279	2,95	373	1,069	1,142
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,55	192	,677	,458	1,55	182	,694	,481	1,55	374	,684	,468
McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	2,70	191	,894	,800	2,75	181	1,017	1,035	2,72	372	,955	,912
McDonald's liegen Ihre Interessen am Herzen.	2,97	190	,940	,883	2,93	178	1,028	1,058	2,95	368	,982	,965
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	2,95	190	,953	,907	2,97	180	1,022	1,044	2,96	370	,986	,971

Quelle: Eigene Darstellung

Tabelle A-1-2: Mittelwert, Standardabweichung, Varianz (Altersklassen)

	Altersklasse											
	18-30 Jahre				31-45 Jahre				älter als 45 Jahre			
	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	2,18	102	,895	,800	2,25	206	1,047	1,097	2,19	68	1,055	1,112
Die Einrichtung ist ansprechend gestaltet.	2,32	102	1,036	1,073	2,53	206	1,107	1,226	2,31	68	,918	,844
Die Angestellten machen einen ordentlichen Eindruck.	2,03	102	,710	,504	2,24	205	,727	,528	1,94	68	,620	,385
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,36	102	,973	,946	2,53	204	,949	,900	2,26	68	,785	,615
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	2,18	102	,849	,721	2,11	202	,707	,499	2,05	66	,593	,352
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,32	102	,846	,716	2,42	198	,795	,631	2,05	65	,717	,513
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	2,02	102	,923	,851	1,92	206	,837	,700	1,79	66	,595	,354
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	2,05	102	,927	,859	1,91	201	,722	,522	1,80	66	,588	,345
McDonald's liefert fehlerfreie Belege.	1,96	100	,898	,806	1,83	199	,886	,785	1,78	67	,714	,510
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,52	101	1,064	1,132	2,61	203	1,039	1,080	2,02	63	,889	,790
Der Gast erhält von den Angestellten einen prompten Service.	2,13	101	,934	,873	2,10	204	,790	,625	1,78	67	,670	,449
Die Mitarbeiter sind immer willens zu helfen.	2,29	101	,792	,627	2,39	205	,807	,651	1,83	66	,646	,418

Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,75	100	,892	,795	2,68	205	,904	,817	2,27	64	,980	,960
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,68	101	,882	,779	2,79	204	,876	,768	2,24	68	,813	,660
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	2,16	101	,857	,735	2,28	203	,920	,847	2,01	68	,702	,492
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	2,23	101	,835	,698	2,33	204	,929	,864	1,90	68	,626	,392
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	2,34	101	,852	,726	2,49	200	,856	,733	1,97	67	,717	,514
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	2,93	101	1,032	1,065	3,10	205	1,096	1,200	2,49	67	,911	,829
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,44	101	,590	,348	1,60	205	,746	,556	1,59	68	,604	,365
McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	2,72	101	,960	,922	2,87	204	,966	,933	2,28	67	,775	,600
McDonald's liegen Ihre Interessen am Herzen.	2,98	101	,938	,880	3,12	201	,988	,976	2,41	66	,841	,707
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	3,07	101	1,022	1,045	3,08	202	,948	,899	2,43	67	,874	,764

Quelle: Eigene Darstellung

Tabelle A-1-3: Mittelwert, Standardabweichung, Varianz (höchster Bildungsabschluss)

	Höchster Bildungsabschluss											
	Hauptschule/Realschule				Abitur				Studium			
	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	2,11	76	,946	,895	2,08	103	,977	,955	2,34	197	1,036	1,073
Die Einrichtung ist ansprechend gestaltet.	2,32	76	,912	,832	2,29	103	1,054	1,110	2,55	197	1,104	1,218

Die Angestellten machen einen ordentlichen Eindruck.	2,04	76	,621	,385	2,09	103	,781	,610	2,19	196	,709	,503
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,34	76	,825	,681	2,42	103	1,015	1,030	2,49	195	,927	,859
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	2,03	73	,645	,416	2,17	101	,825	,681	2,13	196	,708	,502
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,22	72	,773	,598	2,21	101	,840	,706	2,43	192	,790	,624
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	1,78	74	,603	,364	2,10	103	,924	,853	1,88	197	,834	,695
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	1,81	74	,612	,375	1,99	102	,862	,742	1,94	193	,765	,585
McDonald's liefert fehlerfreie Belege.	1,80	74	,776	,602	1,87	100	,761	,579	1,88	192	,941	,885
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,35	71	1,043	1,089	2,57	101	1,023	1,047	2,49	195	1,052	1,107
Der Gast erhält von den Angestellten einen prompten Service.	2,04	75	,845	,715	2,08	102	,886	,786	2,04	195	,779	,607
Die Mitarbeiter sind immer willens zu helfen.	2,12	74	,810	,656	2,24	101	,750	,563	2,33	197	,819	,671
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,45	73	,929	,862	2,71	100	,891	,794	2,65	196	,941	,886
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,49	76	,902	,813	2,74	101	,879	,773	2,68	196	,884	,781
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	2,09	76	,734	,538	2,19	101	,857	,734	2,25	195	,926	,857
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	2,13	76	,838	,702	2,15	101	,817	,668	2,30	196	,904	,817
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	2,16	76	,749	,561	2,31	98	,854	,730	2,45	194	,876	,767
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	2,72	75	,966	,934	2,91	102	1,054	1,111	3,05	196	1,104	1,218

McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,53	76	,577	,333	1,42	101	,682	,465	1,63	197	,714	,510
McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	2,56	75	,904	,817	2,65	102	,816	,666	2,83	195	1,031	1,062
McDonald's liegen Ihre Interessen am Herzen.	2,69	74	,859	,738	2,94	101	,998	,996	3,06	193	1,003	1,007
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	2,70	74	,872	,760	2,87	101	,956	,913	3,10	195	1,020	1,041

Quelle: Eigene Darstellung

Tabelle A-1-4: Mittelwert, Standardabweichung, Varianz (Tätigkeit)

	Tätigkeit															
	Schüler/Student				Hausfrau/Rentner				Voll berufstätig				Teilzeit berufstätig			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	1,90	51	0,878	0,770	2,06	31	1,209	1,462	2,31	255	1,024	1,048	2,21	39	0,801	0,641
Die Einrichtung ist ansprechend gestaltet.	1,98	51	1,068	1,140	2,32	31	1,194	1,426	2,51	255	1,049	1,101	2,59	39	0,850	0,722
Die Angestellten machen einen ordentlichen Eindruck.	1,82	51	0,713	0,508	1,81	31	0,543	0,295	2,25	254	0,710	0,503	2,03	39	0,668	0,447
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,24	51	0,929	0,864	1,97	31	0,657	0,432	2,52	253	0,945	0,893	2,54	39	0,913	0,834
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	2,14	51	0,775	0,601	1,79	29	0,491	0,241	2,16	251	0,754	0,569	2,05	39	0,605	0,366
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,29	51	0,901	0,812	1,93	28	0,604	0,365	2,38	247	0,818	0,669	2,31	39	0,655	0,429
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	2,02	51	0,927	0,860	1,77	30	0,679	0,461	1,94	254	0,839	0,704	1,79	39	0,695	0,483
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	1,96	51	0,799	0,638	1,72	29	0,591	0,350	1,98	250	0,794	0,630	1,72	39	0,605	0,366
McDonald's liefert fehlerfreie Belege.	1,80	51	0,693	0,481	1,87	30	0,900	0,809	1,91	246	0,913	0,833	1,59	39	0,637	0,406
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,44	50	1,146	1,313	1,89	27	0,974	0,949	2,56	251	1,023	1,047	2,46	39	0,969	0,939
Der Gast erhält von den Angestellten einen prompten Service.	2,08	50	0,900	0,810	1,70	30	0,651	0,424	2,12	253	0,829	0,687	1,82	39	0,683	0,467
Die Mitarbeiter sind immer willens zu helfen.	2,22	50	0,815	0,665	1,86	29	0,789	0,623	2,34	254	0,802	0,644	2,13	39	0,695	0,483
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,70	50	0,909	0,827	2,18	28	0,945	0,893	2,69	252	0,909	0,827	2,41	39	0,966	0,933
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,51	49	0,845	0,713	2,13	31	0,885	0,783	2,78	254	0,858	0,736	2,51	39	0,970	0,941
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	2,02	49	0,829	0,687	1,81	31	0,601	0,361	2,28	253	0,893	0,798	2,21	39	0,864	0,746

Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	2,02	49	0,924	0,854	1,97	31	0,836	0,699	2,31	254	0,858	0,737	2,13	39	0,833	0,694
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	2,14	49	0,816	0,667	1,74	31	0,575	0,331	2,45	250	0,841	0,707	2,45	38	0,921	0,849
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	2,78	50	0,975	0,951	2,37	30	0,928	0,861	3,05	254	1,068	1,140	2,92	39	1,156	1,336
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,28	50	0,454	0,206	1,71	31	0,693	0,480	1,57	254	0,712	0,507	1,64	39	0,668	0,447
McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	2,50	50	0,974	0,949	2,23	30	0,774	0,599	2,84	253	0,931	0,866	2,64	39	1,063	1,131
McDonald's liegen Ihre Interessen am Herzen.	2,86	50	1,010	1,021	2,47	30	0,860	0,740	3,00	249	0,994	0,988	3,13	39	0,864	0,746
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	2,84	50	0,976	0,953	2,43	30	0,817	0,668	3,04	251	0,995	0,990	3,00	39	0,946	0,895

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Tabelle A-1-5: Mittelwert, Standardabweichung, Varianz (Erfahrung als Mystery Shopper)

	Mystery Shopping-Erfahrung							
	keine Erfahrung vorhanden				Erfahrung vorhanden			
	Mittelwert	N	Standardabweichung	Varianz	Mittelwert	N	Standardabweichung	Varianz
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	2,31	277	1,041	1,083	1,98	99	,869	,755
Die Einrichtung ist ansprechend gestaltet.	2,51	277	1,079	1,164	2,21	99	,972	,944
Die Angestellten machen einen ordentlichen Eindruck.	2,17	277	,728	,530	2,03	98	,665	,442
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,50	277	,969	,939	2,27	97	,797	,636
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	2,16	271	,741	,549	2,01	99	,692	,479
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,37	267	,832	,693	2,21	98	,722	,521
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	1,97	275	,843	,711	1,80	99	,769	,591
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	2,01	270	,797	,636	1,71	99	,627	,393
McDonald's liefert fehlerfreie Belege.	1,91	270	,906	,820	1,72	96	,706	,499

Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,56	271	1,083	1,173	2,26	96	,885	,784
Der Gast erhält von den Angestellten einen prompten Service.	2,10	274	,839	,705	1,92	98	,755	,571
Die Mitarbeiter sind immer willens zu helfen.	2,26	274	,792	,627	2,27	98	,832	,692
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,64	272	,935	,873	2,59	97	,910	,828
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,69	275	,898	,807	2,58	98	,861	,741
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	2,20	275	,900	,810	2,20	97	,786	,617
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	2,23	275	,893	,798	2,21	98	,803	,644
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	2,36	270	,863	,744	2,33	98	,822	,676
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	2,99	275	1,081	1,168	2,82	98	1,029	1,059
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,59	276	,684	,468	1,44	98	,675	,455
McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	2,75	274	,948	,898	2,64	98	,977	,954
McDonald's liegen Ihre Interessen am Herzen.	2,97	271	,975	,951	2,90	97	1,005	1,010
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	2,97	272	,977	,955	2,92	98	1,012	1,024

Quelle: Eigene Darstellung

Tabelle A-1-6: Mittelwert, Standardabweichung, Varianz (Fast Food-Affinität)

	Losgelöst von Ihrer heutigen Aufgabe als Testbesucher, wie gern essen Sie in einem Fast Food Restaurant?																			
	Sehr gern				Eher gern				Teils-teils				Weniger gern				Gar nicht gern			
	MW	N	S	V	MW	N	W	V	MW	N	W	V	MW	N	W	V	MW	N	S	V
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	1,84	19	1,015	1,029	2,23	73	0,936	0,876	2,07	163	0,890	0,792	2,49	82	1,220	1,488	2,47	38	0,979	0,959
Die Einrichtung ist ansprechend gestaltet.	2,00	19	0,882	0,778	2,34	73	0,975	0,951	2,31	163	0,997	0,994	2,72	82	1,220	1,488	2,74	38	1,032	1,064

Die Angestellten machen einen ordentlichen Eindruck.	2,05	19	0,705	0,497	2,12	73	0,666	0,443	2,09	163	0,706	0,499	2,21	81	0,720	0,518	2,18	38	0,834	0,695
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,32	19	0,946	0,895	2,49	71	0,924	0,854	2,31	163	0,938	0,881	2,55	82	0,905	0,819	2,71	38	0,927	0,860
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	2,00	19	0,667	0,444	2,11	72	0,912	0,833	2,11	161	0,716	0,512	2,10	79	0,652	0,426	2,24	38	0,590	0,348
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,42	19	1,017	1,035	2,29	70	0,745	0,555	2,28	160	0,818	0,669	2,35	77	0,757	0,573	2,47	38	0,830	0,688
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	1,95	19	0,848	0,719	1,93	73	0,903	0,815	1,98	162	0,912	0,832	1,85	81	0,654	0,428	1,79	38	0,577	0,333
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	1,84	19	0,765	0,585	1,92	72	0,931	0,866	1,97	160	0,772	0,597	1,89	79	0,679	0,461	1,89	38	0,559	0,313
McDonald's liefert fehlerfreie Belege.	1,74	19	0,806	0,649	1,70	71	0,818	0,668	1,87	159	0,832	0,693	1,86	78	0,864	0,746	2,05	38	0,928	0,862
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,53	19	0,964	0,930	2,45	71	1,039	1,080	2,45	159	1,060	1,123	2,62	79	1,078	1,162	2,34	38	0,938	0,880
Der Gast erhält von den Angestellten einen prompten Service.	2,05	19	0,970	0,942	2,12	73	0,816	0,665	2,09	161	0,872	0,760	1,93	80	0,689	0,475	1,97	38	0,788	0,621
Die Mitarbeiter sind immer willens zu helfen.	2,11	19	0,737	0,544	2,29	72	0,721	0,519	2,32	161	0,876	0,768	2,15	81	0,673	0,453	2,29	38	0,898	0,806
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,58	19	0,902	0,813	2,69	71	0,872	0,760	2,64	160	0,942	0,887	2,50	80	0,941	0,886	2,71	38	0,956	0,914
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,58	19	0,838	0,702	2,71	72	0,813	0,660	2,62	161	0,928	0,862	2,74	82	0,858	0,736	2,58	38	0,976	0,953
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	2,05	19	0,705	0,497	2,04	72	0,759	0,576	2,20	161	0,934	0,873	2,37	81	0,858	0,736	2,21	38	0,875	0,765
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	1,84	19	0,765	0,585	2,33	72	0,839	0,704	2,20	161	0,916	0,839	2,21	82	0,733	0,537	2,32	38	1,016	1,033
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	2,00	19	0,667	0,444	2,32	71	0,713	0,508	2,41	160	0,934	0,873	2,32	81	0,788	0,621	2,31	36	0,822	0,675
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	2,74	19	0,991	0,982	3,03	73	0,942	0,888	2,85	160	1,117	1,248	3,11	82	1,030	1,062	2,89	38	1,158	1,340
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,05	19	0,229	0,053	1,57	72	0,709	0,502	1,52	162	0,680	0,462	1,66	82	0,741	0,549	1,68	38	0,574	0,330
McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	2,53	19	0,841	0,708	2,76	72	0,896	0,803	2,66	160	1,033	1,068	2,80	82	0,895	0,801	2,79	38	0,905	0,819
McDonald's liegen Ihre Interessen am Herzen.	2,84	19	1,119	1,251	2,99	72	0,864	0,746	2,84	158	1,031	1,064	3,08	80	0,938	0,880	3,13	38	0,991	0,982
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	2,79	19	1,084	1,175	3,06	72	0,918	0,842	2,84	158	1,006	1,013	3,11	82	0,916	0,840	2,97	38	1,052	1,107

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Tabelle A-1-7: Mittelwert, Standardabweichung, Varianz (Restaurant-Affinität)

	Und wie gern essen Sie bei McDonald's?																			
	Sehr gern				Eher gern				Teils-teils				Weniger gern				Gar nicht gern			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	2,21	24	1,021	1,042	2,11	99	0,891	0,794	2,15	143	0,922	0,849	2,38	74	1,202	1,444	2,42	31	1,205	1,452
Die Einrichtung ist ansprechend gestaltet.	2,58	24	0,974	0,949	2,26	99	0,996	0,992	2,40	143	1,008	1,016	2,55	74	1,218	1,483	2,71	31	1,131	1,280
Die Angestellten machen einen ordentlichen Eindruck.	2,13	24	0,741	0,549	2,08	99	0,665	0,442	2,11	143	0,662	0,438	2,22	73	0,786	0,618	2,16	31	0,898	0,806
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,38	24	0,875	0,766	2,44	98	0,953	0,909	2,36	142	0,910	0,828	2,50	74	0,954	0,911	2,68	31	1,013	1,026
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	2,08	24	0,881	0,775	2,07	97	0,794	0,630	2,13	142	0,693	0,480	2,17	71	0,717	0,514	2,13	31	0,619	0,383
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,38	24	0,770	0,592	2,22	96	0,823	0,678	2,33	139	0,784	0,614	2,41	70	0,825	0,681	2,39	31	0,844	0,712
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	2,00	24	1,022	1,043	1,89	99	0,891	0,794	1,95	142	0,811	0,657	1,99	73	0,790	0,625	1,71	31	0,588	0,346
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	1,83	24	0,963	0,928	1,97	98	0,831	0,690	1,95	140	0,771	0,595	1,89	72	0,640	0,410	1,90	30	0,662	0,438
McDonald's liefert fehlerfreie Belege.	1,58	24	0,654	0,428	1,83	96	0,867	0,751	1,85	138	0,801	0,641	1,99	72	0,957	0,915	1,84	31	0,860	0,740
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,57	23	1,080	1,166	2,48	96	1,046	1,094	2,45	141	1,031	1,063	2,62	71	1,033	1,068	2,32	31	1,045	1,092
Der Gast erhält von den Angestellten einen prompten Service.	2,35	23	1,152	1,328	2,03	99	0,762	0,581	2,08	142	0,834	0,696	1,97	74	0,702	0,492	1,97	29	0,906	0,820
Die Mitarbeiter sind immer willens zu helfen.	2,26	23	0,864	0,747	2,26	98	0,803	0,646	2,25	142	0,794	0,630	2,30	73	0,739	0,547	2,26	31	0,965	0,931
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,59	22	0,854	0,729	2,75	97	0,902	0,813	2,58	141	0,871	0,759	2,59	73	0,969	0,940	2,61	31	1,145	1,312
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,70	23	0,974	0,949	2,69	97	0,858	0,737	2,62	143	0,863	0,745	2,76	74	0,919	0,844	2,61	31	0,955	0,912
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	1,96	23	0,825	0,680	2,10	97	0,835	0,698	2,24	143	0,864	0,746	2,42	73	0,881	0,775	2,06	31	0,964	0,929
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	2,04	23	0,767	0,589	2,18	97	0,842	0,709	2,25	143	0,876	0,767	2,31	74	0,810	0,655	2,26	31	1,125	1,265
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	2,17	23	0,937	0,877	2,32	96	0,761	0,579	2,37	141	0,814	0,663	2,41	73	0,910	0,829	2,17	30	0,791	0,626
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	3,04	23	1,224	1,498	2,91	99	1,001	1,002	2,86	141	1,066	1,137	3,19	74	1,016	1,032	2,84	31	1,214	1,473
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,22	23	0,422	0,178	1,52	98	0,662	0,438	1,58	143	0,726	0,527	1,61	74	0,679	0,461	1,68	31	0,702	0,492
McDonald's hat Mitarbeiter, die Ihnen persönliche	2,83	23	1,230	1,514	2,70	98	0,840	0,705	2,68	141	0,966	0,933	2,86	74	0,941	0,886	2,65	31	1,018	1,037

Aufmerksamkeit widmen.																				
McDonald's liegen Ihre Interessen am Herzen.	2,87	23	1,217	1,482	2,88	96	0,932	0,868	2,91	139	0,982	0,964	3,08	74	0,962	0,925	3,03	31	0,983	0,966
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	3,09	23	1,164	1,356	2,90	96	0,946	0,894	2,90	141	0,913	0,833	3,05	74	1,071	1,148	2,97	31	1,048	1,099

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Tabelle A-1-8: Mittelwert, Standardabweichung, Varianz (Häufigkeit eines Fast Food-Restaurantbesuchs)

	FastFood Häufigkeitslevel															
	keine Besuche				1-12 mal im Jahr				13-52 mal im Jahr				mehr als 52 mal im Jahr			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	1,86	14	0,663	0,440	2,30	186	1,032	1,066	2,19	150	1,008	1,016	2,00	26	0,938	0,880
Die Einrichtung ist ansprechend gestaltet.	2,29	14	0,994	0,989	2,56	186	1,100	1,209	2,31	150	0,990	0,979	2,31	26	1,123	1,262
Die Angestellten machen einen ordentlichen Eindruck.	1,93	14	0,730	0,533	2,17	186	0,707	0,500	2,11	149	0,740	0,548	2,04	26	0,599	0,358
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,36	14	0,745	0,555	2,55	185	0,944	0,891	2,32	149	0,901	0,812	2,42	26	1,065	1,134
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	2,00	13	0,408	0,167	2,17	183	0,631	0,398	2,06	148	0,851	0,724	2,12	26	0,766	0,586
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,00	13	0,408	0,167	2,31	179	0,772	0,596	2,41	148	0,832	0,693	2,16	25	0,987	0,973
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	1,86	14	0,363	0,132	1,91	184	0,777	0,604	1,90	150	0,896	0,802	2,15	26	0,925	0,855
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	1,86	14	0,363	0,132	1,98	181	0,703	0,494	1,87	149	0,836	0,698	1,96	25	0,935	0,873
McDonald's liefert fehlerfreie Belege.	2,14	14	0,949	0,901	1,92	180	0,851	0,725	1,74	147	0,853	0,727	1,96	25	0,889	0,790
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,08	13	0,954	0,910	2,44	180	1,042	1,086	2,55	148	1,032	1,065	2,62	26	1,134	1,286
Der Gast erhält von den Angestellten einen prompten Service.	1,71	14	0,611	0,374	2,02	182	0,814	0,662	2,13	150	0,833	0,694	1,96	26	0,871	0,758
Die Mitarbeiter sind immer willens zu helfen.	2,07	14	0,730	0,533	2,21	183	0,801	0,641	2,38	149	0,818	0,669	2,08	26	0,688	0,474
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,21	14	0,802	0,643	2,59	182	0,958	0,918	2,71	147	0,868	0,753	2,62	26	1,061	1,126
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,36	14	0,842	0,709	2,68	185	0,911	0,829	2,70	150	0,880	0,775	2,46	24	0,779	0,607
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	2,00	14	0,784	0,615	2,29	185	0,867	0,751	2,13	149	0,864	0,747	2,04	24	0,955	0,911
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	1,86	14	0,663	0,440	2,24	185	0,847	0,718	2,26	150	0,923	0,851	2,08	24	0,776	0,601
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	2,00	14	0,555	0,308	2,33	183	0,879	0,773	2,44	147	0,845	0,713	2,17	24	0,761	0,580
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	2,57	14	0,852	0,725	2,91	183	1,086	1,179	3,01	150	1,068	1,141	3,00	26	1,058	1,120
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,71	14	0,611	0,374	1,65	185	0,692	0,479	1,46	149	0,663	0,439	1,31	26	0,679	0,462

McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	2,50	14	0,760	0,577	2,73	183	0,990	0,980	2,77	150	0,970	0,942	2,52	25	0,653	0,427
McDonald's liegen Ihre Interessen am Herzen.	2,85	13	0,801	0,641	2,90	181	0,967	0,934	3,01	149	0,986	0,973	3,04	25	1,172	1,373
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	2,62	13	1,044	1,090	2,94	182	0,953	0,908	3,01	150	0,983	0,966	2,96	25	1,207	1,457

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Tabelle F-1-9: Mittelwert, Standardabweichung, Varianz (spezifische Restaurant-Besuchshäufigkeit)

	Spezifische Restaurant-Besuchshäufigkeit															
	keine Besuche				1-12 mal im Jahr				13-52 mal im Jahr				mehr als 52 mal im Jahr			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Die Betriebsausrüstung des McDonald's-Restaurants sieht modern aus.	1,97	36	1,000	0,999	2,30	255	1,045	1,092	2,11	80	0,871	0,759	1,80	5	0,837	0,700
Die Einrichtung ist ansprechend gestaltet.	2,36	36	0,990	0,980	2,49	255	1,104	1,219	2,35	80	0,943	0,889	1,60	5	0,548	0,300
Die Angestellten machen einen ordentlichen Eindruck.	1,89	36	0,667	0,444	2,19	254	0,739	0,547	2,09	80	0,640	0,410	1,80	5	0,447	0,200
Materialien, die mit der Dienstleistung in Verbindung stehen (z.B. Displays, Broschüren, Speisekarten) sind visuell ansprechend.	2,25	36	0,732	0,536	2,49	254	0,940	0,883	2,37	79	0,989	0,979	2,40	5	0,894	0,800
Wenn McDonald's verspricht, etwas zu einem bestimmten Zeitpunkt zu tun, dann wird es auch getan.	1,97	34	0,521	0,272	2,15	251	0,711	0,505	2,11	80	0,857	0,734	1,60	5	0,548	0,300
Hat ein Gast ein Problem, dann zeigt McDonald's ein ernstes Interesse, es zu lösen.	2,09	34	0,668	0,447	2,35	247	0,802	0,643	2,41	79	0,855	0,731	1,80	5	0,837	0,700
McDonald's führt die Bestellung gleich beim ersten Mal richtig aus.	1,81	36	0,525	0,275	1,92	253	0,858	0,736	1,96	80	0,834	0,695	2,00	5	1,000	1,000
Die Bestellung wird zum versprochenen Zeitpunkt ausgeführt.	1,82	34	0,626	0,392	1,96	251	0,734	0,539	1,89	79	0,906	0,820	1,60	5	0,894	0,800
McDonald's liefert fehlerfreie Belege.	1,92	36	0,806	0,650	1,91	246	0,895	0,800	1,68	79	0,777	0,604	1,60	5	0,548	0,300
Die Mitarbeiter geben dem Gast genau darüber Auskunft, wann die Bestellung ausgeführt wird.	2,18	34	1,086	1,180	2,50	249	1,013	1,025	2,58	79	1,105	1,221	2,40	5	1,140	1,300
Der Gast erhält von den Angestellten einen prompten Service.	1,67	33	0,645	0,417	2,10	254	0,806	0,650	2,06	80	0,905	0,819	2,00	5	0,707	0,500
Die Mitarbeiter sind immer willens zu helfen.	1,91	34	0,712	0,507	2,32	253	0,795	0,633	2,25	80	0,819	0,671	1,80	5	0,837	0,700
Die Mitarbeiter sind nie zu beschäftigt, um auf Kundenwünsche einzugehen.	2,09	34	0,900	0,810	2,71	251	0,908	0,825	2,62	79	0,924	0,854	2,40	5	1,140	1,300
Die Mitarbeiter haben ein Vertrauen erweckendes Verhalten.	2,28	36	0,779	0,606	2,72	254	0,883	0,779	2,65	78	0,937	0,879	2,40	5	0,548	0,300
Als Gast fühlen Sie sich bei einem Essen bei McDonald's sicher.	1,94	36	0,791	0,625	2,28	253	0,866	0,750	2,09	78	0,900	0,810	1,60	5	0,548	0,300
Mitarbeiter von McDonald's sind durchweg höflich zu Ihnen.	1,86	36	0,762	0,580	2,28	254	0,855	0,732	2,24	78	0,942	0,888	1,80	5	0,447	0,200
Die Mitarbeiter verfügen über die notwendigen Kenntnisse, um Ihre Fragen beantworten zu können.	1,97	36	0,654	0,428	2,42	250	0,871	0,759	2,30	77	0,844	0,712	2,20	5	0,447	0,200
McDonald's gibt Ihnen individuelle Aufmerksamkeit.	2,71	35	1,202	1,445	2,95	253	1,030	1,061	3,06	80	1,118	1,249	2,40	5	1,140	1,300
McDonald's hat Öffnungszeiten, die für alle Kunden geeignet sind.	1,61	36	0,645	0,416	1,58	253	0,677	0,459	1,45	80	0,727	0,529	1,20	5	0,447	0,200
McDonald's hat Mitarbeiter, die Ihnen persönliche Aufmerksamkeit widmen.	2,46	35	1,039	1,079	2,78	253	0,950	0,903	2,67	79	0,944	0,890	2,60	5	0,548	0,300
McDonald's liegen Ihre Interessen am Herzen.	2,71	34	0,970	0,941	2,97	250	0,979	0,959	3,03	79	1,000	0,999	2,60	5	0,894	0,800
Die Mitarbeiter von McDonald's verstehen Ihre speziellen Bedürfnisse.	2,65	34	1,041	1,084	2,98	252	0,951	0,904	3,04	79	1,043	1,088	2,80	5	1,304	1,700

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Ermittelte Daten für die Operationalisierung der Kontaktpunkte

Tabelle A-2-1: Mittelwert, Standardabweichung, Varianz (Geschlecht)

	Geschlecht											
	Männlich				Weiblich				Insgesamt			
	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz
Wie ist der Sauberkeitszustand der näheren Umgebung des Restauranteingangs?	1,75	193	,694	,482	1,64	183	,826	,682	1,69	376	,762	,581
Wie ist der Zustand von Fenstern und Fassaden?	1,45	193	,636	,405	1,56	183	,738	,544	1,51	376	,689	,475
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	1,98	193	,760	,578	2,14	182	,868	,753	2,06	375	,817	,667
Wie ist der Sauberkeitszustand des Bodens?	1,84	192	,731	,534	1,95	182	,796	,633	1,89	374	,764	,584
Wie ist der Zustand des Sitzbereichs?	1,65	193	,735	,540	1,70	182	,713	,508	1,68	375	,724	,524
Wie ist der Zustand der Servicestation?	1,48	192	,655	,429	1,56	180	,662	,438	1,52	372	,658	,434
Wie ist der Zustand des Thekenbereichs?	1,32	192	,512	,262	1,28	180	,499	,249	1,30	372	,505	,255
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	1,99	193	,770	,594	2,12	182	,839	,704	2,05	375	,806	,650
Beim Herantreten an die Theke werde ich freundlich begrüßt.	2,15	192	,987	,973	2,18	182	1,138	1,295	2,16	374	1,062	1,127
Ich werde freundlich nach meinem Bestellwunsch gefragt.	2,01	192	,865	,749	2,06	181	1,042	1,086	2,03	373	,954	,910
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	2,02	192	,973	,947	2,02	182	1,107	1,226	2,02	374	1,039	1,080
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,29	192	1,134	1,286	2,25	179	1,197	1,434	2,27	371	1,164	1,354
Der Bezahlvorgang verläuft reibungslos.	1,49	192	,772	,597	1,54	181	,833	,694	1,51	373	,802	,643

Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	2,10	191	1,103	1,216	2,23	182	1,271	1,615	2,16	373	1,188	1,411
Die Mitarbeiter sind durchweg höflich zu mir.	1,87	191	,805	,647	2,01	181	,969	,939	1,94	372	,890	,791
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	2,31	192	,872	,760	2,32	182	,939	,881	2,32	374	,904	,817
Die Mitarbeiter wirken ordentlich.	2,01	193	,661	,437	2,01	183	,741	,549	2,01	376	,700	,491
Die Mitarbeiter wirken hektisch und überlastet.	3,74	192	,870	,757	3,66	183	1,031	1,062	3,70	375	,951	,905
Die Mitarbeiter machen einen freundlichen Eindruck.	2,26	192	,762	,581	2,32	182	,928	,861	2,29	374	,846	,716
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,25	189	,712	,507	2,30	183	,926	,858	2,27	372	,823	,678
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,32	192	,694	,482	2,37	183	,794	,630	2,35	375	,744	,553
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,34	192	,743	,551	2,48	180	,787	,620	2,41	372	,766	,587

Quelle: Eigene Darstellung

Tabelle A-2-2: Mittelwert, Standardabweichung, Varianz (Altersklassen)

	Altersklasse											
	18-30 Jahre				31-45 Jahre				älter als 45 Jahre			
	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz
Wie ist der Sauberkeitszustand der näheren Umgebung des Resturanteingangs?	1,80	102	,797	,634	1,68	206	,774	,599	1,57	68	,654	,427
Wie ist der Zustand von Fenstern und Fassaden?	1,52	102	,656	,430	1,57	206	,693	,481	1,29	68	,692	,479
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	2,12	102	,848	,719	2,05	205	,768	,590	2,00	68	,914	,836
Wie ist der Sauberkeitszustand des Bodens?	1,90	102	,777	,604	1,94	204	,737	,543	1,75	68	,817	,668

Wie ist der Zustand des Sitzbereichs?	1,67	102	,694	,482	1,72	205	,738	,545	1,56	68	,720	,519
Wie ist der Zustand der Servicestation?	1,61	100	,764	,584	1,54	204	,614	,377	1,31	68	,580	,336
Wie ist der Zustand des Thekenbereichs?	1,33	101	,550	,302	1,33	203	,513	,263	1,18	68	,384	,147
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	2,07	102	,824	,678	2,11	205	,791	,626	1,87	68	,809	,654
Beim Herantreten an die Theke werde ich freundlich begrüßt.	2,10	102	1,067	1,139	2,28	205	1,056	1,116	1,88	67	1,023	1,046
Ich werde freundlich nach meinem Bestellwunsch gefragt.	2,07	102	,998	,995	2,08	204	,925	,855	1,81	67	,957	,916
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	2,06	102	1,106	1,224	2,05	205	1,058	1,120	1,85	67	,857	,735
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,18	101	1,144	1,308	2,33	204	1,156	1,337	2,21	66	1,222	1,493
Der Bezahlvorgang verläuft reibungslos.	1,47	102	,780	,608	1,53	204	,821	,674	1,54	67	,785	,616
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	2,04	102	1,107	1,226	2,30	204	1,230	1,513	1,93	67	1,132	1,282
Die Mitarbeiter sind durchweg höflich zu mir.	1,85	101	,865	,748	2,06	204	,940	,883	1,70	67	,697	,485
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	2,20	102	,845	,714	2,43	205	,919	,844	2,15	67	,909	,826
Die Mitarbeiter wirken ordentlich.	2,00	102	,660	,436	2,08	206	,715	,511	1,82	68	,690	,476
Die Mitarbeiter wirken hektisch und überlastet.	3,54	102	1,040	1,083	3,71	206	,901	,812	3,93	67	,926	,858
Die Mitarbeiter machen einen freundlichen Eindruck.	2,24	102	,835	,697	2,42	204	,847	,717	2,00	68	,792	,627
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,27	102	,869	,756	2,37	203	,824	,680	1,97	67	,674	,454
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,26	102	,783	,612	2,47	205	,718	,515	2,09	68	,685	,470
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,31	100	,787	,620	2,47	204	,725	,526	2,37	68	,845	,714

Quelle: Eigene Darstellung

Tabelle A-2-3: Mittelwert, Standardabweichung, Varianz (höchster Bildungsabschluss)

	Höchster Bildungsabschluss											
	Hauptschule/Realschule				Abitur				Studium			
	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz	Mittelwert	N	Standard- abweichung	Varianz
Wie ist der Sauberkeitszustand der näheren Umgebung des Resturanteingangs?	1,63	76	,709	,502	1,77	103	,819	,671	1,68	197	,752	,566
Wie ist der Zustand von Fenstern und Fassaden?	1,37	76	,746	,556	1,49	103	,739	,546	1,57	197	,632	,400
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	2,04	75	,861	,742	2,01	103	,786	,618	2,10	197	,818	,669
Wie ist der Sauberkeitszustand des Bodens?	1,96	75	,892	,796	1,84	102	,714	,510	1,89	197	,738	,545
Wie ist der Zustand des Sitzbereichs?	1,63	75	,712	,507	1,70	103	,725	,526	1,69	197	,730	,533
Wie ist der Zustand der Servicestation?	1,35	75	,507	,257	1,58	101	,738	,545	1,55	196	,658	,433
Wie ist der Zustand des Thekenbereichs?	1,23	75	,421	,178	1,35	101	,573	,329	1,31	196	,496	,246
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	1,95	75	,820	,673	2,04	103	,803	,646	2,10	197	,802	,643
Beim Herantreten an die Theke werde ich freundlich begrüßt.	2,13	75	1,095	1,198	2,10	103	1,062	1,128	2,20	196	1,052	1,107
Ich werde freundlich nach meinem Bestellwunsch gefragt.	2,01	75	,966	,932	2,01	103	,934	,872	2,05	195	,965	,931
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	2,01	75	,937	,878	2,14	103	1,121	1,256	1,96	196	1,032	1,065
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,34	74	1,231	1,514	2,38	102	1,305	1,704	2,18	195	1,054	1,110

Der Bezahlvorgang verläuft reibungslos.	1,49	75	,760	,578	1,50	103	,726	,527	1,53	195	,857	,735
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	2,19	75	1,205	1,451	2,17	103	1,192	1,420	2,15	195	1,186	1,406
Die Mitarbeiter sind durchweg höflich zu mir.	2,04	75	,922	,850	1,87	103	,836	,700	1,93	194	,905	,819
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	2,40	75	,915	,838	2,29	103	,870	,757	2,30	196	,919	,845
Die Mitarbeiter wirken ordentlich.	1,93	76	,680	,462	1,98	103	,767	,588	2,06	197	,672	,451
Die Mitarbeiter wirken hektisch und überlastet.	3,71	76	,907	,822	3,68	103	1,040	1,083	3,71	196	,924	,853
Die Mitarbeiter machen einen freundlichen Eindruck.	2,20	75	,870	,757	2,24	103	,834	,695	2,35	196	,843	,711
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,17	75	,760	,578	2,31	101	,925	,855	2,29	196	,792	,628
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,29	76	,797	,635	2,28	102	,736	,542	2,40	197	,726	,527
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,47	76	,739	,546	2,29	101	,739	,547	2,45	195	,787	,620

Quelle: Eigene Darstellung

Tabelle A-2-4: Mittelwert, Standardabweichung, Varianz (Tätigkeitsgruppen)

	Tätigkeit															
	Schüler/Student				Hausfrau/Rentner				Voll berufstätig				Teilzeit berufstätig			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Wie ist der Sauberkeitszustand der näheren Umgebung des Restauranteingangs?	1,59	51	0,698	0,487	1,45	31	0,675	0,456	1,78	255	0,787	0,619	1,44	39	0,641	0,410
Wie ist der Zustand von Fenstern und Fassaden?	1,33	51	0,589	0,347	1,23	31	0,560	0,314	1,60	255	0,696	0,484	1,31	39	0,731	0,534
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	1,84	51	0,731	0,535	1,81	31	0,749	0,561	2,14	255	0,804	0,646	2,05	38	0,985	0,970
Wie ist der Sauberkeitszustand des Bodens?	1,78	51	0,673	0,453	1,65	31	0,839	0,703	1,94	254	0,757	0,573	1,95	38	0,837	0,700
Wie ist der Zustand des Sitzbereichs?	1,63	51	0,720	0,518	1,55	31	0,810	0,656	1,69	255	0,733	0,538	1,76	38	0,590	0,348
Wie ist der Zustand der Servicestation?	1,47	51	0,578	0,334	1,13	31	0,562	0,316	1,59	253	0,687	0,472	1,41	37	0,498	0,248

Wie ist der Zustand des Thekenbereichs?	1,30	50	0,544	0,296	1,06	31	0,359	0,129	1,34	253	0,516	0,266	1,24	38	0,431	0,186
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	1,88	51	0,840	0,706	1,81	31	0,910	0,828	2,11	255	0,774	0,599	2,13	38	0,844	0,712
Beim Herantreten an die Theke werde ich freundlich begrüßt.	2,06	51	1,156	1,336	1,81	31	0,910	0,828	2,25	254	1,069	1,144	2,00	38	0,930	0,865
Ich werde freundlich nach meinem Bestellwunsch gefragt.	2,02	51	1,104	1,220	1,94	31	0,929	0,862	2,07	253	0,944	0,892	1,84	38	0,823	0,677
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	2,14	51	1,249	1,561	1,74	31	0,729	0,531	2,02	254	1,033	1,067	2,05	38	0,985	0,970
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,18	51	1,276	1,628	2,35	31	1,404	1,970	2,26	252	1,123	1,262	2,38	37	1,089	1,186
Der Bezahlvorgang verläuft reibungslos.	1,35	51	0,559	0,313	1,39	31	0,667	0,445	1,57	253	0,877	0,770	1,47	38	0,603	0,364
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	1,92	51	1,093	1,194	2,19	31	1,352	1,828	2,21	253	1,194	1,426	2,18	38	1,136	1,289
Die Mitarbeiter sind durchweg höflich zu mir.	1,84	51	0,925	0,855	1,77	31	0,762	0,581	1,98	253	0,904	0,817	1,89	37	0,843	0,710
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	2,10	51	0,964	0,930	2,19	31	0,833	0,695	2,39	254	0,916	0,840	2,21	38	0,741	0,549
Die Mitarbeiter wirken ordentlich.	1,88	51	0,683	0,466	1,74	31	0,631	0,398	2,09	255	0,712	0,508	1,87	39	0,615	0,378
Die Mitarbeiter wirken hektisch und überlastet.	3,59	51	1,080	1,167	3,94	31	0,814	0,662	3,73	254	0,932	0,869	3,46	39	0,969	0,939
Die Mitarbeiter machen einen freundlichen Eindruck.	2,14	51	0,872	0,761	2,13	31	0,922	0,849	2,36	253	0,821	0,675	2,21	39	0,894	0,799
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,32	50	0,957	0,916	1,97	30	0,669	0,447	2,30	253	0,804	0,646	2,28	39	0,857	0,734
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,18	50	0,850	0,722	2,19	31	0,792	0,628	2,41	255	0,720	0,519	2,26	39	0,677	0,459
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,16	50	0,766	0,586	2,39	31	0,844	0,712	2,43	253	0,756	0,572	2,58	38	0,722	0,521

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Tabelle A-2-5: Mittelwert, Standardabweichung, Varianz (Erfahrung als Mystery Shopper)

	Mystery Shopping-Erfahrung							
	keine Erfahrung vorhanden				Erfahrung vorhanden			
	Mittelwert	N	Standardabweichung	Varianz	Mittelwert	N	Standardabweichung	Varianz
Wie ist der Sauberkeitszustand der näheren Umgebung des Resturanteingangs?	1,69	277	,789	,622	1,72	99	,686	,470
Wie ist der Zustand von Fenstern und Fassaden?	1,53	277	,699	,489	1,43	99	,657	,432
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	2,13	276	,809	,655	1,86	99	,808	,653
Wie ist der Sauberkeitszustand des Bodens?	1,92	276	,785	,616	1,81	98	,698	,488
Wie ist der Zustand des Sitzbereichs?	1,71	276	,720	,518	1,58	99	,730	,532
Wie ist der Zustand der Servicestation?	1,56	275	,688	,473	1,39	97	,551	,303
Wie ist der Zustand des Thekenbereichs?	1,30	275	,504	,254	1,32	97	,511	,261

Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	2,10	276	,819	,670	1,93	99	,759	,577
Beim Herantreten an die Theke werde ich freundlich begrüßt.	2,16	276	1,079	1,165	2,17	98	1,016	1,031
Ich werde freundlich nach meinem Bestellwunsch gefragt.	2,09	275	,996	,992	1,87	98	,808	,652
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	2,05	276	1,062	1,129	1,94	98	,972	,945
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,32	273	1,220	1,489	2,14	98	,984	,969
Der Bezahlvorgang verläuft reibungslos.	1,51	276	,816	,665	1,53	97	,765	,585
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	2,15	275	1,192	1,422	2,19	98	1,181	1,395
Die Mitarbeiter sind durchweg höflich zu mir.	1,95	275	,899	,808	1,91	97	,867	,752
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	2,33	276	,913	,834	2,27	98	,880	,774
Die Mitarbeiter wirken ordentlich.	2,03	277	,719	,517	1,97	99	,646	,417
Die Mitarbeiter wirken hektisch und überlastet.	3,70	277	,979	,958	3,71	98	,873	,763
Die Mitarbeiter machen einen freundlichen Eindruck.	2,34	276	,865	,748	2,16	98	,782	,612
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,30	273	,820	,673	2,20	99	,833	,693
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,39	276	,753	,566	2,22	99	,708	,501
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,49	274	,776	,602	2,18	98	,694	,481

Quelle: Eigene Darstellung

Tabelle A-2-6: Mittelwert, Standardabweichung, Varianz (Fast Food-Affinität)

	Losgelöst von Ihrer heutigen Aufgabe als Testbesucher, wie gern essen Sie in einem Fast Food Restaurant?																			
	Sehr gern				Eher gern				Teils-teils				Weniger gern				Gar nicht gern			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Wie ist der Sauberkeitszustand der näheren Umgebung des Resturanteingangs?	1,53	19	0,697	0,485	1,66	73	0,692	0,478	1,74	163	0,734	0,538	1,68	82	0,901	0,812	1,66	38	0,745	0,555
Wie ist der Zustand von Fenstern und Fassaden?	1,47	19	0,772	0,596	1,44	73	0,552	0,305	1,48	163	0,706	0,498	1,57	82	0,770	0,593	1,63	38	0,633	0,401

Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	1,79	19	0,631	0,398	1,86	73	0,751	0,564	2,10	162	0,839	0,703	2,07	82	0,766	0,587	2,37	38	0,942	0,888
Wie ist der Sauberkeitszustand des Bodens?	1,68	19	0,582	0,339	1,82	73	0,805	0,648	1,93	161	0,815	0,664	1,89	82	0,737	0,543	1,97	38	0,592	0,351
Wie ist der Zustand des Sitzbereichs?	1,58	19	0,507	0,257	1,49	73	0,556	0,309	1,69	162	0,798	0,637	1,79	82	0,749	0,561	1,79	38	0,664	0,441
Wie ist der Zustand der Servicestation?	1,37	19	0,496	0,246	1,52	71	0,694	0,482	1,54	162	0,697	0,486	1,60	82	0,645	0,416	1,30	37	0,463	0,215
Wie ist der Zustand des Thekenbereichs?	1,21	19	0,419	0,175	1,21	71	0,445	0,198	1,35	162	0,551	0,304	1,35	81	0,504	0,254	1,24	38	0,431	0,186
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	1,68	19	0,820	0,673	1,93	73	0,751	0,565	2,03	162	0,759	0,577	2,16	82	0,867	0,752	2,34	38	0,878	0,772
Beim Herantreten an die Theke werde ich freundlich begrüßt.	1,95	19	0,848	0,719	2,33	73	1,094	1,196	2,11	161	1,043	1,087	2,23	82	1,081	1,168	1,92	38	1,024	1,048
Ich werde freundlich nach meinem Bestellwunsch gefragt.	2,05	19	0,970	0,942	2,08	73	0,909	0,826	1,99	161	0,929	0,862	2,09	81	1,063	1,130	1,92	38	0,882	0,777
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	2,00	19	1,054	1,111	2,12	73	1,079	1,165	2,08	161	1,095	1,200	1,95	82	0,942	0,886	1,68	38	0,873	0,762
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,05	19	1,079	1,164	2,38	72	1,168	1,364	2,30	161	1,220	1,488	2,22	81	1,095	1,200	2,05	37	1,026	1,053
Der Bezahlvorgang verläuft reibungslos.	1,37	19	0,684	0,468	1,56	72	0,902	0,814	1,56	161	0,886	0,786	1,49	82	0,653	0,426	1,37	38	0,541	0,293
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	1,95	19	1,129	1,275	2,14	73	1,122	1,259	2,20	160	1,212	1,469	2,23	82	1,220	1,489	1,95	38	1,114	1,240
Die Mitarbeiter sind durchweg höflich zu mir.	1,79	19	0,855	0,731	1,97	73	0,866	0,749	1,93	160	0,922	0,850	2,00	81	0,851	0,725	1,82	38	0,865	0,749
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	1,95	19	0,970	0,942	2,27	73	0,821	0,674	2,39	161	0,903	0,815	2,29	82	0,936	0,876	2,26	38	0,921	0,848
Die Mitarbeiter wirken ordentlich.	2,05	19	0,780	0,608	1,96	73	0,696	0,484	2,01	163	0,733	0,537	2,06	82	0,654	0,428	1,97	38	0,636	0,405
Die Mitarbeiter wirken hektisch und überlastet.	3,53	19	1,073	1,152	3,52	73	1,002	1,003	3,72	162	0,893	0,798	3,87	82	0,926	0,858	3,76	38	0,971	0,942
Die Mitarbeiter machen einen freundlichen Eindruck.	2,11	19	0,875	0,766	2,27	73	0,750	0,563	2,29	163	0,844	0,713	2,35	80	0,873	0,762	2,29	38	0,984	0,968
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,21	19	0,787	0,620	2,39	72	0,761	0,579	2,22	161	0,806	0,650	2,22	81	0,822	0,675	2,34	38	0,994	0,988
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,26	19	0,872	0,760	2,30	73	0,681	0,463	2,37	162	0,755	0,570	2,35	82	0,726	0,528	2,32	38	0,775	0,600
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,05	19	0,524	0,275	2,13	72	0,649	0,421	2,41	161	0,754	0,568	2,62	81	0,768	0,589	2,66	38	0,909	0,826

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Tabelle A-2-7: Mittelwert, Standardabweichung, Varianz (Restaurantaffinität)

	Und wie gern essen Sie bei McDonald's?																			
	Sehr gern				Eher gern				Teils-teils				Weniger gern				Gar nicht gern			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Wie ist der Sauberkeitszustand der näheren Umgebung des Restauranteingangs?	1,50	24	0,511	0,261	1,79	99	0,732	0,536	1,72	143	0,791	0,625	1,65	74	0,818	0,669	1,58	31	0,720	0,518
Wie ist der Zustand von Fenstern und Fassaden?	1,58	24	0,776	0,601	1,49	99	0,612	0,375	1,44	143	0,708	0,502	1,58	74	0,740	0,548	1,65	31	0,661	0,437
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	1,92	24	0,717	0,514	2,02	99	0,782	0,612	2,04	142	0,820	0,673	2,11	74	0,820	0,673	2,39	31	0,955	0,912
Wie ist der Sauberkeitszustand des Bodens?	1,67	24	0,637	0,406	1,97	99	0,897	0,805	1,83	142	0,694	0,482	1,96	73	0,807	0,651	1,90	31	0,597	0,357
Wie ist der Zustand des Sitzbereichs?	1,71	24	0,806	0,650	1,66	99	0,717	0,514	1,69	142	0,716	0,513	1,65	74	0,730	0,532	1,77	31	0,762	0,581
Wie ist der Zustand der Servicestation?	1,42	24	0,584	0,341	1,54	98	0,789	0,622	1,55	141	0,603	0,363	1,55	73	0,668	0,446	1,32	31	0,475	0,226
Wie ist der Zustand des Thekenbereichs?	1,17	24	0,381	0,145	1,26	98	0,524	0,274	1,36	141	0,511	0,261	1,32	73	0,524	0,274	1,29	31	0,461	0,213
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	1,96	24	0,955	0,911	2,02	99	0,795	0,632	2,02	142	0,748	0,560	2,08	74	0,790	0,623	2,39	31	0,989	0,978
Beim Herantreten an die Theke werde ich freundlich begrüßt.	2,21	24	1,103	1,216	2,17	99	1,031	1,062	2,13	141	1,041	1,084	2,20	74	1,072	1,150	2,13	31	1,147	1,316
Ich werde freundlich nach meinem Bestellwunsch gefragt.	2,13	24	1,076	1,158	2,05	99	0,930	0,865	1,93	140	0,845	0,714	2,15	74	1,081	1,169	2,10	31	1,044	1,090
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	2,04	24	1,122	1,259	2,04	99	0,989	0,978	2,06	141	1,113	1,240	1,97	74	0,979	0,958	1,77	31	0,920	0,847
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,33	24	1,274	1,623	2,38	98	1,214	1,475	2,12	141	1,118	1,250	2,36	72	1,079	1,164	2,26	31	1,210	1,465
Der Bezahlvorgang verläuft reibungslos.	1,63	24	1,013	1,027	1,53	99	0,849	0,721	1,54	140	0,826	0,682	1,54	74	0,744	0,553	1,23	31	0,425	0,181
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	2,33	24	1,341	1,797	2,17	98	1,167	1,361	2,10	141	1,136	1,290	2,23	74	1,234	1,522	1,97	31	1,169	1,366
Die Mitarbeiter sind durchweg höflich zu mir.	2,08	24	1,100	1,210	1,89	99	0,819	0,671	1,88	139	0,888	0,789	2,07	74	0,881	0,776	1,90	31	0,944	0,890
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	2,17	24	1,049	1,101	2,33	99	0,857	0,735	2,26	141	0,825	0,681	2,43	74	0,994	0,989	2,29	31	1,006	1,013
Die Mitarbeiter wirken ordentlich.	2,29	24	0,859	0,737	1,97	99	0,677	0,458	1,98	143	0,687	0,471	2,05	74	0,719	0,518	2,00	31	0,632	0,400
Die Mitarbeiter wirken hektisch und überlastet.	3,63	24	0,875	0,766	3,56	99	0,992	0,984	3,77	142	0,928	0,861	3,84	74	0,876	0,768	3,65	31	1,018	1,037
Die Mitarbeiter machen einen freundlichen Eindruck.	2,29	24	0,908	0,824	2,20	99	0,808	0,653	2,26	143	0,785	0,616	2,42	74	0,876	0,767	2,45	29	1,121	1,256
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,29	24	0,806	0,650	2,25	97	0,750	0,563	2,24	142	0,798	0,637	2,32	73	0,880	0,774	2,39	31	1,022	1,045
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,46	24	0,932	0,868	2,31	98	0,709	0,503	2,32	143	0,678	0,459	2,41	74	0,792	0,628	2,35	31	0,839	0,703
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,13	23	0,694	0,482	2,23	98	0,701	0,491	2,43	141	0,719	0,518	2,55	74	0,813	0,661	2,71	31	0,973	0,946

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Tabelle F-2-2: Mittelwert, Standardabweichung, Varianz (Häufigkeit eines Fast Food-Restauranbesuchs)

	FastFood Häufigkeitslevel															
	keine Besuche				1-12 mal im Jahr				13-52 mal im Jahr				mehr als 52 mal im Jahr			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Wie ist der Sauberkeitszustand der näheren Umgebung des Resturanteingangs?	1,64	14	0,745	0,555	1,65	186	0,772	0,596	1,76	150	0,757	0,573	1,65	26	0,745	0,555
Wie ist der Zustand von Fenstern und Fassaden?	1,43	14	0,514	0,264	1,48	186	0,699	0,489	1,58	150	0,717	0,514	1,27	26	0,452	0,205
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	2,21	14	0,802	0,643	2,12	185	0,864	0,747	2,03	150	0,785	0,617	1,77	26	0,587	0,345
Wie ist der Sauberkeitszustand des Bodens?	2,00	14	0,555	0,308	1,96	184	0,785	0,616	1,82	150	0,715	0,511	1,77	26	0,951	0,905
Wie ist der Zustand des Sitzbereichs?	1,79	14	0,699	0,489	1,72	185	0,757	0,573	1,62	150	0,682	0,465	1,65	26	0,745	0,555
Wie ist der Zustand der Servicestation?	1,21	14	0,426	0,181	1,51	184	0,644	0,415	1,55	148	0,642	0,412	1,58	26	0,902	0,814
Wie ist der Zustand des Thekenbereichs?	1,14	14	0,363	0,132	1,31	184	0,498	0,248	1,31	148	0,506	0,256	1,31	26	0,618	0,382
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	2,21	14	0,975	0,951	2,14	185	0,822	0,676	1,96	150	0,732	0,535	1,88	26	0,952	0,906
Beim Herantreten an die Theke werde ich freundlich begrüßt.	1,93	14	1,141	1,302	2,19	184	1,097	1,204	2,20	150	1,056	1,114	1,85	26	0,732	0,535
Ich werde freundlich nach meinem Bestellwunsch gefragt.	1,79	14	0,699	0,489	2,07	183	0,987	0,974	2,02	150	0,959	0,919	1,96	26	0,824	0,678
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	1,93	14	1,072	1,148	2,05	184	0,965	0,932	1,97	150	1,120	1,254	2,15	26	1,084	1,175
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,14	14	1,099	1,209	2,34	183	1,151	1,326	2,25	148	1,217	1,481	1,92	26	0,935	0,874
Der Bezahlvorgang verläuft reibungslos.	1,36	14	0,497	0,247	1,55	184	0,745	0,555	1,49	149	0,875	0,765	1,50	26	0,906	0,820
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	1,93	14	1,072	1,148	2,24	184	1,210	1,464	2,11	149	1,203	1,448	2,04	26	0,999	0,998
Die Mitarbeiter sind durchweg höflich zu mir.	1,71	14	0,611	0,374	1,98	183	0,886	0,785	1,95	149	0,943	0,889	1,69	26	0,679	0,462
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	2,14	14	0,949	0,901	2,37	184	0,949	0,901	2,29	150	0,886	0,786	2,15	26	0,613	0,375
Die Mitarbeiter wirken ordentlich.	2,00	14	0,784	0,615	1,99	186	0,701	0,492	2,03	150	0,723	0,523	2,04	26	0,528	0,278
Die Mitarbeiter wirken hektisch und überlastet.	3,64	14	0,929	0,863	3,73	185	0,922	0,850	3,71	150	0,966	0,934	3,50	26	1,105	1,220
Die Mitarbeiter machen einen freundlichen Eindruck.	2,07	14	0,829	0,687	2,34	184	0,885	0,784	2,27	150	0,835	0,697	2,15	26	0,613	0,375
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,21	14	0,802	0,643	2,25	185	0,809	0,655	2,32	148	0,867	0,751	2,16	25	0,688	0,473
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,29	14	0,726	0,527	2,37	186	0,790	0,624	2,36	150	0,717	0,514	2,12	25	0,526	0,277
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,50	14	0,760	0,577	2,54	184	0,829	0,687	2,26	148	0,664	0,440	2,27	26	0,724	0,525

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Tabelle F-2-9: Mittelwert, Standardabweichung, Varianz (spezifische Restaurant-Besuchshäufigkeit)

	Spezifische Restaurant-Besuchshäufigkeit															
	keine Besuche				1-12 mal im Jahr				13-52 mal im Jahr				mehr als 52 mal im Jahr			
	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V	MW	N	SD	V
Wie ist der Sauberkeitszustand der näheren Umgebung des Resturanteingangs?	1,53	36	0,810	0,656	1,68	255	0,752	0,565	1,81	80	0,765	0,585	1,80	5	0,837	0,700
Wie ist der Zustand von Fenstern und Fassaden?	1,36	36	0,683	0,466	1,53	255	0,698	0,487	1,54	80	0,674	0,454	1,00	5	0,000	0,000
Welchen Eindruck macht der äußere Zustand des Restaurants insgesamt auf Sie?	2,17	36	0,845	0,714	2,05	254	0,825	0,681	2,09	80	0,783	0,613	1,40	5	0,548	0,300
Wie ist der Sauberkeitszustand des Bodens?	1,81	36	0,668	0,447	1,96	253	0,783	0,614	1,74	80	0,725	0,525	1,40	5	0,548	0,300
Wie ist der Zustand des Sitzbereichs?	1,58	36	0,806	0,650	1,70	254	0,735	0,541	1,63	80	0,644	0,415	1,80	5	0,837	0,700
Wie ist der Zustand der Servicestation?	1,25	36	0,439	0,193	1,55	251	0,633	0,401	1,56	80	0,793	0,629	1,40	5	0,548	0,300
Wie ist der Zustand des Thekenbereichs?	1,17	36	0,378	0,143	1,35	252	0,525	0,276	1,22	79	0,443	0,197	1,40	5	0,894	0,800
Welchen Eindruck macht der Zustand der Innenräume des Restaurants insgesamt auf Sie?	2,03	36	0,971	0,942	2,10	254	0,779	0,606	1,90	80	0,821	0,673	2,20	5	0,447	0,200
Beim Herantreten an die Theke werde ich freundlich begrüßt.	1,97	36	1,082	1,171	2,22	253	1,079	1,165	2,05	80	1,005	1,010	2,20	5	0,837	0,700
Ich werde freundlich nach meinem Bestellwunsch gefragt.	1,89	36	1,008	1,016	2,06	252	0,955	0,912	1,99	80	0,948	0,899	2,00	5	0,707	0,500
Der Mitarbeiter nimmt beim Bestellgespräch Augenkontakt mit mir auf.	1,94	36	0,924	0,854	2,03	253	1,002	1,003	1,98	80	1,180	1,392	2,80	5	1,304	1,700
Der Mitarbeiter nennt mir in höflicher Weise den Preis.	2,21	34	1,149	1,320	2,36	253	1,166	1,359	2,03	79	1,154	1,333	1,80	5	0,837	0,700
Der Bezahlvorgang verläuft reibungslos.	1,50	36	0,655	0,429	1,55	253	0,842	0,709	1,42	79	0,744	0,554	1,40	5	0,548	0,300
Bei der Übergabe meiner Bestellung werde ich freundlich verabschiedet.	1,92	36	1,131	1,279	2,24	252	1,231	1,515	2,05	80	1,090	1,187	1,80	5	0,447	0,200
Die Mitarbeiter sind durchweg höflich zu mir.	1,67	36	0,756	0,571	2,01	252	0,910	0,829	1,87	79	0,868	0,753	1,40	5	0,548	0,300
Welchen Eindruck macht das Kundengespräch insgesamt auf Sie?	2,17	36	1,056	1,114	2,37	253	0,906	0,821	2,21	80	0,837	0,701	2,40	5	0,548	0,300
Die Mitarbeiter wirken ordentlich.	1,86	36	0,723	0,523	2,04	255	0,712	0,507	2,00	80	0,656	0,430	2,00	5	0,707	0,500
Die Mitarbeiter wirken hektisch und überlastet.	3,86	36	0,833	0,694	3,71	254	0,975	0,950	3,58	80	0,938	0,880	4,20	5	0,447	0,200
Die Mitarbeiter machen einen freundlichen Eindruck.	2,15	34	1,019	1,038	2,34	255	0,835	0,697	2,24	80	0,815	0,664	1,80	5	0,447	0,200
Die Zusammenarbeit der Mitarbeiter macht einen guten Eindruck.	2,08	36	0,874	0,764	2,34	253	0,827	0,684	2,17	78	0,780	0,608	2,00	5	0,707	0,500
Welchen Eindruck machten die Mitarbeiter insgesamt auf Sie?	2,19	36	0,856	0,733	2,39	255	0,739	0,546	2,30	79	0,722	0,522	2,00	5	0,000	0,000
Welchen Eindruck macht der Besuch insgesamt auf Sie?	2,44	36	0,939	0,883	2,48	253	0,780	0,608	2,17	78	0,590	0,348	2,20	5	0,447	0,200

MW = Mittelwert; S = Standardabweichung; V = Varianz

Quelle: Eigene Darstellung

Anhang 3 – Abgrenzung zu der Arbeit von Schmidt (2007)

Sowohl in dieser als auch in der Untersuchung von Schmidt wird der SERVPERF-Fragebogen eingesetzt. Dies erleichtert eine Vergleichbarkeit der ermittelten Ergebnisse im Rahmen zukünftiger Forschungsvorhaben. Zudem wurde SERVPERF bereits in vergangenen Mystery Shopping-Untersuchungen im Einzelhandel eingesetzt⁴⁷⁷, so dass vier Arbeiten für solche Vergleichsstudien zur Verfügung stehen. Die Darstellung der theoretischen Grundlagen ist aus Gründen der Vollständigkeit erforderlich und erfolgt in dieser Arbeit in ähnlicher Weise wie bei Schmidt, allerdings deutlich kompakter.

Schmidts Arbeit widmet sich insbesondere den Aspekten der Reliabilität und Validität des Dienstleistungsqualitätsurteils von Testkunden in Reisebüros. Dabei werden u.a. Interraterreliabilität und Beurteilerübereinstimmung empirisch überprüft. Die theoretischen Grundlagen und die praktische Umsetzung der Messung dieser Gütekriterien werden ohne vergleichbares Beispiel von Wirtz/Caspar dargestellt.⁴⁷⁸ Schmidt überträgt dies auf den Einsatz bei Testkunden. In der vorliegenden Arbeit werden diese (neben anderen) als Komponenten der Ergebnisqualität herangezogen. Dabei erfolgt eine sehr kompakte Darstellung der theoretischen Grundlagen mit Verweis auf Wirtz/Caspar sowie auf Schmidts Leistungen.

Die Anforderungen und Voraussetzungen für den erfolgreichen Einsatz von Mystery Shopping fußen insbesondere auf den Arbeiten von Matzler et al.⁴⁷⁹ Schmidt greift diese ebenfalls auf. In der vorliegenden Arbeit war eine Aufnahme unerlässlich. Dabei wurde die Wiedergabe dieser Grundlagen um zahlreiche Beispiele ergänzt, die vor allem bei den praktischen Beiträgen zu finden sind. Ferner wurden neuartige Überlegungen zu unterschiedlichen Aspekten angestellt. In diesem Zusammenhang ist auch die Würdigung der historischen Entwicklung von Mystery Shopping in der Literatur anzuführen.⁴⁸⁰

Insgesamt befasst sich die vorliegende Arbeit im Vergleich zu der Arbeit Schmidts deutlich intensiver mit divergenten Facetten der Mystery Shopping-Thematik. Die empirische Analyse in dieser Arbeit ist hinsichtlich Fragestellung und Methodik überhaupt nicht mit derjenigen in der Arbeit Schmidts vergleichbar.

⁴⁷⁷ Vgl. Lowndes/Dawes (1999) und Dawes/Sharp (2000)

⁴⁷⁸ Vgl. Wirtz/Caspar (2002)

⁴⁷⁹ Vgl. Matzler/Kittinger-Rosanelli (2000: 226ff.), Matzler/Pechlaner/Kohl (2000: 164ff.)

⁴⁸⁰ Die historische Entwicklung ist lediglich bei Deckers (1999) zu finden, den Schmidt in seiner Arbeit nicht anführt.

Literaturverzeichnis

- ACHELIS, T. (1975). Was leistet die Bankberatung? Erfolgskontrolle mit fingierten Schaltergesprächen. In: *Bank und Markt* 4 (1), S. 15-16
- ACHELIS, T. (1985). Testgespräche am Bankschalter: Aufwand lohnt sich. In: *Bank und Markt* 14 (2), S. 26-28
- ALBERS, S.; HILDEBRANDT, L. (2006). Methodische Probleme bei der Erfolgsfaktorenforschung : Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells. In: *Zfbf* 58 (1), S. 2-33
- ALMHOFER, H. (2007). Mystery Shopping – “just in time” Information mit dem Augen des Kunden. In: *Verband der Marktforscher Österreichs (Hrsg.). Handbuch der Marktforschung.* Wien: Facultas.wuv, S. 33-38
- ANDERSON, D.N.; GROVES, D.L.; LENGFELDER, J.; TIMOTHY, D. (2001). A Research Approach to Training: A Case Study of Mystery Guest Methodology. In: *International Journal of Contemporary Hospitality Management* 13 (2), S. 93-102
- ANDERSON, E.W.; MITTAL, V. (2000): Strengthening the Satisfaction-Profit Chain. In: *Journal of Service Research* 3 (2), S. 107-120.
- ARBEITSKREIS DEUTSCHER MARKT- UND SOZIALFORSCHER (2006). Richtlinie für Beobachtungen bei demoskopischen Untersuchungen. Elektronisch veröffentlicht unter der URL http://www.adm-ev.de/pdf/R02_D.pdf (Stand 02.08.2007)
- BACKHAUS, K.; ERICHSON, B.; PLINKE, W.; WEIBER, R. (2006). *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung.* Berlin: Springer
- BAUER, F., SCHNEIDER, P., HERRMANN, T. (2002). Kundenbindung durch Servicequalität. In: *Planung & Analyse* 29 (4), S. 12-18
- BAUER, F., URBAHN, J., MARKART, V. (2003). Die zehn häufigsten Missverständnisse zum Thema Mystery Analysen. In: *Planung & Analyse* 30 (5), S. 16-23
- BECK, J.; MIAO, L. (2003). Mystery Shopping in Lodging Properties as a Measurement of Service Quality. In: *Journal of Quality Assurance in Hospitality and Tourism* 4 (1/2), S. 1-22
- BECKER, W. (1973). *Beobachtungsverfahren in der demoskopischen Marktforschung.* Stuttgart: Ulmer
- BECKMANN, J. (2003) Mystery Calls zur Überprüfung der Service-Qualität für Call-Center. In: *Planung & Analyse* 30 (5), S.39-44
- BEISTER, P. (2003). *Testkäufe.* Stuttgart: DSV
- BEREKOVEN, L.; ECKERT, W.; ELLENRIEDER, P. (2004): *Marktforschung. Methodische Grundlagen und praktische Anwendung.* Wiesbaden: Gabler
- BERGER, R. (1958). Der Kunden-Test. *Jahrbuch der Absatz- und Verbrauchsforschung* 4, S. 122-136
- BEUTIN, N. (2006). Verfahren zur Messung der Kundenzufriedenheit im Überblick. In: *Homburg, C. (Hrsg.). Kundenzufriedenheit : Konzepte – Methoden – Erfahrungen.* Wiesbaden : Gabler, S. 121-169
- BORTZ, J.; DÖRING, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Heidelberg: Springer.

- BÖHM, D.N.; FISCHL, C.; RENNHAKE, C. (2006). Kundenzufriedenheitsmessung mittels Mystery Shopping. In: Rennhak, C. (Hrsg.). Herausforderung Kundenbindung. Wiesbaden: Dt. Univ.-Verlag. S. 87-93
- BÖSSOW, O., ZINNEGER, U. (2003). Auf den richtigen Service kommt es an. In: Planung & Analyse 30 (5), S.66-69
- BRADY, M.K.; CRONIN, J.J.; BRAND, R.R. (2002): Performance-only Measurement of Service Quality: A Replication and Extension. In: Journal of Business Research 55 (1), S. 17-35
- BRANDSTÄTTER, H. (1970). Die Beurteilung von Mitarbeitern. In: Mayer, A.; Herwig, B. (Hrsg.). Handbuch der Psychologie, Bd. 9: Betriebspsychologie. Göttingen: Hogrefe, S. 668-734
- BRÜCHER, H. (1962). Überwachung der Preisbindung durch Testkäufe. In: Betriebs-Berater 17 (3), S.119-121
- BRUHN, M. (2000): Qualitätssicherung im Dienstleistungsmarketing. Eine Einführung in die theoretischen und praktischen Probleme. In: Bruhn, M., Stauss, B. (Hrsg.). Dienstleistungsqualität. Grundlagen – Konzepte – Methoden. Wiesbaden: Gabler, S. 21–48
- BRUHN, M. (2001). Marketing – Grundlagen für Studium und Praxis. Wiesbaden: Gabler
- BRUHN, M. (2003). Messung der Anforderungen an die Dienstleistungsqualität. In: Hansen, W.; Kamiske, G. F. (Hrsg.): Qualitätsmanagement im Dienstleistungsbereich. Düsseldorf: Symposium, S. 7-43
- BRUHN, M. (2006): Qualitätsmanagement für Dienstleistungen: Grundlagen – Konzepte – Methoden. Berlin: Springer
- BRUHN, M., GEORGI, D. (1998). Kundenbezogene Wirtschaftlichkeitsanalyse des Qualitätsmanagements für Dienstleistungen; Konzept, Modellrechnung und Fallbeispiel. In: Marketing ZFP 20 (2), S. 98-108
- BUHL, H.U.; MELLWIG, N.; RENZ, A. (2003). Ein Modell zur Bewertung unterschiedlicher Zufriedenheitsquellen unter besonderer Berücksichtigung des M- und E-Business. Diskussionspapier WI-129 der Universität Augsburg
- CALVERT, P. (2005). It's a Mystery – Mystery Shopping in New Zealand's Public Libraries. In: Library Review 54 (1), S. 24-35
- CARLZON, J. (1987). Moments of Truth. Cambridge: Ballinger
- CHIN, W.W.; NEWSTED, P.R. (1999). Structural Equation Modelling Analysis with Small Samples Using Partial Least Squares. In: Hoyle, R.H. (Hrsg.). Statistical Strategies for Small Sample Research. Thousand Oaks: Sage, S. 307-341
- CHIN, W.W. (2004). Frequently Asked Questions – Partial Least Squares & PLS-Graph. <http://disc-nt.cba.uh.edu/chin/plsfaq/plsfaq.htm>
- CHURCHILL, G.A. JR. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. In: Journal of Marketing Research 16 (1), S. 64-73
- COLLINS, J. (2003). Mystery Shopping: Was steckt wirklich dahinter? In: Planung & Analyse 30 (5), S. 30-33
- COOK, L.S.; BOWEN, D.E.; CHASE, R.B.; DASU, S.; STEWART, D.M.; TANSIK, D.A. (2002). Human Issues in Service Design. In: Journal of Operations Management 20 (2), S. 159-174
- CORSTEN, H. (2001). Dienstleistungsmanagement. München: Oldenbourg

- CRONBACH, L.J. (1955). Processes Affecting Scores on "Understanding of Others" and "Assumed Similarity". In: *Psychological Bulletin* 52 (3), S. 177-197
- CRONIN, J.J.; TAYLOR, S.A. (1992). Measuring Service Quality: A Reexamination and Extension. In: *Journal of Marketing* 56 (3), S. 55-68
- CRONIN, J.J.; TAYLOR, S.A. (1994). SERVPERF Versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality. In: *Journal of Marketing* 58 (1), S. 125-131
- CZOPEK, V. (1998). Using Mystery Shoppers to Evaluate Customer Service in Public Libraries. In: *Public Libraries* 37 (6), S. 370-375
- DAWES, J.; SHARP, B. (2000). The Reliability and Validity of Objective Measures of Customer Service: "Mystery Shopping". In: *Australian Journal of Market Research* 8 (1), S. 29-46
- DAWSON, J.; HILLIER, J. (1995). Competitor Mystery Shopping: Methodological Considerations and Implications for the MRS Code of Conduct. In: *Journal of the Market Research Society* 37 (4), S. 417-427
- DECKERS, R. (1996). Der Soziologe als Methodenspezialist. In: *Sozialwissenschaften und Berufspraxis* 19 (1), S. 25-31
- DECKERS, R. (1999). Das Testkundenverfahren: Bestandsaufnahme, Methodenprobleme, Qualitätssicherung. Dissertation: Köln
- DECKERS, R. (2003). Taking the Mystery out of Mystery Shopping. In: *Planung & Analyse* 30 (5), S.34-38
- DECKERS, R., DECKERS, S. (2004). Die Beteiligungsrechte des Betriebsrats beim Testkauf. In: *Neue Zeitschrift für Arbeitsrecht* 21 (3), S. 139-142
- DECKERS, R.; HEINEMANN, G. (2006). *Mystery Shopping: mit Testkäufern Verkauf und Service nachhaltig verbessern*. Göttingen: BusinessVillage
- DEGES, F. (1992). Der Einsatz von Testkunden im Einzelhandel. In: *Jahrbuch der Absatz- und Verbrauchsforschung* 38 (1), S. 85-100
- DEUTSCHE GESELLSCHAFT FÜR QUALITÄT E.V. (1995): *Begriffe zum Qualitätsmanagement*. DGQ-Schrift, Nr. 11-04. Frankfurt/Main: Beuth
- DIAMANTOPOULOS, A.; WINKLHOFER, H. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. In: *Journal of Marketing Research* 38 (2), S. 269-277
- DIEHL, J.M.; STAUFENBIEL, T. (2007). *Statistik mit SPSS für Windows*. Eschborn: Klotz
- DIEKMANN, A. (2007). *Empirische Sozialforschung : Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt
- DOBBELSTEIN, T.; WINDBACHER, T. (2007). Mystery Shopping – Ziele, Prozess und Qualität eines Verfahrens zum Controlling der Dienstleistungsqualität. In: Schuckel, M.; Toporowski, W. (Hrsg.). *Theoretische Fundierung und praktische Relevanz der Handelsforschung*. Wiesbaden: Dt. Univ.-Verlag
- DORMAN, K.G. (1994). Mystery Shopping Results Can Shape Your Future. In: *Bank Marketing* 26 (8), S. 17-21
- DOSTERT, E. (2005). Noten vom Detektiv. In: *Süddeutsche Zeitung* vom 11.3.2005, elektronisch veröffentlicht unter <http://www.sueddeutsche.de/wirtschaft/artikel/282/49233/> (Stand 18.5.2007)

- DOUGLAS, A.; DOUGLAS, J. (2006). Campus Spies? Using Mystery Students to Evaluate University Performance. In: Educational Research 48 (1), S. 111-119
- DREES, N.; SCHILLER, S. (2000). Mystery Shopping - Ein Instrument zur Überprüfung der Servicequalität. In: Absatzwirtschaft 43 (9), S. 66-71
- DREES, N.; SCHILLER, S. (2001). Mystery Shopping: Wie Dienstleistungsqualität messbar wird. In: Transfer – Werbeforschung & Praxis 46 (1), S. 6-11
- DREES, N.; SCHILLER, S. (2003). Mystery Shopping – Ein Instrument zur systematischen Optimierung von Kundenzufriedenheit im Dienstleistungsbereich. In: Kamenz, U. (Hrsg.). Applied Marketing. Berlin: Springer, S. 159-172
- DREYER, A.; LINNE, M. (2004). Der Einsatz von Mystery Guests in touristischen Destinationen – Ein Beitrag zur Qualitätsentwicklung. In: Hinterhuber, H. et al. (Hrsg.). Kundenmanagement als Erfolgsfaktor – Grundlagen des Tourismusmarketing. Berlin: Schmidt
- DUTTENHÖFER, S.; LERCH, S. (2000). Anlageberatung: Qualitätssicherung durch Testberatungen. In: bank & markt 29 (11), S. 30-34
- DYKE, T.P. VAN; KAPPELMAN, L.A.; PRYBUTOK, V.R. (1997). Measuring Information Systems Service Quality: Concerns on the Use of the SERVQUAL Questionnaire. In: MIS Quarterly 21 (2), S. 195-208
- DYKE, T.P. VAN; PRYBUTOK, V.R.; KAPPELMAN, L.A. (1999). Cautions on the Use of SERVQUAL Measure to Assess the Quality of Information Systems Services. In: Decision Sciences 30 (3), S. 877-891
- EBERL, M. (2006). Formative und reflektive Konstrukte und die Wahl des Strukturgleichungsverfahrens : eine statistische Entscheidungshilfe. In: Die Betriebswirtschaft 66 (6), S. 651-668
- EBERS, M.; GOTSCH, W. (2006). Institutionenökonomische Theorien der Organisation. In: Kieser, A.; Ebers, M. (Hrsg.). Organisationstheorien. Stuttgart: Kohlhammer
- ECKHARDT, V. (1968). Die Befugnis des Preisbinders zur Überwachung durch Testkäufe. Dissertation: Göttingen:
- EICHHOLZ, R. (1966). Testkunden + Kundentests. In: Der Marktforscher 10 (Februar), S. 34-40
- EIKENROTH, O. (2004). Mystery Customer – Möglichkeiten und Grenzen in der Marktforschung. Göttingen: ACU
- ERSTAD, M. (1998) Mystery Shopping Programmes and Human Resource Management. In: International Journal of Contemporary Hospitality Management 10 (1), S. 34-38
- ESOMAR (2005). ESOMAR World Research Codes and Guidelines: Mystery Shopping Studies. Elektronisch veröffentlicht unter der URL „http://www.esomar.org/web/show/file/id=47043/filename=ps_cg_mysteryshopping.pdf/page=43240“ (Stand 18.06.2007)
- ESOMAR (2007). Global Market Research 2007 – ESOMAR Industry Report. ESOMAR/KPMG Business Advisory Services
- FASSOTT, G. (2006). Operationalisierung latenter Variablen in- Strukturgleichungsmodellen: Eine Standort-bestimmung. In: Zfbf 58 (1), S. 67-88
- FERCHLAND, H.; PAUL, O. (2006). Sind Testkäufe ein geeignetes Instrument im Neukundengeschäft? In: Bank und Markt 36 (8), S. 34-37

- FINN, A.; KAYANDÉ, U. (1999). Unmasking a Phantom: A Psychometric Assessment of Mystery Shopping. In: *Journal of Retailing* 75 (2), S. 195-217
- FINN, A. (2001) Mystery Shopper Benchmarking of Durable-Goods Chains and Stores. In: *Journal of Service Research* 3 (4), S. 310-320
- FISCHER, M.; HERRMANN, A.; HUBER, F. (2001). Return on Customer Satisfaction. In: *Zeitschrift für Betriebswirtschaft* 71 (10), S. 1161-1190.
- GEIGER, S. (1962). Testkäufe und Verkaufstests. In: *Der Marktforscher* 6 (1), S. 101-102
- GEIBLER, J.; DINGLER, D. (1969). Ist Freundlichkeit messbar? Bericht über ein neuartiges Testinstrument. In: *Supermarkt und Selbstbedienung* 8 (8), S. 59-62
- GEIBLER, C. (2003). Mystery Shopper. In: *Harvard Business Manager* 5/2003, elektronisch veröffentlicht unter der URL http://www.harvardbusinessmanager.de/go/see!_DIG26963824 (Stand 17.12.2006)
- GEIBLER, H.; THOMAS, C.; DUSCHA, A. (2005). Emails unter der Lupe. In: *Planung & Analyse* 32 (1), S. 64-69
- GIRTLER, R. (1992). *Methoden der qualitativen Sozialforschung*. Wien
- GÖTZ, O.; LIEHR-GOBBER, K. (2004). Analyse von Strukturgleichungsmodellen mit Hilfe der Partial-Least-Squares(PLS)-Methode. In: *Die Betriebswirtschaft* 64 (6), S. 714-738
- GREVE, W.; WENTURA, D. (1991). *Wissenschaftliche Beobachtung in der Psychologie: Eine Einführung*. München: Quintessenz
- GROVE, S.J.; FISK, R.P. (1992). Observational Data Collection Methods for Services Marketing: An Overview. In: *Journal of the Academy of Marketing Science* 20 (3), S. 217-224
- GOHR, S. (2006). Mystery Calls - das Konzept zählt! Kundenservice. In: *Call-Center-Profi* 5/2006, S. 32-36
- GONCOURT, E., GONCOURT, J. (1983). *Tagebücher. Aufzeichnungen aus den Jahren 1851-1870*. Frankfurt/M.: Insel
- GUBER, A.; KERPER, M. (2003). Mystery Shopping als Neuland bei Versicherungen: Gothaer macht den Anfang. In: *Planung & Analyse* 30 (5), S.62-65
- GÜTHOFF, J. (1995). *Qualität komplexer Dienstleistungen*. Dissertation: Rostock
- HAAS, A. (2001). Wie entsteht Beratungszufriedenheit? : Ergebnisse einer Mystery Shopping-Studie im Gebrauchsgüterhandel. In: Diller, H. (Hrsg.). *Der moderne Verbraucher - neue Befunde zum Einkaufsverhalten*. Nürnberg: GIM, S.87-106
- HAAS, A. (2002). Analyse von Verkaufssituationen mit Mystery Shopping. In: *Jahrbuch der Absatz- und Verbrauchsforschung* 48 (3), S.277-294
- HAAS, A. (2003). Wie valide sind Urteile von Mystery Shoppern? In: *Planung & Analyse* 30 (5), S. 24-29
- HAAS, A. (2006). Bestimmungsfaktoren des Beratungserfolges: Eine informationsökonomische Betrachtung und empirische Analyse im Handel. In: *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 58 (8), S. 638-664
- HAASE, A. (2002). *Make-or-buy-Entscheidung für die Unternehmensberatung : ein Prinzipal-Agent-theoretischer Strukturierungsansatz*. Dissertation: Marburg
- HACKMANN, C.; SCHÄBE, H. (1996). Testkunden statt Kundenumfragen. In: *Qualität und Zuverlässigkeit* 41 (10). S. 1136-1140

- HALL, E.J. (1993). Smiling, Deferring and Flirting: Doing Gender by Giving Good Service. In: *Work & Occupations* 20 (4). S. 452-471
- HARTUNG, J.; ELPELT, B.; KLÖSENER, K.-H. (2005). *Statistik: Lehr- und Handbuch der angewandten Statistik*. München: Oldenbourg
- HAZLERIG, A. (2007). *Introduction to Mystery Shopping: The Perfect Home-based Business*. Charleston: BookSurge
- HENTSCHEL, B. (1990). Die Messung wahrgenommener Dienstleistungsqualität mit SERVQUAL. Eine kritische Auseinandersetzung. In: *Marketing – Zeitschrift für Forschung und Praxis* 12 (4), S. 230-240
- HENTSCHEL, B. (1992). *Dienstleistungsqualität aus Kundensicht: Vom merkmals- zum ereignisorientierten Ansatz*. Dissertation: Eichstätt
- HENTSCHEL, B. (2000). Multiattributive Messung von Dienstleistungsqualität. In: Bruhn, M., Stauss, B. (Hrsg.). *Dienstleistungsqualität. Grundlagen – Konzepte – Methoden*. Wiesbaden: Gabler, S. 289-320
- HERRMANN, A.; HUBER, F; KRESSMANN, F. (2006). Varianz- und kovarianzbasierte Strukturgleichungsmodelle : ein Leitfaden zu deren Spezifikation, Schätzung und Beurteilung. In: *Zfbf* 58 (1), S. 34-66
- HESELINK, M.; WIELE, T. VAN DER (2003). *Mystery Shopping : In-Depth Measurement of Customer Satisfaction*. Erasmus Report Series Research in Management, ERS-2003-020-ORG
- HESELINK, M.; IWAARDEN, J. VAN; WIELE, T. VAN DER (2004) *Mystery Shopping: A Tool to Develop Insight into Customer Service Provision*. Erasmus Report Series Research in Management, ERS-2004-082-ORG
- HESSLER, A. (1999). *Mystery Shopping – Was falsch läuft im Verkaufsgespräch*. In: *Absatzwirtschaft* 42 (11), S. 60-63
- HÖHNER, J. (2003). *Mystery Research als Controlling-Instrument im Beschwerdemanagement*. In: *Planung & Analyse* 30 (5), S.51-54
- HÖHNER, J., SCHAPER, C. (2004): *Vom Kundenfrust zur Kundenlust: Mystery Research als Instrument zur Messung der Servicequalität am Point-of-Sale*. In: *Planung & Analyse* 31 (4), S. 32-37
- HOLTZ, M.J. (2003). *Stop der Aldisierung mit Hilfe von Mystery Shopping*. In: *Planung & Analyse* 30 (5), S.75-77
- HOMBURG, C.; BUCERIUS, M. (2006). *Kundenzufriedenheit als Managmentherausforderung*. In: Homburg, C. (Hrsg.). *Kundenzufriedenheit. Konzepte – Methoden – Erfahrungen*. Wiesbaden: Gabler, S. 53-89
- HOMBURG, C.; HERRMANN, A.; PFLESSER, C. (2000). *Methoden der Datenanalyse im Überblick*. In: *Marktforschung: Methoden, Anwendung, Praxisbeispiele*. Wiesbaden: Gabler, S. 101-125
- HOMBURG, C.; KROHMER, H. (2006). *Marketingmanagement. Strategie - Instrumente - Umsetzung – Unternehmensführung*. Wiesbaden: Gabler
- HOMBURG, C.; PFLESSER, C. (2000). *Strukturgleichungsmodelle mit latenten Variablen: Kausalanalyse*. In: *Marktforschung: Methoden, Anwendung, Praxisbeispiele*. Wiesbaden: Gabler, S. 633-659

- HOMBURG, C.; RUDOLPH, B. (1995): Theoretische Perspektiven zur Kundenzufriedenheit. In: Simon, H.; Homburg, Ch. (Hrsg.): Kundenzufriedenheit: Konzepte – Methoden – Erfahrungen. Wiesbaden: Gabler, S. 29-49
- HOMBURG, C.; RUDOLPH, B.; WERNER, H. (1995): Messung und Management von Kundenzufriedenheit in Industriegüterunternehmen In: Simon, H.; Homburg, Ch. (Hrsg.): Kundenzufriedenheit: Konzepte – Methoden – Erfahrungen. Wiesbaden: Gabler, S. 313-340
- HUBER, F.; HEITMANN, M.; HERRMANN, A. (2006). Ansätze zur Kausalmodellierung mit Interaktionseffekten : ein Überblick. In: Die Betriebswirtschaft 66 (6), S. 696-710
- HUBER, F.; HERRMANN, A.; MEYER, F.; VOGEL, J.; VOLLHARDT, K. (2007). Kausalmodellierung mit Partial Least Squares. Wiesbaden: Gabler
- HÜRLIMANN, F.W.; VON HEBENSTREIT, B. (1987). Verkehrssicherheitsarbeit in der Praxis. Bern: Huber
- INFAS TTR (2005). Wie viel Marktforschung braucht und nutzt der Vertrieb? Der Weg zum Vertriebsserfolg. Frankfurt: Eigene Studie
- JAUSCHOWETZ, D. (1978). Mitarbeiterforschung bei Kreditinstituten. Bank und Markt 7 (6), S. 34-37
- JOHANNES, H. (1961). Die Überwachung der Preisbindung durch Testkäufe. In: Betriebs-Berater 16 (15), S.577-580
- KELLER, H. U. (1982). Der Schaltertest. Instrument zur Kontrolle des persönlichen Verkaufs im Bankmarketing. Dissertation: St. Gallen
- KELLER, B.; STÜCKEN, M. (2004). Mystery Shopping (Testberatungen) als Instrument des Vertriebscontrollings. In: Duttonhöfer, S; Keller, B.; Beitner, R. P. (Hrsg.). Handbuch Vertriebsmanagement Finanzdienstleistungen. Frankfurt: Knapp, S.525-548
- KIM, Y.P.; LEE, S.H.; YUN, D.G. (2004). Integrating Current and Competitive Service-Quality Level Analyses for Service-Quality Improvement Programs. In: Managing Service Quality 14 (4), S. 288-296
- KOHLBACHER, J., EGLOFF, N. (2003). Mystery Shopping als Instrument strategischer Unternehmensführung. In: Planung & Analyse 30 (5), S.70-74
- KRAFFT, M. (1999). Der Kunde im Fokus: Kundennähe, Kundenzufriedenheit, Kundenbindung – und Kundenwert? In: Die Betriebswirtschaft 59 (4), S. 511-530
- KROMREY, H. (2006). Empirische Sozialforschung: Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung. Stuttgart: Lucius & Lucius
- KRÜGER, S.; STÜCKEN, M. (2003). Testberatungen: wo sind die Grenzen? In: Bank und Markt 33 (9), S. 32-36
- KÜPPERS, B. (1977). Erfahrungen mit Schaltertests bei Kreditinstituten. In: Bank und Markt 6 (3), S. 32-33
- LAMMERS, M., SCHUBERT, A. (2003). Am Puls der Dienstleistungsqualität mit Mystery Shopping. In: Planung & Analyse 30 (5), S.56-60
- LANDGRAF, P. (1995): Ansätze zur Verbesserung der Dienstleistungsqualität am Beispiel der Schalterdienste der Deutsche Post AG. Dissertation: München
- LANDWEHR, F. (1984). Einsatz von Kontrollkäufern. In: Dynamik im Handel 28 (2), S. 23-26
- LEEDS, B. (1995). Mystery Shopping: From Novelty to Necessity. In: Bank Marketing 27 (6), S. 17-23

- LINGENFELDER, M.; WIESEKE, J.; SCHMIDT, K. (2003). Dienstleistungsqualität von Reisebüro-Unternehmen: Modellierung und Messung mittels Mystery Shopping; am Fallbeispiel einer im Franchisesystem organisierten Reisebüro-Kette. In: *Tourismus-Journal* 7 (3), S.283-306
- LINGENFELDER, M.; SCHMIDT, K.; WIESEKE, J. (2005). Mitarbeiter-Performance im Servicekontakt - Modellierung und Messung mittels Mystery Shopping im Tourismus. In: Bruhn, M. (Hrsg.). *Forum Dienstleistungsmanagement*. Wiesbaden: Gabler, S.181-210
- LORENTZ, F. (2001). Dienstleister unter der Lupe. In: *Die Welt am Sonntag* vom 28.1.2001, elektronisch veröffentlicht unter der URL <http://www.wams.de/data/2001/01/28/469872.html> (Stand 16.12.2006)
- LOWNDES, M. (2000). The Impact of Negatively Directed Statements on Service Quality Ratings. Conference Paper of Australian and New Zealand Marketing Academy Conference, Gold Coast, S. 724-727
- LOWNDES, M.; DAWES, J. (1999). Do SERVPERF Dimensions Emerge from Mystery Shopping Data? Conference Paper of Australian and New Zealand Marketing Academy Conference, Sydney
- LUBIN, P.C. (2001). What's the Mystery? In: *ABA Bank Marketing* 33 (7), S. 28-33
- MARCUS, S.; SCHULER, H. (2006). Leistungsbeurteilung. In: Schuler, H. (Hrsg.). *Lehrbuch Personalpsychologie*. Göttingen: Hogrefe, S. 433-470
- MARKET RESEARCH SOCIETY (2003): Guidelines on Mystery Customer Research. Elektronisch veröffentlicht unter der URL: „<http://www.mrs.org.uk/standards/downloads/revise/active/Mystery%20shopping%20guidelines.doc>“ (Stand 18.06.2007)
- MATZLER, K.; KITTINGER-ROSANELLI, C. (2000): Mystery Shopping als Instrument zur Messung der wahrgenommenen Dienstleistungsqualität von Banken. *Jahrbuch der Absatz- und Verbrauchsforschung*, Bd. 46 (2000), 3, S.220-241
- MATZLER, K.; PECHLANER, H.; KOHL, M. (2000). Formulierung von Servicestandards für touristische Dienstleistungen und Überprüfung durch den Einsatz von Mystery Guests. In: *Tourismus Journal* 4 (2), S. 157-176
- MAY, M. (2007). Urlaubsberater oder Reiseverkäufer. In: *Planung & Analyse* 34 (1), S. 42-44
- MAYER, D. (2007). Was nun, Kunde? Elektronisch veröffentlicht unter der URL http://www.e-benchmarking.de/Was_nun/body_was_nun.htm (Stand 12.7.2007)
- MCDONALD'S (2007). Jahresbericht 2006. Elektronisch veröffentlicht unter der URL http://www.mcdonalds.de/presse/basis/publikationen/mcd_jahresbericht_2006.pdf (Stand: 8.9.2007)
- MEFFERT, H., BRUHN, M. (2006). *Dienstleistungsmarketing: Grundlagen – Konzepte – Methoden*. Wiesbaden: Gabler
- MELCHERS, K.G.; KLEINMANN, M. (2007). Beurteilungsakkuratheit und Beurteilertraining. In: Schuler, H.; Sonntag, K. (Hrsg.). *Handbuch der Arbeits- und Organisationspsychologie*. Göttingen: Hogrefe, S. 561-566
- MEYER, J.A.; HOFFMANN, F. (1997). *Erfolgsfaktoren in der Gastronomie*. Berlin: Berliner Wissenschafts-Verlag
- MICHEL, J. PRYKOP, C. (2006). Wertvolle Client Insights durch Testkäufe. In: *Planung & Analyse* 33 (3), S. 52-54

- MICHELSON, M.L. (2005). Taking the Mystery out of Mystery Shopping. Elektronisch veröffentlicht unter der URL <http://www.mspa-eu.org/about/MysteryShopping1.ppt> (Stand: 18.12.2006)
- MORRALL, K. (1994). Mystery Shopping. In: Bank Marketing 26 (2), S. 13-17
- MORRALL, K. (1995). Building the Bank on Total Quality Management. In: Bank Marketing 27 (3), S. 64-68
- MORRISON, L.J.; COLMAN, A.M.; PRESTON C. C. (1997). Mystery Customer Research: Cognitive Processes Affecting Accuracy. In: Journal of the Market Research Society 39 (2), S. 349-361
- MÜLLER, F.; NEUHAUS, M.; LÖHNERT, R. (2006). Mystery Shopping – ein neuer Weg in der Revisionsarbeit. In: Betriebswirtschaftliche Blätter 55 (1), S. 24-27
- MYSTERY SHOPPING PROVIDERS ASSOCIATION (2006). Pressemitteilung vom 19.05.2006 zum Marktvolumen in Europa.
- NERDINGER, F.W. (1994). Zur Psychologie der Dienstleistung: Theoretische und empirische Studien zu einem wirtschaftspsychologischen Forschungsgebiet. Stuttgart: Schäffer-Poeschel
- NEWHOUSE, I. (2004). Mystery Shopping Made Simple. New York: McGraw-Hill
- NIENHÜSER, W. ; JANS, M. (2004): Grundbegriffe und Grundideen der Transaktionskostentheorie. Working Paper der Universität Essen
- OLAVARRIA, M. (1999). Erfolgreiche Marketingstrategien für Business to Consumer-Dienstleistungen. Dissertation: Berlin
- OLIVER, R.L. (1980): A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. In: Journal of Marketing Research 17 (4), S. 460-469
- PAERSCH, E. (2006). Kundenbindung – In geheimer Mission. In: Der Handel, Nr. 07-08, S. 28
- PARASURAMAN, A.; ZEITHAML, V.A.; BERRY, L. (1985): A Conceptual Model of Service Quality and its Implications for Future Research. In: Journal of Marketing 49 (4), S. 41-50
- PARASURAMAN, A.; ZEITHAML, V.A.; BERRY, L.L. (1988). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. In: Journal of Retailing 64 (1), S. 12-40
- PERGANDÉ, N., STÜCKEN, M. (2004). Mystery Research: Wegweiser für das kundenorientierte Unternehmen. In: Spalink, H. (Hrsg.): Kundenparadies Deutschland. Aktuelle Spitzenleistungen und Konzepte für die Zukunft. Berlin: Springer, S. 203-224
- PICOT, A. (1991). Ein neuer Ansatz zur Gestaltung der Leistungstiefe. In: Zeitschrift für betriebswirtschaftliche Forschung 43 (4), S. 336-357
- PLATZEK, T. (1997): Mystery Shopping – „Verdeckte Ermittler“ im Kampf um mehr Kundenorientierung. In: Wirtschaftswissenschaftliches Studium 26 (7), S.364-366
- POYNTER, J. (2002). Mystery Shopping: Get Paid to Shop: Denver: Leromi Publishing
- PULFORD, S. (2005). Method and System for Evaluating Quality Services. United States Patent No. US 6952679 B1, genehmigt 4. Oktober 2005
- RAFAELLI, A ; SUTTON, R.I. (1990). Busy Stores and Demanding Customers: How Do They Affect the Display of Positive Emotion? In: Academy of Management Journal 33 (3), S. 623-637
- REICHARDT, H.; REICHARDT, A. (2000). Statistische Methodenlehre für Wirtschaftswissenschaftler. Wiesbaden: Gabler

- RINGLE, C.M. (2004a). Messung von Kausalmodellen: Ein Methodenvergleich. Arbeitspapier Nr. 14 der Universität Hamburg, Institut für Industriebetriebslehre und Organisation, Arbeitsbereich Industrielles Management
- RINGLE, C.M. (2004b). Gütemaße für den Partial Least Squares-Ansatz zur Bestimmung von Kausalmodellen. Arbeitspapier Nr. 16 der Universität Hamburg, Institut für Industriebetriebslehre und Organisation, Arbeitsbereich Industrielles Management
- RINGLE, C.M.; BOYSEN, N., WENDE, S.; WILL, A. (2006). Messung von Kausalmodellen mit dem Partial-Least-Squares-Verfahren. In: *Das Wirtschaftsstudium* 35 (1), S. 81-88
- RINGLE, C.M.; SPREEN, F. (2007). Beurteilung der Ergebnisse von PLS-Pfadanalysen. In: *Das Wirtschaftsstudium* 36 (2), S. 211-216
- RINGLE, C.M.; WENDE, S.; WILL, A. (2005). SmartPLS 2.0. <http://www.smartpls.de>
- ROBERTS, B.; CAMPBELL, R.C. (2007). Being New-Customer Friendly: Determinants of Service Perceptions in Retail Banking. In: *International Journal of Bank Marketing* 25 (1), S. 56-67
- ROSENBERGER, G. (2000). Messung der Dienstleistungsqualität durch die Stiftung Warentest. In: Bruhn, M.; Strauss, B. (Hrsg.). *Dienstleistungsqualität. Konzepte – Methoden – Erfahrungen*. Wiesbaden: Gabler, S. 341-359
- RÜGER, B. (2002). Test- und Schätztheorie, Bd. 2, Statistische Tests. München: Oldenbourg
- RUST, R. T.; ZAHORIK, A. J. (1993). Customer Satisfaction, Customer Retention and Market Share. In: *Journal of Retailing* 69 (2), S. 145-156
- RUST, R. T.; ZAHORIK, A. J.; KEININGHAM, T.L. (1995). Return on Quality (ROQ): Making Service Quality Financially Accountable. In: *Journal of Marketing* 59 (4), S. 58-70
- SALCHER, E. F.; ESSER, H. (1987). Anonyme Testkäufe am Schalter – eine Methode der Erfolgskontrolle im Vertrieb. In: *Bank und Markt* 16 (11), S. 17-21
- SANDER, I. (2000). Who Watches the Watchers. In: *Planung & Analyse* 27 (5), S.30-36
- SCHLICH, A. (2006). Ein Qualitätsinstrument kommt in Mode: Mystery Shopping bewertet Servicequalität von Banken. In: *Betriebswirtschaftliche Blätter* 55 (5), S. 274-277
- SCHMIDT, K. (2007). *Mystery Shopping – Leistungsfähigkeit eines Instruments zur Messung der Dienstleistungsqualität*. Dissertation: Marburg (Anmerkung: Der Beitrag wurde vom Autor vorab als PDF-Datei zur Verfügung gestellt. Seitenzahlen können demnach nach Drucklegung variieren.)
- SCHOTT, A. (2007). Mysteriöse Kunden. In: *Frankfurter Allgemeine Zeitung* Nr. 135 vom 14.6.2007, S. 25
- SCHRADER, U. (2006). Ignorant Advice – Customer Advisory Service for Ethical Investment Funds. In: *Business Strategy and Environment* 15 (3), S. 200-214
- SCHULER, H. (2004). Der Prozess der Urteilsbildung und die Qualität der Beurteilungen. In: Schuler, H. (Hrsg.). *Beurteilung und Förderung beruflicher Leistung*. Göttingen: Hogrefe, S. 33-60
- SCHULER, H.; HÖFT, S. (2007). Diagnose beruflicher Eignung und Leistung. In: Schuler, H. (Hrsg.). *Lehrbuch Organisationspsychologie*. Bern: Huber, S. 289-343
- SCHULTE, K.; PRINS, S. (2007). Lust und Frust in der Bankberatung. In: *Planung & Analyse* 34 (6), S. 65-68

- SCHUMECKERS, G. (1980). Schaltertests helfen weiter – Testgespräche verdeckt. In: Geldprofi 4, S. 8-9 (zitiert bei DECKERS (1999))
- SCHÜTZE, R. (1992). Kundenzufriedenheit. Wiesbaden: Gabler
- SCHWEISHEIMER, W. (1962). Wilmark erzieht zum richtigen Verkaufen. In: Textilzeitung 135: 12
- SEMEL, J. (2006). Mystery Shopping. Qualitätskontrolle durch anonyme Testkäufer. Marburg: Tectum
- SHING, M.N.K.; SPENCE, L. (2002). Investigating the Limits of Competitive Intelligence Gathering: Is Mystery Shopping Ethical? In: Business Ethics: A European Review 11 (4), S. 343-353
- SHOSTACK, G.L. (1984). Designing Services that Deliver. In: Harvard Business Review 62 (January/February), S. 133-139
- SIEFKE, A. (1998). Zufriedenheit mit Dienstleistungen: Ein phasenorientierter Ansatz zur Operationalisierung und Erklärung der Kundenzufriedenheit im Verkehrsbereich auf empirischer Basis. Frankfurt/M.: Lang
- STAUSS, B. (1997). Process-Oriented Measurement of Service Quality. Applying the Sequential Incident Technique. In: European Journal of Marketing 31 (1), S. 33-55
- STAUSS, B. (2000). „Augenblicke der Wahrheit“ in der Dienstleistungserstellung – Ihre Relevanz und ihre Messung mit Hilfe der Kontaktpunkt-Analyse. In: Bruhn, M., Stauss, B. (Hrsg.). Dienstleistungsqualität. Grundlagen – Konzepte – Methoden. Wiesbaden: Gabler, S. 321-340
- STAUSS, B.; HENTSCHEL, B. (1990). Verfahren der Problementdeckung- und analyse im Qualitätsmanagement von Dienstleistungsunternehmen. In: Jahrbuch der Absatz- und Verbrauchsforschung 36 (3), S. 232-259
- STAUSS, B.; HENTSCHEL, B. (1992). Attribute-Based Versus Incident-Based Measurement of Service Quality: Results of an Empirical Study in the German Car Service Industry. In: Kunst, P.; Lemming, J. (Hrsg.). Quality Management in Services. Assen: Von Gorcum, S. 59-78
- STAUSS, B.; SEIDEL, W. (2006). Prozessuale Zufriedenheitsermittlung und Zufriedenheitsdynamik bei Dienstleistungen. In: Homburg, C. (Hrsg.). Kundenzufriedenheit: Konzepte – Methoden – Erfahrungen. Wiesbaden: Gabler, S. 171-195
- STEAD, B.A.; ZINKHAN, G.M. (1986). Service Priority in Department Stores: The Effects of Customer Gender and Dress. In: Sex Roles 15 (11-12), S. 601-611
- STÜCKEN, M. (2003). Mystery Research oder Kundenzufriedenheitsbefragung? In: Planung & Analyse 30 (5), S. 45-50
- SULSKY, L.M.; BALZER, W.K. (1988). Meaning and Measurement of Performance Rating Accuracy: Some Methodological and Theoretical Concerns. In: Journal of Applied Psychology 73 (3), S. 497-506
- SUTTON, R.I.; RAFAELI, A. (1988). Untangling the Relationship Between Displayed Emotions and Organizational Sales. The Case of Convenience Stores. In: Academy of Management Journal 31 (2), S. 461-489.
- TEMME, D.; KREIS, H.; HILDEBRANDT, L. (2006). PLS Path Modeling – A Software Review. Discussion Paper 2006-084 der Humboldt-Universität zu Berlin, Sonderforschungsbereich 649: Ökonomisches Risiko

- TIMPE, M. (2005). Service-Sündern mit Testkäufern auf der Spur. In: Die Geschäftsidee 30 (3), S. 14-27
- TNS EMNID (2003). Semiometrie – Positionierungsanalyse Schnellrestaurants. Elektronisch veröffentlicht unter der URL http://www.tns-emnid.com/pdf/top-thema-semiometrie/Archiv2003/TopThema_04-03.pdf (Stand 17.08.2007)
- TRIBE, J. (2002). Mystery Shopping: Theory and Practice. In: Laws, E. (Hrsg.). Tourism Marketing: Quality and Service Management Perspectives. London: Sontinum, S. 75-86
- TROMMSDORFF, V. (2004). Konsumentenverhalten. Stuttgart: Kohlhammer
- TURNER, H. (2007). Mystery Shopping. In: Hamersveld, M. van ; Bont, C. de (Hrsg.) Handbook of Market Research. Chichester: Wiley, S. 333-346
- VILLWOCK, B. (2004). Neue Testberatungs-Tools im Automobil- und Finanzdienstleistungssektor. In: Planung & Analyse 31 (4), S. 54-57
- VOSS, C.A.; ALEDA, V.; ROTH, E.D.; ROSENZWEIG, K.B.; RICHARD B.C. (2004). A Tale of Two Countries' Conservatism, Service Quality, and Feedback on Customer Satisfaction. In: Journal of Service Research 6 (3), S. 212-230.
- WARMUTH, D.P.; WEINHOLD, M. (2005). Damit der Kunde nicht mehr stört. Kundenorientierte Führung durch Mystery Shopping. Wien: Linde
- WIESEKE, J.; SCHMIDT, K.; LINGENFELDER, M. (2006). Was leistet Mystery Shopping? In: Absatzwirtschaft 49 (6), S. 42-44
- WILLIAMSON, O.E. (1975). Markets and Hierarchies. Analysis and Antitrust Implications. NewYork/London
- WILLIAMSON, O.E. (1985). The Economic Institutions of Capitalism. Firms, Markets, Relational Contracting. New York u.a.
- WILSON, A.M. (1998a). The Use of Mystery Shopping in the Measurement of Service Delivery. In: Service Industries Journal 18 (3), S. 148-163
- WILSON, A.M. (1998b) The Role of Mystery Shopping in the Measurement of Service Performance. In: Managing Service Quality 8 (6), S. 414-420
- WILSON, A. M. (2001). Using Deception to Measure Service Performance. In: Psychology & Marketing 18 (7), S. 721-734
- WILSON, A. M.; GUTMANN, J. (1998). Public Transport: The Role of Mystery Shopping in Investment Decisions. In: Journal of the Market Research Society 40 (4), S. 285-293
- WINTER, S. (2005). Kundenbefragung. In: Jöns, I.; Bungard, W. (Hrsg.). Feedbackinstrumente im Unternehmen : Grundlagen, Gestaltungshinweise, Erfahrungsberichte. Wiesbaden: Gabler, S. 177-194
- WIRTZ, M.; CASPAR, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität. Göttingen: Hogrefe
- ZANGER, C.; SISTENICH, F. (1996). Qualitative Marktforschung. Struktur, Methoden und Anwendungsraum des hermeneutischen Ansatzes. In: Wirtschaftswissenschaftliches Studium 25 (7), S. 351-354
- ZENZ, S.; SCHWEIGER, G. (2003). Mystery Shopping – Mysterium oder Hilfsmittel zur Verbesserung der Unternehmensrentabilität. In: Transfer – Werbeforschung & Praxis, 48 (2), S. 2-6

ZINKHAN, G.A.; STOIADIN, L.F. (1984). Impact of Sex Role Stereotypes on Service Priority in Department Stores. In: Journal of Applied Psychology 69 (4), S. 691-693

ZÖFEL, P. (2003). Statistik für Psychologen. München: Pearson