

**Entwicklung eines Wortschatztests
für gehörlose Kinder im Grundschulalter
in Gebärden-, Schrift- und Lautsprache**

Inauguraldissertation
zur Erlangung des akademischen Grades
des Doktors der Philosophie
am Fachbereich Erziehungswissenschaft
der Universität Hamburg

vorgelegt von

Sibylle Bizer und Anne-Katrin Karl

Hamburg, 2002

Gutachter:

Prof. Dr. Günther, I 5

Prof. Dr. Wudtke, I 3

Prof. Dr. Prillwitz, FB 07

Prof. Dr. Burisch, FB 16

Prof. Dr. Dehn, I 7

Letzte mündliche Prüfung
am 20. November 2002

Erklärung zur Arbeitsteilung im Rahmen einer Gemeinschaftsarbeit

Die vorliegende Arbeit ist gemeinschaftlich von Frau Sibylle Bizer und Frau Anne-Katrin Karl erstellt worden. Im Folgenden wird eine Erklärung darüber abgegeben, welche Anteile von wem erarbeitet und niedergeschrieben wurden.

Vorwegnehmend sei gesagt, dass beide Autorinnen grundsätzlich an sämtlichen inhaltlichen sowie formalen Elementen der Arbeit beteiligt waren. D. h., dass alle zu treffenden Entscheidungen zur Entwicklung des Tests und der Elemente qualitativer Bewertungsversuche, zur Durchführung der Untersuchung sowie zur Erstellung des Konzeptes der Dissertationsschrift (unabhängig von der Aufteilung der Niederschrift) gemeinsam diskutiert und erarbeitet wurden. Dabei wurden zwei Schwerpunktbereiche festgelegt, für die jeweils eine der Autorinnen die tragende Verantwortung übernimmt. Dies ist für Frau Bizer der Bereich der Anwendung der klassischen Testtheorie auf das vorliegende Instrument, die Durchführung entsprechender statistischer Berechnungen sowie die Auswertung der Ergebnisse im Sinne der Erstellung übersichtlicher Darstellungen in Form von Tabellen und in Texten. Frau Karl bearbeitete schwerpunktmäßig die theoretische Grundlegung im Sinne einer Einordnung des Projekts in den bestehenden Forschungsstand und seiner Herleitung daraus, die Erarbeitung und Darstellung seiner Grundelemente auf diesem Hintergrund sowie die Auswertung der Ergebnisse im Sinne einer (vor allem pädagogischen) Bewertung. (Ergänzend sei dazu angemerkt, dass sich diese Aufteilung mit den jeweiligen Hintergründen der Autorinnen deckt. So widmete sich Frau Bizer im Rahmen des sonderpädagogischen Studiums einem naturwissenschaftlichen Bereich (Unterrichtsfach Mathematik) und Frau Karl einem geisteswissenschaftlichen (Unterrichtsfach ev. Theologie)).

Die Niederschrift wurde gemäß der Bearbeitungsschwerpunkte wie folgt aufgeteilt:

Gliederungs- punkt	Überschrift	Nieder- schrift	Anmerkung
Teil A	Ausgangslage und Ziel der Untersuchung		
A 1.	Problembereich der Untersuchung	Frau Karl	
A 2.	Forschungsstand	Frau Karl	
A 3.	Zielsetzung und Ziel	Frau Karl	
Teil B	Testaufbau und Durch- führung der Untersuchung		
B 1.	Grundkonzeption des Tests	Frau Karl	
B 2.	Pilotstudie	Frau Karl	Tabellen: Frau Bizer
B 3.	Hauptuntersuchung	Frau Bizer	B 3.6: Frau Karl
Teil C	Auswertung der Untersuchungsergebnisse		
C 1.	Überprüfung pädagogischer Fragestellungen	Frau Bizer	Formulierung der Begründungszusammenhänge für die einzelnen Annahmen gemeinsam
C 2.	Pädagogische Folgerungen	Frau Karl	
C 3.	Resümee klassischer testtheoretischer Aspekte	Frau Karl	
C 4.	Resümee der qualitativen Auswertung	Frau Karl	
C 5.	Ausblick	Frau Karl	
Teil D	Anwenderhandbuch		
D 1.	Beschreibung des Tests	Frau Karl	Zusammenstellung und Anpassung der Formulierung für das Handbuch: Frau Bizer
D 2.	Normierung	Frau Bizer	
D 3.	Gütekriterien	Frau Bizer	
D 4.	Durchführungsanleitung	Frau Karl	Zusammenstellung und Anpassung der Formulierung für das Handbuch: Frau Bizer
D 5.	Auswertungsanleitung	Frau Bizer	
D 6.	Literaturverzeichnis des Anwenderhandbuchs	Frau Karl	
Literatur- verzeichnis		Frau Karl	
Anhang 1		Frau Bizer	

Sibylle Bizer

Anne-Katrin Karl

Verzeichnis der für diese Arbeit eingeführten Abkürzungen

GS	Gebärdensprache
SS	Schriftsprache
LS	Lautsprache
GSN	Subtestteil Gebärdensprache Nomen
GSV	Subtestteil Gebärdensprache Verben
GSA	Subtestteil Gebärdensprache Adjektive
SSN	Subtestteil Schriftsprache Nomen
SSV	Subtestteil Schriftsprache Verben
SSA	Subtestteil Schriftsprache Adjektive
LSN	Subtestteil Lautsprache Nomen
LSV	Subtestteil Lautsprache Verben
LSA	Subtestteil Lautsprache Adjektive
I	Distraktor mit Ähnlichkeit im Inhalt
G	Distraktor mit Ähnlichkeit in der Gebärde
M	Distraktor mit Ähnlichkeit im Mundbild
Z	Distraktor mit Ähnlichkeit in der Zeichnung
HRP	Distraktor höchster relativer Plausibilität

Tabellenverzeichnis

laufende Nummer	Bezeichnung	Kapitel, Seite
Tab. 1-A 2.4.1	Itemkorpus des FTF-W	A 2.4.1, S. 18
Tab. 2-A 2.4.1	Itemkorpus des WSS 1	A 2.4.1, S. 19
Tab. 3-A 2.4.1	Itemkorpus des PPVT	A 2.4.1, S. 20
Tab. 4-B 1.3	Wortliste Günther (1989 b)	Anh. 1: B 1.3a, S. 1
Tab. 5-B 1.3	Itemkorpus Pilotversion	B 1.3, S. 44
Tab. 6a/b/c-B 1.3	Distraktorenähnlichkeiten Pilotversion N/V/A	Anh. 1: B 1.3b, S. 12
Tab. 7a/b-B 2.3.1	Klasseninterner Vergleich der Leistungen Gehörloser Klasse 3/4	Anh. 1: B 2.3.1, S. 16
Tab. 8a/b-B 2.3.2	Klasseninterner Vergleich der Zeiten Gehörloser Klasse 3/4	Anh. 1: B 2.3.2, S. 18
Tab. 9a/b-B 2.3.3	Gesamtergebnisse Gehörloser Klasse 3/4	Anh. 1: B 2.3.3, S. 20
Tab. 10a/b/c-B 2.3.4	Ergebnisse/Fehleranalyse Hörender	Anh. 1: B 2.3.4, S. 21
Tab. 11-B 2.4.1	Fehlerverteilung Gehörloser	Anh. 1: B 2.4.1, S. 26
Tab. 12a-k-B 2.4.2	Rating Testpersonen 1-11, Klasse 3/4	Anh. 1: B 2.4.2, S. 27
Tab. 13a/b/c-B 2.4.5	Fehleranalyse Gehörloser N/V/A	Anh. 1: B 2.4.5, S. 48
Tab. 14a/b/c-B 2.4.7	Itemrangliste N/V/A	Anh. 1: B 2.4.7, S. 51
Tab. 15a/b/c-B 2.7	Auswahl und Distraktorenähnlichkeit zusätzlicher und schwerer Items N/V/A	B 2.7, S. 91
Tab. 16a/b/c-B 2.8	Itemkorpus N/V/A Version I	B 2.8, S. 92
Tab. 17-B 3.1	Anzahl der Testpersonen	B 3.1, S. 96
Tab. 18-B 3.1	Geschlecht der Testpersonen	B 3.1, S. 96
Tab. 19-B 3.1	Unterrichtskonzepte der Sprachvermittlung	B 3.1, S. 96
Tab. 20a/b/c-B 3.1	Alter der Testpersonen beim Subtest GS/SS/LS	B 3.1, S. 97
Tab. 21-B 3.2.1	Testpunktwerte	Anh. 1: B 3.2.1, S. 53
Tab. 22a/b/c-B 3.2.2	Distraktorenverteilung N/V/A	Anh. 1: B 3.2.2, S. 56
Tab. 23-B 3.3	Punkteverteilung in Kategorien	Anh. 1: B 3.3a, S. 64
Tab. 24-B 3.3	Schiefe und Exzess	B 3.3, S. 100
Tab. 25a/b/c-B 3.4.2	Schwierigkeit und Trennschärfe N/V/A Version I	B 3.4.2, S. 104
Tab. 26-B 3.4.3	Interkorrelationen	Anh. 1: B 3.4.3, S. 78
Tab. 27a-i-B 3.4.4	Reliabilitätswerte der Testaufgaben GSN - LSA	Anh. 1: B 3.4.4, S. 79
Tab. 28a/b/c-B 3.4.4	Gesamtreliabilität N/V/A	B 3.4.4, S. 109
Tab. 29a-e-B 3.4.5	Distraktorentrennschärfewerte	Anh. 1: B 3.4.5a, S. 85
Tab. 30a-e-B 3.4.5	Korrelation Distraktor–Raven	Anh. 1: B 3.4.5b, S. 90

laufende Nummer	Bezeichnung	Kapitel, Seite
Tab. 31-B 3.5.2	Eliminierte Aufgaben	B 3.5.2, S. 114
Tab. 32a/b/c-B 3.5.2	Gründe für Eliminierung von N/V/A	B 3.5.2, S. 115
Tab. 33a/b/c-B 3.5.3	Itemkorpus N/V/A Version II	B 3.5.3, S. 120
Tab. 34-B 3.5.	Testergebnisse Version I-6	Anh. 1: B 3.5.3, S. 95
Tab. 35a/b/c-B 3.5.4	Neue Platzverteilung aufgrund der Schwierigkeit N/V/A	Anh. 1: B 3.5.4a, S. 98
Tab. 36-B 3.5.4	Mittelwerte der Schwierigkeitsrangfolge für Testheft I	Anh. 1: B 3.5.4b, S. 101
Tab. 37a/b-B 3.5.4	Gliederung des Itemkorpus in Testheft I/II	B 3.5.4, S. 123
Tab. 38-B 3.5.4	Fehlerverteilung	B 3.5.4, S. 125
Tab. 39a/b/c-B 3.5.5	Distraktorenverteilung N/V/A Version II	B 3.5.5, S. 126
Tab. 40a/b/c-B 3.5.6	Schwierigkeit und Trennschärfe Version II N/V/A	Anh. 1: B 3.5.6a, S. 111
Tab. 41-B 3.5.6	Mittlere Trennschärfe im Vergleich Version I und II	B 3.5.6, S. 130
Tab. 42a-j-B 3.5.7	Modifiziertes Testprotokoll Deckblatt, GSN - LSA	Anh. 1: B 3.5.7, S. 123
Tab. 43a/b/c-B 3.6	Ikonizität der N/V/A	B 3.6, S. 133
Tab. 44a/b/c-B 3.6	Ausschließlich in GS richtig beantwortete N/V/A	B 3.6, S. 136
Tab. 45-B 3.6	Ausschließlich in GS richtig beantwortete Items mit und ohne Ikonizität	B 3.6, S. 139
Tab. 46-B 3.7.2	Aufteilung der Items in Testhälften A und B	Anh. 1: B 3.7.2a, S. 133
Tab. 47-B 3.7.2	Gesamtwerte der Testhälften A und B nach Halbierung über statistische Kennwerte	Anh. 1: B 3.7.2b, S. 134
Tab. 48-B 3.7.2	Reliabilitätskoeffizienten der Split-half-Methode nach statistischen Kennwerten	Anh. 1: B 3.7.2c, S. 136
Tab. 49-B 3.7.2	Gesamtwerte der Testhälften A und B nach Halbierung über die odd-even-Methode	B 3.7.2, S. 143
Tab. 50-B 3.7.2	Reliabilitätskoeffizienten	B 3.7.2, S. 145
Tab. 51-B 3.7.3	Korrelationen Wortschatztest – Faktoren	B 3.7.3, S. 149
Tab. 52-B 3.7.3	Korrelation Wortschatztest – Raven	B 3.7.3, S. 150
Tab. 53-B 3.7.3	Lehrer-Rating der passiven Wortschatzkenntnis	Anh. 1: B 3.7.3, S. 137
Tab. 54-B 3.7.3	Korrelation Wortschatztest– Rating	B 3.7.3, S. 151
Tab. 55a-i-B 3.8.4	Normierung über Flächentransformation GSN - LSA	Anh. 1: B 3.8.4, S. 139
Tab. 56a-i-B 3.8.4	Normierung bei Normalverteilung GSN - LSA	B 3.8.4, S. 157

laufende Nummer	Bezeichnung	Kapitel, Seite
Tab. 57a/b/c-B 3.8.5	Normierung aufgrund modifizierter Fehlerbewertung N/V/A	B 3.8.5, S. 162
Tab. 58-B 3.9.1	Standardmessfehler der Subtestteile	B 3.9.1, S. 164
Tab. 59-B 3.9.2	Standardabweichung je zweier Subtestteile für die Beurteilung intraindividueller Unterschiede	Anh. 1: B 3.9.2a, S. 144
Tab. 60-B 3.9.2	Reliabilität je zweier Subtestteile für die Beurteilung intraindividueller Unterschiede	Anh. 1: B 3.9.2b, S. 145
Tab. 61-B 3.9.2	Kritische Differenzen je zweier Subtestteile	B 3.9.2, S. 165
Tab. 62-B 3.9.3	Standardabweichung, Reliabilität und kritische Differenz für die Beurteilung interindividueller Unterschiede	B 3.9.3, S. 166
Tab. 63a/b/c-B 3.10.1	Gesamtfehlerverteilung der besten und schlechtesten Testpersonen GS/SS/LS	B 3.10.1, S. 170
Tab. 64a/b/c-B 3.10.1	Gesamtfehlerverteilung GS/SS/LS – ausgewählte Ablenkerarten entsprechend dem quantitativen Testergebnis	B 3.10.1, S. 171
Tab. 65a-i-B 3.10.2	Distraktoren-Normierung GSN - LSA	B 3.10.2, S. 173
Tab. 66a-e-B 3.10.3	Standardabweichung je zweier Subtestteile HRP - I	Anh. 1: B 3.10.3a, S. 146
Tab. 67a-e-B 3.10.3	Reliabilität je zweier Subtestteile HRP - I	Anh. 1: B 3.10.3b, S. 148
Tab. 68a-e-B 3.10.3	Standardmessfehler und kritische Differenzen HRP - I	B 3.10.3, S. 178
Tab. 69a-e-B 3.10.3	Kritische Differenzen je zweier Subtestteile HRP – I	B 3.10.3, S. 180
Tab. 70a/b-C 1.2.1	Fehlervergleich Klasse 3/4	C 1.2.1, S. 184
Tab. 71a/b-C 1.2.2	Fehlervergleich Gesamtergebnisse Gehörlose/Hörende	C 1.2.2, S. 186
Tab. 72-C 1.2.3	Vergleich Hörende Lautsprache vs. Gehörlose Gebärdensprache	C 1.2.3, S. 186
Tab. 73a/b/c-C 1.4.1	Ergebnisse N/V/A durch drei Modalitäten	C 1.4.1, S. 193
Tab. 74a/b/c-C 1.4.2	Anteil der N/V/A, die nur in Gebärdensprache richtig beantwortet wurden	C 1.4.2, S. 195
Tab. 75-C 1.5.1	Testergebnisse der Testpersonen nach Konzepten sortiert	C 1.5.1, S. 201
Tab. 76-C 1.5.2	Klassenzugehörigkeit der besten und schlechtesten Testpersonen in allen Subtestteilen	C 1.5.2, S. 205

laufende Nummer	Bezeichnung	Kapitel, Seite
Tab. 77-C 1.6.1	Leistungsmittelwerte der Klassenstufen	C 1.6.1, S. 207
Tab. 78-C 1.6.1	Leistungsmittelwerte und signifikante Unterschiede der Konzepte	C 1.6.1, S. 207
Tab. 79-C 1.6.1	Differenzen der Konzeptmittelwerte	C 1.6.1, S. 209
Tab. 80-C 1.6.1	Leistungsmittelwerte und signifikante Unterschiede der IQ-Gruppen	C 1.6.1, S. 210
Tab. 81-C 1.6.1	Differenzen der IQ-Gruppenmittelwerte	C 1.6.1, S. 211
Tab. 82-C 1.6.1	Leistungsmittelwerte der Testpersonen mit hörenden/gehörlosen Eltern	C 1.6.1, S. 212
Tab. 83a/b-C 1.6.2	Einfluss von Faktorenkombinationen mit 4/5 Variablen	C 1.6.2, S. 214
Tab. 84-C 1.7.1	Verteilung der Testpersonen auf sechs Cluster	C 1.7.1, S. 216
Tab. 85-C 1.7.2	Teilleistungsprofile innerhalb der Modalitäten	C 1.7.2, S. 222
Tab. 86-C 1.7.2	Teilleistungsprofile innerhalb der Modalität Gebärdensprache	C 1.7.2, S. 225
Tab. 87-C 1.7.2	Teilleistungsprofile innerhalb der Modalität Schriftsprache	C 1.7.2, S. 228
Tab. 88-C 1.7.2	Teilleistungsprofile innerhalb der Modalität Lautsprache	C 1.7.2, S. 231
Tab. 89-C 1.7.3	Teilleistungsprofile innerhalb der Wortarten	C 1.7.3, S. 234
Tab. 90-C 1.8.1	Durchschnittliche Wahlhäufigkeit der Distraktoren	C 1.8.1, S. 238
Tab. 91a/b/c-C 1.8.2	Clusteranalyse bzgl. der Distraktorenwahl GSN-GSA	C 1.8.2, S. 238
Tab. 92a/b/c-C 1.8.3	Clusteranalyse bzgl. der Distraktorenwahl SSN-SSA	C 1.8.3, S. 246
Tab. 93a/b/c-C 1.8.4	Clusteranalyse bzgl. der Distraktorenwahl LSN-LSA	C 1.8.4, S. 254
Tab. 94-C 1.8.5	Clusteranalyse bzgl. der Distraktorenwahl Gesamttest	C 1.8.5, S. 262
Tab. 95a/b/c-D 1.4	Itemkorpus N/V/A	D 1.4, S. 277
Tab. 96a/b-D 1.4	Gliederung des Itemkorpus in Testheft I/II	D 1.4, S. 279
Tab. 97a/b/c-D 2.2	Schwierigkeit und Trennschärfe N/V/A	D 2.2, S. 282
Tab. 98-D 3.2.1	Reliabilitätskoeffizienten	D 3.2.1, S. 288
Tab. 99-D 3.2.2	Standardmessfehler der Subtestteile	D 3.2.2, S. 288
Tab. 100-D 3.3.2	Korrelationen Wortschatztest – Faktoren	D 3.2.2, S. 291

laufende Nummer	Bezeichnung	Kapitel, Seite
Tab. 101-D 3.3.3	Korrelation Wortschatztest – Rating	D 3.2.3, S. 293
Tab. 102-D 3.3.4	Korrelation Wortschatztest – Raven	D 3.2.4, S. 294
Tab. 103a/b/c-D 5.1.1	Normierung Subtestteile GSN - LSA	D 5.1.1, S. 301
Tab. 104a/b/c-D 5.1.1	Normierung Gesamtwerte Subtests GS/SS/LS	D 5.1.1, S. 304
Tab. 105a/b/c-D 5.1.2	Normierung aufgrund modifizierter Fehlerbewertung	D 5.1.2, S. 307
Tab. 106a/b/c-D 5.2	Distraktorenverteilung N/V/A	D 5.2, S. 310
Tab. 107a-i-D 5.2	Distraktoren-Normierung GSN - LSA	D 5.2, S. 315
Tab. 108-D 5.2	Standardmessfehler im Distraktorenprofil	D 5.2, S. 319

Abbildungsverzeichnis

laufende Nummer	Bezeichnung	Kapitel, Seite
Abb. 1-A 2.3	Das semiotische Dreieck	A 2.3, S. 15
Abb. 2-A 2.4.1	Item Nr. 23 des FTF-W	A 2.4.1, S. 18
Abb. 3-A 2.4.1	Item Nr. 23 des WSS 1	A 2.4.1, S. 19
Abb. 4-A 2.4.1	Item Nr. 23 des PPVT	A 2.4.1, S. 20
Abb. 5-B 1.5	Aufbauschema Testheft; Vierfelder-Wahltafel	B 1.5, S. 51
Abb. 6-B 1.5	Schema des Ergebnisprotokolls Pilotversion	B 1.5, S. 52
Abb. 7-B 2.4.2	Rating zur Arbeitshaltung	B 2.4.2, S. 65
Abb. 8a/b-B 2.4.2	Kategorisierung Ergebnisse Gehörloser Klasse 3/4	B 2.4.2, S. 67
Abb. 9a-i-B 2.4.3	Schwierigkeit–Trennschärfe Gehörlose Pilotstudie GSN-LSA	Anh. 1: B 2.4.3a, S. 33
Abb. 10a-f-B 2.4.3	Schwierigkeit–Trennschärfe Hörende Pilotstudie SSN-LSA	Anh. 1: B 2.4.3b, S. 42
Abb. 11a-i-B 3.3	Kategorisierte Verteilungskurven GSN - LSA	Anh. 1: B 3.3b, S. 65
Abb. 12a-i-B 3.4.2	Schwierigkeit–Trennschärfe Version I GSN - LSA	Anh. 1: B 3.4.2, S. 69
Abb. 13-B 3.5.4	Gebärdenfotos und Ausführungsanweisungen	Anh. 1: B 3.5.4, S. 102
Abb. 14a-i-B 3.5.6	Schwierigkeit–Trennschärfe Version II GSN - LSA	Anh. 1: B 3.5.6b, S. 114
Abb. 15-B 3.7.3	Ratingskala	B 3.7.3, S. 151
Abb. 16-B 3.10	Schaubild zur Informationsgewinnung	B 3.10, S. 167
Abb. 17a-f-C 1.7.1	Graphische Darstellung Cluster 1-6	C 1.7.1, S. 216
Abb. 18a/b-C 1.7.2	Profilverläufe innerhalb der Modalitäten	C 1.7.2, S. 223
Abb. 19a/b-C 1.7.2	Profilverläufe innerhalb der Modalität Geb.sprache	C 1.7.2, S. 226
Abb. 20a/b-C 1.7.2	Profilverläufe innerhalb der Modalität Schriftsprache	C 1.7.2, S. 229
Abb. 21a/b-C 1.7.2	Profilverläufe innerhalb der Modalität Lautsprache	C 1.7.2, S. 232
Abb. 22a/b-C 1.7.3	Profilverläufe innerhalb der Wortarten	C 1.7.3, S. 235
Abb. 23-C 1.7.4	Profilverläufe Konzepte	Anh. 1: C 1.7.4, S. 150
Abb. 24-C 1.7.5	Profilverläufe Klassen	Anh. 1: C 1.7.5, S. 151
Abb. 25-C 1.8.1	Fehleranzahl	Anh. 1: C 1.8.1, S. 152

INHALT

A. AUSGANGSLAGE UND ZIEL DER UNTERSUCHUNG.....	4
1. PROBLEMBEREICH DER UNTERSUCHUNG	4
2. FORSCHUNGSSTAND.....	6
2.1 <i>Bilingualer Schulversuch Hamburg</i>	6
2.2 <i>Gehörlosenpädagogische Methodendiskussion</i>	9
2.3 <i>Begriffsbildung und Wortschatz</i>	13
2.4 <i>Diagnostik lexikalisch-semantischer Fähigkeiten</i>	17
2.4.1 Untersuchung der kommunikativ-sprachlichen Entwicklung.....	17
2.4.2 Testinventar.....	22
2.4.3 Folgerungen	27
2.5 <i>Testdiagnostik</i>	28
3. ZIELSETZUNG UND ZIELENTWICKLUNG	32
B. TESTAUFBAU UND DURCHFÜHRUNG DER UNTERSUCHUNG	34
1. GRUNDKONZEPTION DES TESTS.....	34
1.1 <i>Auswahl der Sprachmodalitäten</i>	34
1.2 <i>Auswahl und Gewichtung der Wortarten</i>	38
1.3 <i>Auswahl von Aufgabentyp, Testitems und Distraktoren</i>	42
1.4 <i>Zeichnerische Gestaltung</i>	47
1.5 <i>Testheft und Ergebnisprotokoll</i>	50
1.6 <i>Testsituation und -durchführung</i>	54
2. PILOTSTUDIE	59
2.1 <i>Stichprobe und Durchführung der Untersuchung</i>	59
2.2 <i>Testdurchführung mit hörenden Testpersonen</i>	61
2.3 <i>Ergebnisse der Pilotstudie</i>	62
2.3.1 <i>Klasseninterner Vergleich der Leistungen Gehörloser</i>	62
2.3.2 <i>Klasseninterner Vergleich der Zeiten Gehörloser</i>	63
2.3.3 <i>Gesamtergebnisse Gehörloser</i>	63
2.3.4 <i>Ergebnisse Hörender</i>	64
2.4 <i>Datenaufbereitung unter verschiedenen Fragestellungen der Testentwicklung</i>	64
2.4.1 <i>Fehlerverteilung Gehörlose</i>	64
2.4.2 <i>Rating zur Arbeitshaltung</i>	65
2.4.3 <i>Schwierigkeit und Trennschärfe</i>	68
2.4.4 <i>Distraktoren-Ähnlichkeiten</i>	69
2.4.5 <i>Fehleranalyse Gehörlose</i>	69
2.4.6 <i>Fehleranalyse Hörende</i>	70
2.4.7 <i>Itemrangliste</i>	70
2.5 <i>Vorgehen bei der Modifikation der Pilotversion</i>	71
2.6 <i>Analyse und Modifikation zur Testversion I</i>	73
2.6.1 <i>Diskussion der Testaufgaben/Nomen</i>	73
2.6.2 <i>Diskussion der Testaufgaben/Verben</i>	81
2.6.3 <i>Diskussion der Testaufgaben/Adjektive</i>	85
2.7 <i>Ergänzung zusätzlicher und schwererer Items</i>	90
2.8 <i>Testversion I: Itemkorpus</i>	92
3. HAUPTUNTERSUCHUNG	95
3.1 <i>Stichprobe</i>	95
3.2 <i>Untersuchungsergebnisse</i>	98
3.2.1 <i>Darstellung der Testpunktwerte</i>	98
3.2.2 <i>Darstellung der Distraktorenwahl</i>	99
3.3 <i>Überprüfung auf Normalverteilung</i>	99
3.4 <i>Aufgabenanalyse</i>	100
3.4.1 <i>Qualitätskriterien der Testaufgaben</i>	100
3.4.2 <i>Aufgabentrennschärfe und –schwierigkeit</i>	103
3.4.3 <i>Aufgaben-Interkorrelationen</i>	107
3.4.4 <i>Aufgabenstreuung und –reliabilität</i>	108
3.4.5 <i>Distraktorenanalyse</i>	111
3.4.6 <i>Sonstige Kriterien</i>	113
3.5 <i>Konsequenzen der Aufgabenanalyse</i>	113
3.5.1 <i>Aufgabenselektion</i>	114
3.5.2 <i>Aufgabenrevision</i>	116
3.5.3 <i>Modifizierter Itemkorpus</i>	120

3.5.4 Modifizierte Reihenfolge	122
3.5.5 Modifizierte Distraktorenverteilung	125
3.5.6 Neue Trennschärfewerte.....	129
3.5.7 Modifiziertes Testprotokoll.....	130
3.6 Exkurs: Ikonizität von Gebärdensprache.....	131
3.7 Testgütekriterien	140
3.7.1 Objektivität	140
3.7.2 Reliabilität.....	140
3.7.3 Validität.....	145
3.8 Normierung	152
3.8.1 Skalenniveaus	153
3.8.2 Prozentrangskalen.....	154
3.8.3 Lineare Standardwerte.....	155
3.8.4 Normierungstabellen	156
3.8.5 Normierung aufgrund modifizierter Fehlerbewertung.....	162
3.9 Möglichkeiten zur Beurteilung von intra- und interindividuellen Leistungsunterschieden.....	164
3.9.1 Standardmessfehler.....	164
3.9.2 Kritische Differenz für die Beurteilung intraindividueller Unterschiede.....	165
3.9.3 Kritische Differenz für die Beurteilung interindividueller Unterschiede.....	166
3.10 Distraktorenprofil	166
3.10.1 Überprüfung der qualitativen Distraktorenabstufung	168
3.10.2 Distraktoren-Normierungstabellen	172
3.10.3 Standardmessfehler und kritische Differenzen im Distraktorenprofil	178
C. AUSWERTUNG DER UNTERSUCHUNGSERGEBNISSE.....	182
1. ÜBERPRÜFUNG PÄDAGOGISCHER FRAGESTELLUNGEN.....	182
1.1 Schwierigkeitsgefälle der Modalitäten im Rahmen der Pilotstudie.....	182
1.2 Vergleich der Kompetenzen mit neuer Fehlerbewertung im Rahmen der Pilotstudie.....	183
1.2.1 Überprüfung I.....	184
1.2.2 Überprüfung II.....	186
1.2.3 Überprüfung III.....	186
1.3 Schwierigkeitsgefälle der Modalitäten im Rahmen der Hauptuntersuchung.....	187
1.3.1 Überprüfung I.....	188
1.3.2 Überprüfung II.....	188
1.3.3 Überprüfung III.....	190
1.4 Vergleich der Kompetenzen mit modifizierter Fehlerbewertung.....	193
1.4.1 Überprüfung I.....	193
1.4.2 Überprüfung II.....	195
1.5 Bilinguale Schüler an der Leistungsspitze.....	197
1.5.1 Überprüfung I.....	197
1.5.2 Überprüfung II.....	202
1.6 Korrelationen.....	206
1.6.1 Überprüfung I.....	206
1.6.2 Überprüfung II.....	213
1.7 Typische Profilverläufe.....	215
1.7.1 Überprüfung I.....	215
1.7.2 Überprüfung II.....	221
1.7.3 Überprüfung III.....	233
1.7.4 Überprüfung IV.....	236
1.7.5 Überprüfung V.....	237
1.8 Häufig vorkommende typische Profilverläufe bzgl. der Distraktorenwahl.....	237
1.8.1 Überprüfung I.....	237
1.8.2 Überprüfung II.....	238
1.8.3 Überprüfung III.....	246
1.8.4 Überprüfung IV.....	254
1.8.5 Überprüfung V.....	261
2. PÄDAGOGISCHE FOLGERUNGEN	264
3. RESÜMEE KLASSISCHER TESTTHEORETISCHER ASPEKTE.....	268
4. RESÜMEE DER QUALITATIVEN AUSWERTUNG.....	269
5. AUSBLICK.....	272
D. ANWENDERHANDBUCH	274
1. BESCHREIBUNG DES TESTS	274
1.1 Vorgeschichte und Aufgabe des Tests.....	274
1.2 Anwendungsbereich.....	275

1.3 Eigenschaften des Tests	275
1.4 Testmaterial.....	277
1.5 Zeichnerische Gestaltung	280
2. NORMIERUNG	281
2.1 Stichprobe.....	281
2.2 Aufgabenanalyse	282
2.3 Prozentrangskalen und lineare Standardwerte	284
3. GÜTEKRITERIEN	286
3.1 Objektivität.....	286
3.2 Reliabilität.....	286
3.2.1 Innere Konsistenz und Split-half.....	286
3.2.2 Standardmessfehler.....	288
3.3 Validität	289
3.3.1 Inhaltliche Validität	289
3.3.2 Konstruktvalidität	290
3.3.3 Übereinstimmungsvalidität (Ratingverfahren).....	292
3.3.4 Übereinstimmungsvalidität (Raven)	294
4. DURCHFÜHRUNGSANLEITUNG	295
4.1 Allgemeine Instruktionen	295
4.2 Das Testprotokoll.....	297
4.3 Arbeitsanweisung.....	298
5. AUSWERTUNGSANLEITUNG	300
5.1 Quantitative Auswertung	300
5.1.1 Reguläre Normierungstabellen.....	300
5.1.2 Normierungstabellen aufgrund modifizierter Fehlerbewertung	307
5.2 Qualitative Auswertung	309
6. LITERATURVERZEICHNIS DES ANWENDERHANDBUCHS.....	320

VERZEICHNIS DER BENUTZTEN LITERATUR

A. Ausgangslage und Ziel der Untersuchung

1. Problembereich der Untersuchung

Die vorliegende Arbeit setzt sich mit der Erfassung lexikalisch-semantischer Fähigkeiten¹ gehörloser Grundschüler² auseinander. Sie versteht sich als ergänzendes und fortsetzendes Element der wissenschaftlichen Begleitforschung des bilingualen Schulversuchs³ in Hamburg. Den Hintergrund sowohl dieses Schulversuchs als auch der vorliegenden Arbeit bildet die kontroverse Diskussion in Deutschland um die Qualität unterschiedlicher gehörlosenpädagogischer Methoden.

Einer 1999 (Günther 1999 a) erschienenen Dokumentation können die Konzeption des Hamburger Bilingualismus, die damit bisher gemachten unterrichtlichen Erfahrungen sowie erste Ergebnisse der Begleitforschung entnommen werden. Sämtliche darin dokumentierte Untersuchungen, zu denen auch das vorliegende Projekt zählt, zielen auf die Klärung von Fragen der Sprachentwicklung. Dabei werden jeweils unterschiedliche Aspekte der Sprachentwicklung gehörloser Grundschüler unter bilingualen schulischen Sozialisationsbedingungen beleuchtet.

Die wissenschaftliche Begleitforschung stand vor der besonderen Schwierigkeit, dass empirisch fundierte Belege fast gänzlich fehlten. So wurde zwar über lange Zeit eine Methodendiskussion geführt, die sich jedoch lediglich auf Fragmente eines wissenschaftlich gesicherten Rahmen beziehen konnte. Die verschiedenen den Schulversuch begleitenden Forschungsprojekte sollen eine breite empirische Basis erarbeiten. Dazu wurde es vielfach zunächst notwendig, Instrumentarien und Verfahren zu entwickeln, mit denen das gewünschte Datenmaterial erhoben werden konnte.

Wie die Dokumentation darlegt, besteht auf Grund der Forschungstätigkeiten im Rahmen der wissenschaftlichen Begleitung derzeit ein z. T. vorläufiger, z. T. bereits fortgeschrittener Kenntnisstand zur Thematik des Hamburger bilingualen Ansatzes. Die vorliegende Auseinandersetzung mit dem sprachlichen Entwicklungsniveau gehörloser Kinder ergänzt diesen Kenntnisstand zunächst um weiteres im Rahmen einer umfangreichen Untersuchung erhobenes Datenmaterial. Auf diesem somit vervollständigten Hintergrund wissenschaftlich gesicherter Ergebnisse wird vorliegend eine neue Grundlage zur differenzierten Betrachtung und

¹Definitische Ausführungen zu dem Begriff der ‚lexikalisch-semantischen Fähigkeiten‘, sowie angrenzenden Begriffen finden sich unter A 2.3. Begriffsbildung und Wortschatz

² In Zusammenhang mit Personen bezeichnenden Substantiven, die ausschließlich in der männlichen Pluralform vorkommen, wird für die vorliegende Arbeit Gleichberechtigung der damit bezeichneten männlichen und weiblichen Personen festgestellt.

³ Erläuterung unter A 2.1 Bilingualer Schulversuch Hamburg

Bewertung nicht nur des bilingualen, sondern sämtlicher Ansätze zur Sprachvermittlung an Gehörlose geschaffen. Auf diese Weise eröffnet sich die Möglichkeit für die Anfang/Mitte der Neunzigerjahre kontrovers-polemisch und somit fruchtlos gewordene Methodendiskussion, in eine neue, konstruktive Phase der Auseinandersetzung eintreten zu können. Dabei sollte die Bereitstellung gesicherter Fakten das Feld persönlicher Einschätzungen im Rahmen der Methodendiskussion schmälern. Dies sollte dazu führen, dass die Weiterentwicklung und Verbesserung der jeweiligen Methoden wissenschaftlich unterstützt und somit effektiver als bisher vorgenommen werden können. Zudem bzw. durch diese Möglichkeit sollte der wissenschaftliche Diskurs durch Objektivität und somit entstehende Akzeptanz differierenden Ansätzen gegenüber bereichert werden.

Eine derartige Fortführung der Diskussion erscheint aus mehreren Gründen erforderlich. So betrifft zunächst die Entscheidung über die Methode der Sprachvermittlung an gehörlose Kinder nicht isoliert deren Kompetenzen in den verschiedenen sprachlichen Modalitäten Gebärden-, Schrift- und Lautsprache⁴. Der Lern- und Entwicklungsbereich Sprache steht vielmehr in engen wechselseitigen Beziehungen zu der kognitiven Entwicklung, der Identitätsbildung, den sozial-kulturellen Entfaltungsmöglichkeiten u. a. Auf Grund dieser Abhängigkeitsverhältnisse kommt der sprachlichen Entwicklung des Einzelnen und damit der Methode zur Anbahnung kommunikativer Kompetenzen eine besondere Bedeutung zu. Zudem bringt die auf Grund der Kontroverse ungeklärte wissenschaftliche Landschaft konkrete Orientierungsprobleme Betroffener bzw. für betroffene Minderjährige Verantwortlicher mit sich. So unterscheiden sich die Beratungen, die diese erhalten, je nach Bundesland, Profession und persönlicher Einstellung der jeweiligen Berater u. U. deutlich voneinander oder widersprechen einander gar grundlegend. Dabei muss es als die Pflicht der Vertreter unterschiedlicher Positionen bzw. der Teilnehmer an dem wissenschaftlichen Diskurs angesehen werden, ihrer jeweiligen Beweispflicht nachzukommen, um so auf der Ebene wissenschaftlich-theoretischer Grundlegung in eine sachliche Diskussion eintreten und in der Praxis Informationen und Orientierungshilfen geben zu können. Dies ist bislang unzureichend realisiert worden. Ziel der vorliegenden Erarbeitung einer gehörlosensspezifischen Datenerfassungsmethode ist es, an der Aufarbeitung dieser Versäumnisse mitzuwirken.

⁴ Ist im Folgenden von ‚den Sprachmodalitäten‘ die Rede, so meint dieser Begriff Gebärden-, Schrift- und Lautsprache.

2. Forschungsstand

2.1 Bilingualer Schulversuch Hamburg

Im Sommer 1991 wurde der erste, in Deutschland bislang einmalige, bilinguale Schulversuch durch eine Elterninitiative ins Leben gerufen und begann an der Hamburger Gehörlosenschule. Er wurde von Anfang an wissenschaftlich begleitet, wobei sowohl die theoretische bilinguale Konzeption als auch deren praktische Umsetzung aus der Praxis des Schulversuchs selbst entwickelt wurden. Dies versteht sich auf dem Hintergrund des Erkundungscharakters, der den Schulversuch kennzeichnete.

Grundlage war die in der Entwicklung befindliche bilinguale Methode nach Günther, Prillwitz, Wudtke u. a., die aus dem lautsprachbegleitenden Ansatz nach Prillwitz, Wudtke, Wisch u. a. hervorgeht und die Deutsche Gebärdensprache (im Folgenden ‘DGS’) konzeptionell einbezieht. Sie basiert, neben Erkenntnissen unterschiedlicher wissenschaftlicher Disziplinen, auf linguistischen Forschungsergebnissen, welche die DGS als vollwertige, der Lautsprache ebenbürtige Sprache ausweisen (Prillwitz 1975, 1982). Dies ermöglicht ihren Einsatz in der Erziehung gehörloser Kinder parallel zu Lautsprachanbahnung und Artikulationsunterricht.

Im Gegensatz zu der früheren, oralen Theorien⁵ zu Grunde liegenden Annahme wird davon ausgegangen, dass der Einsatz der DGS im Prozess der Sprachanbahnung gehörloser Kinder förderliche Auswirkungen auf die Lautsprach- und die gesamte psycho-soziale Entwicklung hat. Da den Betroffenen mit der vergleichsweise leicht zu erlernenden DGS ein Kommunikationssystem als Basis für Sprachlernprozesse zur Verfügung steht, wird angenommen, dass deren Lautsprachentwicklung begünstigt wird. Zudem wird in Hinblick auf die psycho-soziale Entwicklung der Betroffenen bevorzugt, Entscheidungen über spätere Integrationsmuster nicht bereits durch die Sprachanbahnungsmethode zu bestimmen, sondern offen zu halten. Auf diese Weise besteht für den erwachsenen Gehörlosen die Wahlmöglichkeit zwischen der Gehörlosengemeinschaft und der hörenden Gesellschaft. Demgegenüber bedeutet das singuläre Erlernen der Lautsprache i. d. R. eine Festlegung auf die hörende Gesellschaft als Bezugsgruppe bzw. eine Entfremdung von der Gehörlosengemeinschaft. Schwer wiegende psycho-soziale Folgen hat dies insbesondere bei unvollständiger Beherrschung der Lautsprache, da somit eine tragfähige Kommunikation für die Betroffenen unmöglich ist. Da entsprechende Fälle nachweislich (Günther 1999 b, 177) zahlreich vorkommen, besteht Anlass zum Umdenken.

⁵ Erläuterungen unter A 2.2 Gehörlosenpädagogische Methodendiskussion

In der wissenschaftlichen Begleitung des Schulversuchs (Günther u. a. 1999 a) konnte zunächst auf keine dokumentierten oder gar gesicherten Erfahrungen mit der Thematik zurückgegriffen werden. Im Verlauf des Versuchs befruchteten Erfahrungen mit vergleichbarer Pionierarbeit in den skandinavischen Ländern die Arbeit. Dabei wurden in Hamburg die Ansätze aus Schweden (s. Ahlgren 1993; die Beiträge in Ahlgren und Hiltensstamm 1994; Heiling 1995; Henning 1997) und Dänemark (s. Hansen 1993; Lewis 1995; s.a. für beide Länder Mashie 1995) nicht einfach adaptiert, sondern eine eigene, sich von diesen unterscheidende Realisierungsform auf dem Hintergrund übereinstimmender Prinzipien entwickelt. Diese zeichnet sich durch einen ausdrücklich betonten Wechsel zwischen Gebärden-, Schrift- und Lautsprache aus. So wird der bilinguale Unterricht im Team von hörender Gehörlosenlehrerin und gehörloser Pädagogin durchgeführt. Dadurch kann die allgemeine Unterrichtssprache als bilingual bezeichnet werden. Die klassische bilinguale Konzeption, von der in den skandinavischen Ländern ausgegangen wird und die bereits auf de l'Epee und Sicard zurückzuführen ist, sieht im Unterschied dazu vor, die Gebärdensprache als Primärsprache entsprechend der Lautsprache bei hörenden Kindern zu erwerben und sich auf der Basis dieser Kompetenz die Laut-/Schriftsprache als Zweitsprache anzueignen.

Diese neue und bisher einzigartige Realisierungsform bilingualer Prinzipien⁶ warf einen Katalog klärungsbedürftiger Fragestellungen auf und erforderte daher eine wissenschaftliche Untersuchung. Kernanliegen dieser Untersuchung stellte die Durchleuchtung der Sprachentwicklung der an dem Versuch teilnehmenden Schüler dar, um auf diesem Hintergrund gesicherte Aussagen über die Stärken sowie die veränderungs- und entwicklungsbedürftigen Schwächen des Ansatzes treffen zu können. Als vorrangiges Anliegen galt dabei die Überprüfung eigener Zielsetzungen anhand der Ergebnisse. Man bewegte sich also bewusst zunächst im Rahmen der eigenen Konzeption, ohne sich bereits auf das zuvor ausgiebig betretene Terrain der kontroversen Diskussion in der deutschen Hörgeschädigtenpädagogik zurückzugeben.

Im Folgenden sollen die Ergebnisse dieser Untersuchungen skizziert werden, um so den wissenschaftlichen Kenntnisstand, in den sich die vorliegende Arbeit eingliedert, zu verdeutlichen. Eine ausführliche Darstellung dieser Ergebnisse findet sich im Zwischenbericht der wissenschaftlichen Begleitung des Schulversuchs (Günther u. a. 1999 a). Zunächst zeigen die Berichte über die bilinguale unterrichtliche Praxis, dass das Unterrichtsangebot in allen Lern- und Leistungsbereichen dem Angebot in Grundschulen entspricht. Somit wurde eine Angleichung an das Niveau altersgleicher, nichtbehinderter Kinder realisiert. Wie den Berichten zu entnehmen ist, erreichen die gehörlosen Schüler auf diesem Angebotshintergrund einen grundlegenden Bildungsstand und entwickeln ihre kommunikativen Fähigkeiten altersgemäß.

⁶ Weitere Erläuterungen unter A 2.2 Gehörlosenpädagogische Methodendiskussion

In einer der Untersuchungen der Begleitforschung wurde der Frage nachgegangen, ob in einem bilingualen Konzept zwischen Schriftspracherwerb und Deutscher Gebärdensprache positive oder negative Interferenzen bestehen. Dazu wurde die systematische Analyse eines von einer bilingual geförderten Schülerin verfassten Textes vorgenommen. Dieser Text fiel auf durch seine Diskrepanz zwischen hohem Textkonstitutionsniveau und zahlreichen Abweichungen von den Normen des Deutschen. Diese Abweichungen, so zeigen die Ergebnisse, sind systematisch zurückzuführen auf die Strukturen der Deutschen Gebärdensprache. Diese übernehmen offensichtlich strukturelle Lücken der Schreiberin im Deutschen. Der Vergleich mit einem später verfassten Text derselben Person unterstützt die Annahme, dass es sich hierbei um eine positiv-produktive Interferenz handelt, da die Übernahme von DGS-Strukturen hier bereits weitgehend durch morpho-syntaktische Strukturen ersetzt wurde. Der Umgang der Verfasserin des Textes mit der DGS als ‚konstruktivem Lückenbüßer‘ für fehlende Kenntnisse im Deutschen verdeutlicht zudem, dass die DGS eine leistungsstarke sprachliche Basis darstellt.

In einer weiteren Untersuchung wurden Textanalysen von freien Texten der bilingual geförderten gehörlosen sowie aural geförderten schwerhörigen bzw. an Taubheit grenzenden Schüler vorgenommen. Als Schreib Anlass wurde eine Bildergeschichte vorgelegt, wobei diese Form freien Schreibens für alle Kinder ungewohnt war. Die textanalytische Untersuchung ergab zunächst eine Vermischung der Kompetenzen auf der inhaltlichen Text- und der formalen Sprachebene. So ist seitens der schwerhörigen Schüler, die lautsprachorientiert unterrichtet werden, festzustellen, dass Morphosyntax und Grammatik i. d. R. sicher von ihnen beherrscht werden und sie zudem über einen differenzierten und umfangreichen Wortschatz verfügen. Dieser gut entwickelten Beherrschung des Deutschen auf formaler Ebene steht ein überwiegend niedriges Textkompetenzniveau gegenüber. Der überwiegende Anteil der gehörlosen Schüler weist ebenfalls ein geringes Textkompetenzniveau auf, während jedoch zwei von ihnen Texte auf mit Abstand höherem Niveau verfassen. Formale Abweichungen vom Deutschen erklären sich dabei durch systematische Rückgriffe auf eine Sprachkompetenz, die auf der DGS basiert. Eine der Grundannahmen des bilingualen Ansatzes, dass die DGS den gehörlosen Schüler eine leistungsstarke sprachliche Basis bereitstellt, bestätigt sich hierin erneut, wie bereits mit der vorab skizzierten Untersuchung belegt.

Schließlich wurde eine Untersuchung zu den Lesekompetenzen bilingual gehörloser sowie schwerhöriger Schüler durchgeführt. Dazu wurde beiden Gruppen der gleiche Text vorgelegt, wobei bei der Auswertung die spezifischen Lesebedingungen Gehörloser (sichtbare Begleitgebärden, Mundbild etc.) berücksichtigt wurden. Im Ergebnis weisen alle Untersuchungspersonen eine dominant grafische Orientierung auf, was verdeutlicht, dass sich die Lesekompetenzentwicklung der Schwerhörigen nicht auf einem den Gehörlosen gegenüber fortgeschrittenen Niveau befindet. Es zeigten sich jedoch auch Unterschiede zwischen den

Gruppen. So orientierten sich die gehörlosen Schüler lesestrategisch eher an semantischen Strukturen, während die schwerhörigen Untersuchungspersonen eine Orientierung an morpho-syntaktischen Strukturen vornahmen. Die Begrenztheit der Stichprobe lässt es nicht zu, diese Schlüsse zu verallgemeinern. Weiter gehende Untersuchungen, bei denen abgefragt wurde, was im Einzelnen von den Schüler inhaltlich verstanden wurde, weisen jedoch wiederum darauf hin, dass die bilingualen Schüler über das tendenziell bessere Textverständnis verfügen. Diesem Ergebnis wird auf Grund des Umstandes besondere Relevanz beigemessen, dass schwerhörigen Schüler mit den ihnen eigenen behinderungsspezifischen Bedingungen weiterreichende Möglichkeiten für das Erlernen von Kompetenzen zur Verfügung stehen als den gehörlosen. Insofern erscheint auch im Bereich des Lesens der bilinguale Ansatz gut zur Förderung der Kompetenzen der Schüler geeignet.

Diese Ergebnisse weisen sämtlich darauf hin, dass der bilinguale Ansatz gut geeignet scheint, die sprachlichen Kompetenzen gehörloser Schüler zu fördern. Dies bildet für das vorliegende Projekt den Ausgangspunkt zur Erarbeitung weiterer Daten und Untersuchungsergebnisse, die zur Erforschung und Weiterentwicklung des bilingualen Ansatzes beitragen.

2.2 Gehörlosenpädagogische Methodendiskussion

Übereinstimmender Ausgangspunkt jedes Sprachvermittlungsverfahrens ist die Frage, auf welche Weise ein gehörloser Mensch zur Sprachfähigkeit und damit zu einer sprachlich strukturierten Denk- und Kommunikationsfähigkeit als wesentlicher Grundlage seiner personalen und sozialen Existenz geführt werden kann. Einigkeit besteht zudem in der Einsicht, dass dem gehörlosen Menschen (wie jedem nichtbehinderten Menschen auch) zur Entfaltung seiner Identität und positiven Gestaltung seines Lebens mindestens ein funktionsfähiges Kommunikationssystem zur Verfügung stehen muss (vgl. Ahrbeck 1992). Obgleich demnach Konsens in wesentlichen grundlegenden Annahmen besteht, differieren die Aufstellungen methodischer Richtlinien zur Sprachvermittlung deutlich. Übereinstimmende Beschreibungen von Ausgangslage und Zielvorstellung führen nicht zu einer einheitlichen Wegbeschreibung.

Die historische Entwicklung gehörlosenpädagogischer Ansätze zur Sprachvermittlung, die im 16. Jahrhundert in Spanien mit dem Engagement Ponce de Leons ihren Anfang nahm, wurde beeinflusst durch die jeweils in ihrer Zeit hinzukommenden wissenschaftlichen Erkenntnisse unterschiedlicher Disziplinen. Der Untersuchungsbereich der Sprachvermittlung an Gehörlose wird neben der Linguistik tangiert von Psychologie und Pädagogik, Medizin und Audiologie. Wie Jann (1994) deutlich herausarbeitet, liegen die Differenzen der ver-

schiedenen Theorien maßgeblich darin begründet, dass ihren Schöpfern jeweils unterschiedliche Wissenschaftsdisziplinen als Ausgangspunkt dienen. Neben solchen sachlichen Bezügen stehen hinter den differierenden Ansätzen nach Günther (1999 b) verschiedene 'Bildungsphilosophien'. Da sich somit die jeweiligen Ausgangspunkte unterscheiden, ist es nicht sachgemäß, eine vergleichende Diskussion mit dem Ziel zu führen, das Versagen des jeweils anderen nachzuweisen.

Bereits vor einem Jahrhundert wurde als Reaktion auf einen langwierigen gehörlosenpädagogischen Methodenstreit der Mailänder Kongress einberufen. Die Beschlüsse dieses Kongresses hatten faktisch zur Folge, dass keine Gebärden im Unterricht mit gehörlosen Kindern verwendet wurden. Dieses Vorgehen wird bis heute von den Vertretern der sog. 'oralen Methode'⁷ praktiziert. Heute befindet sich die Hörgeschädigtenpädagogik wieder in der Situation kontroverser Auseinandersetzungen über die Angemessenheit unterschiedlicher Sprachvermittlungsverfahren. Kernfrage dieser Methodendiskussion ist, ob unter Berücksichtigung des Ziels, gehörlose Menschen gesellschaftlich integriert und ohne gravierende Schwierigkeiten in ihrer emotionalen, kognitiven und sozialen Entfaltung aufwachsen und leben zu lassen, die Lautsprache ausreichend kompetent erlernt werden und sich die Einbeziehung von Gebärden unterstützend auswirken kann.

Eine Versachlichung der Diskussion zwischen den methodischen Lagern erscheint nicht nur auf wissenschaftlicher Ebene angezeigt, sondern vor allem durch ihre verunsichernden Auswirkungen in der Praxis für Eltern, Lehrer und die betroffenen Kinder selbst erforderlich. Die für die gehörlosen Kinder Verantwortung tragenden Eltern und Pädagogen sehen sich in der Situation, in einer wissenschaftlich-methodisch ungeklärten Landschaft existenziell bedeutsame Entscheidungen treffen zu müssen. Es muss daher als zentrale Aufgabe der Gehörlosenpädagogik auf wissenschaftlicher Ebene erachtet werden, ausgehend von gemeinsamen Zielsetzungen über eine sachlich geführte und wissenschaftlich fundierte Diskussion zu einem Konsens zu gelangen.

Da der Sprachstatus der Gebärdensprache heute grundsätzlich belegt ist, entfällt der Ansatz von Legitimation einer gehörlosenpädagogischen Methode allein oder wesentlich durch linguistische Argumente. Stattdessen gewinnen bei der Entscheidung für die Ausrichtung nach bestimmten Zielen und Methoden sozialpsychologische und soziologische Aspekte an Relevanz. Neben dem Bemühen, die Voraussetzungen zu schaffen für eine möglichst ungehinderte sprachliche Kommunikation, darf die Gehörlosenpädagogik nach Voit (1992) nicht die Entscheidung über künftige Integrationsmuster der Betroffenen fällen, sondern sollte vielmehr auf selbstverantwortete Integrationsmuster abzielen. Auf diese Weise würden sich gehörlosenpädagogische Konzepte künftig nicht mehr grundsätzlich, sondern nur noch in Ge-

⁷ Erläuterung weiter unten in diesem Kapitel

wichtung und Reihenfolge ihrer Angebote unterscheiden. So könnte die Vision Voits realisiert werden, sich ausgehend von dem kontroversen Nebeneinander der gegenläufigen Strömungen im Sinne der Dialektik weiterzuentwickeln, um auf höherer Ebene eine Position zu erreichen, in der sich scheinbar Widersprüchliches vereint. In Übereinstimmung mit dieser Position formuliert Günther (1999 a, 14): „Der Hamburger [bilingualen A. d. V.] Philosophie liegt der Versuch zu Grunde, die heutigen aural / oralen und gebärdensprachlichen Möglichkeiten in einem Konzept zu vereinen, das den heranwachsenden hörgeschädigten Menschen als Gesamtpersönlichkeit über die gesamte Lebensspanne sieht.“

Die Geschichte gehörlosenpädagogischer Sprachvermittlungsverfahren im deutschsprachigen Raum seit dem zweiten Weltkrieg weist zunächst einen breiten Sockel oralistischer Methoden auf. Diese differieren in (durchaus wesentlichen) Details, stimmen jedoch, basierend auf der Grundannahme der Unmöglichkeit differenzieller Sprachwahrnehmung der Betroffenen über das Hören, in dem zentralen Punkt der Ablehnung des Einsatzes von Gebärden im Sprachvermittlungsprozess überein. Traditionell wird unter oral ausgerichteten Gehörlosenpädagogen davon ausgegangen, dass ein funktionsfähiges Kommunikationssystem Gehörloser lautsprachlicher Art sein müsse. Dies wird zunächst mit integrativen Erwägungen begründet. Ein gehörloser Mensch muss, so die Position, als Angehöriger einer (gebärdensprachlichen Minderheit dazu befähigt werden, sich in die hörende, lautsprachliche Mehrheit sprachlich und damit auch kulturell-sozial integrieren zu können. Dies sei nur möglich, indem er sich die Lautsprache als Kommunikationsmittel der Mehrheit aneignet. Auf Grund der erheblichen Schwierigkeiten, die dieses Vorhaben für gehörlose Menschen bedeutet, müsse die gesamte innerhalb des Erziehungsprozesses zur Verfügung stehende Zeit und Konzentration auf die Lautsprachanbahnung verwendet werden. Dabei wird davon ausgegangen, dass kein alternatives Kommunikationsmittel zur Verfügung steht, weil die DGS von den Vertretern dieser Position nicht als vollwertige Sprache anerkannt wird. Zudem wird angenommen, dass der Einsatz von Gebärden negative Auswirkungen auf die Lautsprachentwicklung hat. Die orale Position schliesst sich damit der seit den Beschlüssen des Mailänder Kongresses geltenden Vorgabe an, den Einsatz von Gebärden auf Grund seiner angenommenen kontraproduktiven Auswirkungen abzulehnen.

Diesen Prämissen gemäß musste Lautsprache von den betroffenen Gehörlosen über die ungenauen Ablesegestalten aufgenommen werden sowie ihre Sprachproduktionsmöglichkeiten über die unnatürliche Artikulation aufgebaut werden. Die durch diese Praxis entstandenen erheblichen Probleme und, gemessen an der Zielvorgabe eines möglichst differenzierten Spracherwerbs, deutlich unbefriedigenden Erfolgsquoten führten, entgegen der Vorgabe rein oraler Erziehungsmethoden, zu der unkontrollierten und wissenschaftlich nicht begleiteten

Hinzunahme verschiedener manueller Hilfssysteme. Es entstand eine Form unreflektierter ‚Totaler Kommunikation‘⁸.

Auf Grund der größtenteils fragmentarischen Lautsprachkompetenzen rein oral erzogener Gehörloser wurden Zweifel an den methodischen Prinzipien immer lauter. Die sich an die unbefriedigenden Ergebnisse anschließende Erneuerungsbewegung führte zunächst zu einer Ergänzung der oralen Methode durch aurale Elemente⁹ und/oder einer Einbeziehung des lautsprachbegleitenden Gebärdens (LBG). Heute wird die orale Methode in ihrer reinen Form zudem durch entscheidende Fortschritte auf dem Sektor der Entwicklung technischer Hilfsmittel (Hörgeräte, Cochlea-Implantat) zur Nutzbarmachung von Hörresten in Frage gestellt. Daneben bedeutete die Erforschung der Gebärdensprache und die damit einhergehende Entdeckung, es dabei mit einem der Lautsprache ebenbürtigem Sprachsystem zu tun zu haben, eine Wendung sowohl für die Bewertung bestehender Ansätze als auch für die Entwicklung neuer Ideen.

So bildeten sich die beiden bis heute maßgeblichen Methoden der deutschen Hörgeschädigtenpädagogik heraus: Die hörgerichtete Erziehung (auch als auraler Ansatz bezeichnet) nach Schmidt-Giovanini (1996) und Diller (1998) und die bilinguale Erziehung nach Henning (1997) und Heiling (1995), wobei die letztere in ihrer Ursprungsform zunächst in Skandinavien praktiziert wurde. Trotz der gemeinsamen Zielsetzung der Sprachanbahnung gehörloser Kinder und des gemeinsamen Ausgangspunktes der Ablehnung des klassischen oralen Ansatzes grenzten sich diese beiden Positionen nachdrücklich gegeneinander ab.

Die Konzentration auf die Hörerziehung innerhalb des auralen Ansatzes wird mit der Annahme begründet, dass auf dem Hintergrund neurophysiologischer Erkenntnisse eine gewisse Hörbahnreifung möglich ist. Um diese innerhalb einer sensiblen Phase bestehenden Möglichkeiten ausschöpfen zu können, muss nach diesem Ansatz der Übung auditiver Wahrnehmung eine kompromisslose Vorrangstellung eingeräumt werden, was zur Ablehnung des Einsatzes jeglicher manueller Systeme führt, da diese negativ überlagernde Auswirkungen mit sich brächten.

Demgegenüber bezieht die ursprünglich aus dem schwedischen Ansatz hervorgegangene bilinguale Methode den Standpunkt, dass die beiden Sprachmodalitäten Laut- und Gebärdensprache in der dargestellten methodischen Weise gegenseitig befruchtend zum Einsatz kommen sollten. In der Auseinandersetzung mit Vertretern der hörgerichteten Methode wird hervorgehoben, dass seitens des bilingualen Ansatzes das allgemeine Prinzip der Hörresteevo-

⁸ Der Begriff der ‚Totalen Kommunikation‘ wird verwendet für eine Praxis, in der jegliches brauchbar erscheinendes und zur Verfügung stehendes Kommunikationsmittel, sei es lautsprachlicher oder manueller Art, Verwendung findet.

⁹ Erläuterung weiter unten in diesem Unterkapitel

zierung und entsprechender Hör-Spracherziehung nicht nur akzeptiert, sondern als notwendig erachtet wird.

So entwickelte sich in jüngster Zeit, maßgeblich angeregt und formuliert durch Günther (1999 b), der Standpunkt, beide Ansätze nebeneinander gelten zu lassen und die Entscheidung darüber, welche von beiden bzw. in welcher Gewichtung beide im Einzelfall zum Einsatz kommen sollten, im Rahmen parallel bestehender Einsatzmöglichkeiten zu fällen. Die Vision eines sog. Förder- und Ressourcenzentrums für Hörgeschädigte als Institution zur Realisierung eines solchen Ansatzes soll hier nur genannt, nicht aber ausgeführt werden.

Durch die vorliegende Arbeit der Entwicklung eines Wortschatztests in Gebärden-, Schrift- und Lautsprache für gehörlose Kinder im Grundschulalter soll eine angemessene Bewertung der sich unterscheidenden methodischen Wege vorangetrieben bzw. erst ermöglicht werden. Dieser Beitrag versteht sich im Dienste der Optimierung in Entwicklung befindlicher gehörlosenpädagogischer Konzepte und nicht der Widerlegung von Ansätzen, denen man sich nicht verschrieben fühlt. Das Instrument selbst ist grundsätzlich neutral. Die mit ihm zu erhebenden Daten und die sich daraus ergebenden Ergebnisse sollen ihren Nutzen finden in einer gehörlosenpädagogischen Landschaft, die gerade heute durch die aktuellen Entwicklungen neuer Methoden der Sprachvermittlung sowie durch die medizinische Versorgung mit Implantaten u. a. m. Flexibilität erfordert.

2.3 Begriffsbildung und Wortschatz

Nach Piaget sind beim Kleinstkind zunächst ganzheitliche Begriffsstrukturen vorhanden, d. h. es bestehen erste Vorstellungsbilder als früheste Form geistiger Nachkonstruktion von Gegenständen, Handlungen und Ereignissen. Diese werden in tätiger Auseinandersetzung mit der Umwelt durch Assimilation und Akkomodation nach und nach ausdifferenziert und in einzelne Begriffsmerkmale zergliedert. Im Laufe dieses Prozesses wird in einem nächsten Schritt die so entstandene geistige Struktur mit einem Wort assoziiert. Begriffe, d. h. inhaltliche Vorstellungen, sind demnach primär. Sie werden erst in einem zweiten Schritt mit einer sprachlichen Bezeichnung assoziiert. Begriffe hängen somit in ihrer Entstehung nicht direkt von Sprache ab, wenngleich sie in der Regel im Laufe ihrer Entwicklung mit sprachlichen Ausdrücken belegt werden (in Wersich 1992).

Piaget beschreibt als einen entscheidenden frühen Entwicklungsschritt die Unterscheidung zwischen dem Ich und den Objekten einer Welt, die außerhalb des Ich existiert. Beim Kind reift die Erkenntnis, dass diese Objekte auch Bestand behalten, wenn sie sich außerhalb der

direkten Wahrnehmung des Ichs befinden. Für diesen Entwicklungsschritt prägte er den Begriff der ‚Objekt Konstanz‘. Diese sei die erste Manifestation der sich entwickelnden Intelligenz. Damit besteht eine innere Repräsentation für Objekte bei dem Kind, sodass es nun nicht länger nur auf etwas reagieren kann. Dies wird als wesentliche Voraussetzung für symbolische Aktivität verstanden. Ist bei Piaget von Symbolbildung des denkenden Individuums die Rede, so ist unter ‚Symbol‘ ein objektives äußeres Ereignis zu verstehen, das in direkter Beziehung zum Individuum steht und nur indirekt mit dem durch es bezeichneten Objekt verbunden ist. Symbole können bspw. Wörter sein und die Sprache ist ein Symbolsystem. Um sich ein solches Symbolsystem aneignen und es sinnvoll anwenden zu können, müssen die entsprechenden Denkstrukturen beim Kind bereits entwickelt sein.

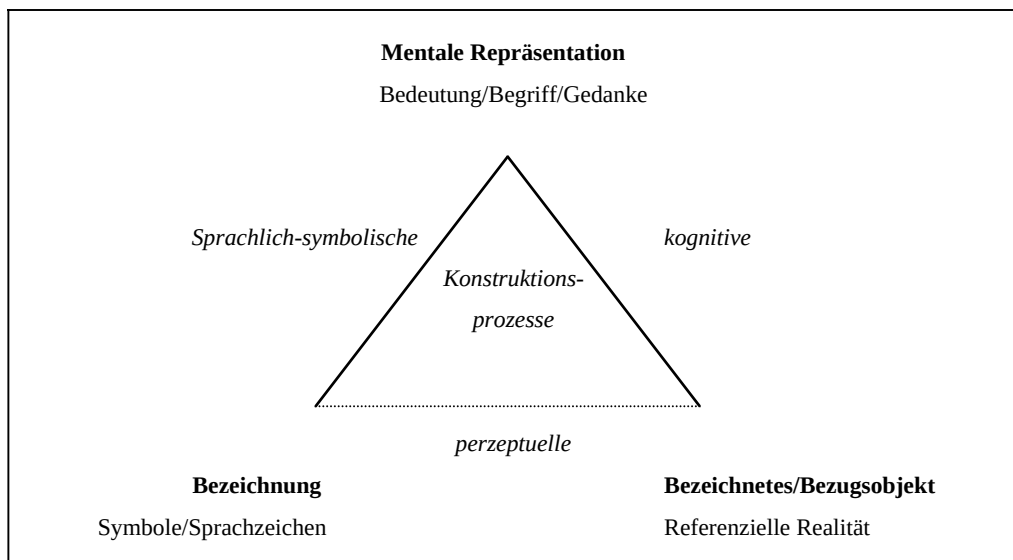
Diese Forschungsergebnisse Piagets bilden die Grundlage der Arbeit Furths (1972), der es sich zur Aufgabe macht, die weit verbreitete These zu überprüfen, dass mangelnde sprachliche Fähigkeiten den Rückschluss auf ein niedriges Intelligenzniveau zulassen. Zentrales Thema Furths ist die Erforschung der spezifischen Situation Gehörloser, da bei diesen einerseits in aller Regel mangelnde (laut)sprachliche Kompetenz vorfindlich ist, andererseits jedoch die davon abgeleitete Annahme mangelnder denkerischer Fähigkeiten unbewiesen existiert. Da die dargestellte Theorie Piagets zu dem Schluss gelangt, dass denkerische Fähigkeiten des Menschen entwicklungstheoretisch unabhängig sind von Sprachkompetenz, ist nach Furth der oben formulierte Rückschluss unzulässig.

So ist Sprache nach Furth (1972, 231) „ein gebrauchsfertiges konventionelles Symbolsystem, das von der Intelligenz je nach ihrer Struktur aufgenommen wird. [...] Sprachkompetenz [...] muss von mehr oder minder intelligentem Gebrauch der Sprache unterschieden werden.“ Gemäß der von Furth (232) herauskristallisierten Erkenntnis, „dass zwischen Sprache und operationalem Denken keine wesentliche Beziehung besteht und während der Entwicklungsphase auch nicht bestehen muss, [...] [erklärt sich die Ableitung, dass] die grundsätzliche Entwicklung und Struktur der Intelligenz bei den Gehörlosen im Vergleich zu den normal Hörfähigen bemerkenswert unberührt von dem Nichtvorhandensein der verbalen Sprache [ist].“ So folgert Furth abschließend, „dass logisches, intelligentes Denken nicht durch ein symbolisches System, wie die lebendige Sprache einer Gesellschaft es darstellt, unterstützt zu werden braucht. Denken ist ohne Zweifel ein inneres System, eine hierarchische Anordnung der Interaktion eines Individuums mit der Welt. Das Symbolsystem der Sprache reflektiert und drückt in gewisser Weise diese innere Organisation aus. Aber die innere Organisation der Intelligenz ist nicht vom Sprachsystem abhängig; im Gegenteil, das Verständnis und die Anwendung der schon gebrauchsfertigen Sprache hängt von der Struktur der Intelligenz ab.“ (233)

Bei Furth (43) wird „das Wort ‚Begriff‘ als künstliche, abstrahierte Einheit des Denkverhaltens verstanden [...]. Man kann sogar immer ‚Begriff‘ durch das Wort ‚Denken‘ ersetzen.“ Der Terminus ‚Begriff‘ findet auch bei Probst Verwendung, wenn er definiert: „Am Anfang steht also etwas ‘Gemeintes’, ein mehr oder weniger klares Bild vom Wesentlichen an einer Reihe von wiederkehrenden Ereignissen. Dies ist der Inhalt des Begriffes. [...] Ich kann diesen Inhalt ‘meinen’, also einen Begriff dieser Klasse von Objekten oder Ereignissen haben, ohne ein Wort dafür zu verwenden oder gar zu kennen. Begriff und Wort sind also nicht dasselbe; das Wort ist auch keine Voraussetzung für den Begriff“ (1981, 9).

Zur Veranschaulichung dieses Zusammenhanges von Begriff und Sprache dient die folgende grafische Darstellung, das sog. „semiotische Dreieck“ nach Ogden/Richards (1923, 74), das auch den Ausführungen Günthers (1991, 167) als das „Standardmodell der semantisch orientierten Linguistik“ zu Grunde gelegt wird:

Abb. 1-A 2.3: Das semiotische Dreieck



Zwischen der mentalen Repräsentation, d. h. dem Gedanken oder Begriff, und dem Symbol oder Sprachzeichen besteht eine kausale Beziehung. Auch zwischen dem Gedanken und dem Bezugsobjekt, also der referenziellen Realität, besteht eine (mehr oder weniger) direkte Beziehung. Zwischen der Bezeichnung, also dem Symbol, und dem Bezeichneten besteht jedoch keine direkte Beziehung, sondern nur die indirekte, über den Gedanken oder Begriff vermittelte. Die entscheidende, über dieses Modell des semiotischen Dreiecks verdeutlichte Einsicht besteht darin, zwischen Bezeichnung und Bezeichnetem eine klare Trennung zu erkennen und gleichzeitig die mentale Repräsentation, also den Begriff, als grundsätzlichen Vermittler zu verstehen. Dies bedeutet, dass eine Person, welche in der Lage ist, die Symbole bzw. Sprachzeichen der durch sie bezeichneten Realität korrekt zuzuordnen, auch um

die Bedeutung dieser Symbole wissen muss, d. h. einen Begriff von dem Bezeichneten haben muss.

Demnach wird deutlich, dass durch kognitive Konstruktionsprozesse real existente Gegenstände, Vorgänge oder Zustände mental repräsentiert werden, d. h. die Person verfügt über eine inhaltlich-begriffliche Vorstellung davon. Diese begrifflichen Vorstellungen werden durch sprachlich-symbolische Konstruktionsprozesse in Form von Sprachzeichen zum Ausdruck gebracht. (Solche Symbole können sowohl lautsprachliche Äußerungen als auch Schrift- oder Gebärdenzeichen sein.) Die Bezeichnungen weisen dann auf die durch sie benannte Realität, wodurch sich der Kreis (bzw. das Dreieck) schließt.

Somit ist die Annahme begründet, dass die Abfrage von sprachlichen Zeichen den direkten Rückschluss auf das Vorhandensein begrifflich-inhaltlicher Vorstellungen zulässt. Trotz dieser Feststellung ist der vorliegende Wortschatztest konzeptionell nicht darauf angelegt, Begriffswissen zu erfassen. Das Instrument leistet dies nach dem vorstehenden Zusammenhang implizit, untersucht es jedoch nicht explizit. Dem Wortschatztest liegt die Fragestellung zu Grunde, ob das lexikalische Element zu dem angebotenen Stimulus verfügbar ist, bewegt sich also auf der Ebene der Bezeichnungen. Dieses Verfahren, auf der Symbolebene zu arbeiten, legitimiert sich nach den dargestellten Zusammenhängen als Methode zur Erfassung lexikalisch-semantischer Fähigkeiten.

Der Begriff der ‚Fähigkeit‘ wird in der Psychologie definiert als die Gesamtheit der zur Ausbildung einer bestimmten Leistung erforderlichen Bedingungen. Das ist abzugrenzen gegenüber den Begriffen ‚Fertigkeit‘ und ‚Kenntnis‘. Fertigkeiten bilden sich auf dem Hintergrund aufgabenübergreifender, personenspezifischer Fähigkeiten durch Übung heraus. Kenntnisse sind dagegen gespeicherte und ggf. aktualisierte Wissensbestände und können z. B. mit Kenntnistests gemessen werden (vergl. Dorsch 1998).

Bezogen auf das Semiotische Dreieck, sind Fertigkeiten und Kenntnisse demnach der Ausdruck vorhandener Fähigkeiten. Wenn im Folgenden der Begriff Wortschatzkenntnis verwendet wird, bewegt man sich auf der Ebene der Bezeichnungen, was jedoch die Ebene der Fähigkeiten voraussetzt und somit einschließt. Alternativ zu den Begriffen ‚Kenntnis‘ und ‚Fähigkeit‘ wird im Folgenden stellenweise der Begriff ‚Kompetenz‘ verwendet, der nach Textor (2000, 165) als „Sachverständnis“ definiert wird. Der Begriff der Fertigkeit wird im Folgenden nicht benutzt, da der Test als passiver Wortschatztest die Kenntnis, nicht aber die aktive Anwendung von Wörtern erfasst.

2.4 Diagnostik lexikalisch-semantischer Fähigkeiten

2.4.1 Untersuchung der kommunikativ-sprachlichen Entwicklung

Prillwitz und Wudtke führen eine Entwicklungsstudie zum Hamburger Früherziehungs- und Vorschulprogramm mit lautsprachbegleitenden Gebärden durch, deren Ergebnisse in ihrer Arbeit zur kommunikativ-sprachlichen Entwicklung gehörloser Kinder bis zum Einschulungsalter (1988) dargestellt werden. Ausgangspunkt ihrer Untersuchung ist die Beobachtung der unzulänglichen Laut- und Schriftsprachkenntnisse Gehörloser und die so entstehende Frage, ob auf diesem Hintergrund auf eine feste Etablierung von Gebärden im Erziehungs- und Bildungsprozess weiter verzichtet werden kann. Ziel ist es, den Einsatz von Gebärden in der Erziehung gehörloser Kinder zu begründen und die Ergebnisse empirisch zu belegen.

Um Antwortansätze zu dieser Fragestellung entwickeln zu können, bedarf es empirischer Daten. Prillwitz / Wudtke unternehmen daher zunächst den Versuch einer Totalerhebung des Wort- / Gebärdenschatzes der betroffenen Kinder, um so eine Dokumentation der vorhandenen lexikalischen Grundlagen der Kommunikationsfähigkeit zu erhalten. Dabei wird die Tragweite dieses Untersuchungsbereiches betont, indem Prillwitz / Wudtke darauf hinweisen, dass Umfang und Qualität des sprachlichen Zeichenrepertoires zum einen über den sprachlichen Entwicklungsstand des Kindes Auskunft geben, zudem jedoch auch Rückschlüsse auf dessen soziale und kognitive Fähigkeiten zulassen. In Ermangelung geeigneter Untersuchungsinstrumente nutzen Prillwitz / Wudtke für die Totalerhebung zunächst folgende Quellen: Direkte Beobachtung, Videoaufnahmen, Tagesinterviews, Befragung und Aufzeichnung der Eltern und der Erzieher im Kindergarten. Neben dem erheblichen Aufwand, den die Nutzung dieser vielfältigen Quellen und deren Auswertung bedeuten, wird es mit zunehmendem Alter der Kinder unmöglich, den Wortschatz auf diese Weise zu erheben, da dessen Umfang anwächst und zu groß und unübersichtlich wird.

Diese Umstände nötigen die Autoren, sich der Durchführung standardisierter Tests zuzuwenden. Ihre so motivierte kritische Betrachtung bestimmter in Deutschland gängig verwendeter Wortschatztests ergibt gravierende Mängel der Instrumente. Die Feststellung dieser Mangelsituation führt hier zu der Forderung nach Entwicklung geeigneter Testinstrumente. Sie verwenden die drei gebräuchlichen Wortschatztests, den FTF-W („Frankfurter Test für Fünfjährige – Wortschatztest“ 1971), den WSS 1 („Wortschatztest für Schulanfänger“ 1970) und den PPVT („Peabody Picture Vocabulary Test“ 1969) und unterziehen die Instrumente schließlich einer kritischen Bewertung. Alle drei Tests wurden für hörende Kinder entwickelt und zur Anwendung bei Gehörlosen in Gebärdensprache übersetzt. Auf Aufbau und Konzeption der Instrumente soll hier insoweit eingegangen werden, als dies auf bestehende Mängel hinweist. Die Kritikpunkte werden von den Autorinnen jeweils an einem Beispiel

verdeutlicht. Um dabei einer einseitigen Auswahl vorzubeugen, wurde durchgehend das Item¹⁰ Nr. 23 zur Verdeutlichung herangezogen.

Der FTF-W ist ein 1971 erschienener aktiver Wortschatztest für Kinder im Vorschulalter. Er umfasst dreißig Items, bei denen es sich ausschließlich um Substantive handelt, die nach Aussage seiner Autoren "repräsentativ für die verschiedenen Lebensbereiche des Kindes sind" (FTF-W 1971, 6). Diese Lebensbereiche umfassen die Kategorien 'Spiel und Freizeit', 'Pflanzen und Tiere', 'Kleidung', 'Technik', 'Haushalt' und 'Haus und Wohnung'. Es handelt sich dabei um folgende Begriffe:

Tab. 1-A 2.4.1: Itemkorpus des FTF-W	
01. Trommel	16. (Glüh)Birne, Glühlampe
02. (Wäsche)klammer	17. Tennis-, Federballschläger
03. Schildkröte	18. Fern-, Opernglas, Fernrohr
04. ...Messer	19. (Teppich)Klopfer
05. Igel	20. ... Zelt
06. Schub-, Schiebkarre	21. (Block)Flöte
07. Hammer	22. Schraube
08. Spielkarte, (Kreuz)Ass	23. Feuerzeug
09. (Kleider-)Bügel	24. Tannenzapfen
10. Würfel	25. (Milch)Kanne
11. Reißverschluss	26. Antenne
12. ...Zange	27. ...Fass, ...Tonne
13. Streichhölzer, Str.schachtel	28. (Hack)beil, Hacke, Axt
14. (Zentral)Heizung ...	29. Schwan
15. Krawatte, Binder, Schlips	30. Korkenzieher, Bohrer

Abb. 2-A 2.4.1: Item Nr. 23 des FTF-W



¹⁰ Der Terminus ‚Item‘ wird in der vorliegenden Arbeit grundsätzlich entsprechend dem deutschen Wort ‚Testaufgabe‘ verwendet. D. h., dass der Terminus den gefragten Begriff sowie die drei diesem zugeordneten Distraktoren umfasst. Eine Ausnahme von dieser Regel stellt seine Verwendung in Kapitel B 1.3 (und einigen darauf Bezug nehmenden Stellen) dar, in dem die Auswahl der Testitems erläutert wird. Hier wird der Terminus für den gesuchten Begriff einer Testaufgabe verwendet. Aus dem jeweiligen Zusammenhang geht jedoch unzweifelhaft hervor, wann welche Verwendung eingesetzt wird. Diese Vorgehensweise wurde der nach Ansicht der Autorinnen verwirrenderen Variante vorgezogen, jedes Mal von ‚dem gefragten Begriff‘ zu sprechen.

Im Gegensatz zum FTF-W handelt es sich bei dem 1970 entwickelten WSS 1 um einen passiven Wortschatztest, der die geforderten Begriffe anhand von Vierfach-Wahlaufgaben erfragt. Der Test, ursprünglich für Schulanfänger konzipiert, eignet sich auch für Vorschüler. Bei der Wortauswahl handelt es sich, ebenso wie bei dem FTF-W, ausschließlich um Substantive:

Tab. 2-A 2.4.1: Itemkorpus des WSS 1	
01. Fass	19. Düsenflugzeug
02. Kalb	20. Gebäck
03. Adler	21. Wasserfall
04. Hummel	22. Streichinstrument
05. Verkehrszeichen	23. Gasse
06. Ananas	24. Lupe
07. Zollstock	25. Insel
08. Gefäß	26. Schneideschere
09. Lebensmittel	27. Harpune
10. Turm	28. Atlas
11. Axt	29. Birke
12. Kaulquappe	30. Elch
13. Gebäude	31. Labor
14. Nagel	32. Steinfrucht
15. Raumkapsel	33. Mohnblume
16. Schraubenzieher	34. Wolkenkratzer
17. Wappen	35. Verließ
18. Mikrofon	

Abb. 3-A 2.4.1: Item Nr. 23 des WSS 1

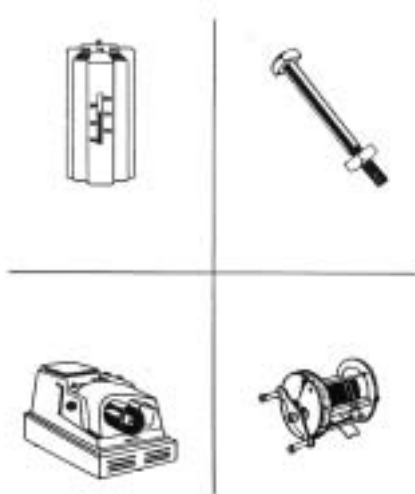


In seiner amerikanischen Originalausgabe von 1959 umfasst der Peabody Picture Vocabulary Test (PPVT) 150 Aufgaben. Bei der hier verwendeten, verkürzten Form handelt es sich um einen Bestandteil der Testbatterie für geistig behinderte Kinder (TBGB), der die Begriffe 31 bis 100 des Originals zur Anwendung bringt. Da der Test nach Schwierigkeiten gestaffelt ist, eignet er sich für eine große Altersspanne von 3;7 bis 17 Jahren. So kommen je nach Alter eine unterschiedliche Anzahl bzw. Auswahl von Begriffen zum Einsatz. Die Itemauswahl umfasst neben 49 Nomen auch 21 Verben. Die Itemauswahl lautet:

Tab. 3-A 2.4.1: Itemkorpus des PPVT

01. Nest	25. ringen	49. bejammern
02. Fahrrad	26. Beförderungsmittel	50. Feder
03. Briefumschlag	27. Küchenschränke	51. Kajak
04. pflücken	28. Feierlichkeit	52. Posten
05. Abzeichen	29. Schote	53. Furche
06. Schutzbrille	30. Schimpanse	54. Balkendecke
07. Pfau	31. umleiten	55. Bruchstück
08. Königin	32. Trichter	56. schweben
09. Kutsche	33. sich freuen	57. Trauer
10. Peitsche	34. Rednerin	58. Felsvorsprung
11. Fangnetz	35. Unterhaltung	59. Schrecken
12. Sommersprossen	36. Bogenschütze	60. untertauchen
13. Adler	37. Sportplatz	61. heruntergehen
14. gedreht	38. ausbaggern	62. Sitzkissen
15. putzen	39. angreifen	63. Hund
16. Telefonscheibe	40. Kunststück	64. untersuchen
17. gähnen	41. Gebäck	65. fischen
18. hinfallen	42. Haushaltsgerät	66. prüfen
19. Lichtsignal	43. Chemiker	67. eingeschlossen
20. hineinstecken	44. Iglu	68. gießen
21. Unterseeboot	45. Zerstörung	69. Giebelseite
22. Thermosflasche	46. Gepäckträger	70. Reptil
23. Projektionsapparat	47. Küste	
24. Gruppe	48. hochhieven	

Abb. 4-A 2.4.1: Item Nr. 23 des PPVT



Die Auflistung des Iteminventars verdeutlicht, dass sowohl der FTF-W als auch der WSS 1 ausschließlich Substantive enthalten. Der PPVT beinhaltet zusätzlich Verben, allerdings in relativ geringer Anzahl (21 von insgesamt 70 Items). In keinem der Tests wird zusätzlich die Wortart der Adjektive berücksichtigt, geschweige denn gleichgewichtet abgefragt. Diese einseitige Auswahl der Wortarten bleibt darüber hinaus von den Autoren unbegründet. Einseitige Auswahl und fehlende Begründung werden von Prillwitz und Wudtke kritisiert.

Sie stellen zudem die Qualität der Tests auf Grund der Itemauswahl in Frage. Die Test-Autoren des FTF-W beispielsweise behaupten, ihre Itemauswahl sei "repräsentativ für die verschiedenen Lebensbereiche des Kindes" (1971, 6), ohne die Gründe dafür transparent zu machen. Welche Wörter im Lebensbereich von Kindern bestimmten Alters vorkommen, lässt sich nach Ansicht der Autorinnen des vorliegenden Tests nur klären, indem entsprechende empirische Erhebungen zum aktiven Wortschatz durchgeführt werden. Deren Ergebnisse müssten, um der zitierten Behauptung zu entsprechen, der Itemauswahl eines Wortschatztests zu Grunde gelegt werden.

Die Verwendbarkeit insbesondere des FTF- W, jedoch auch der anderen beiden Tests, wird von den Autoren zudem auf Grund markanter Ähnlichkeiten vieler gegenständlicher Gebärden zu den bezeichneten Objekten in Frage gestellt. Da diesem Aspekt der Ikonizität von Gebärden in der vorliegenden Arbeit weiterführende Untersuchungen gewidmet werden, wird an dieser Stelle auf einen entsprechenden Exkurs unter B 3.6 hingewiesen.

Die Summe dieser Mängel sowie die aufgezeigten Schwierigkeiten im Untersuchungsverlauf Prillwitz' / Wudtkes (1988) verdeutlichen die entsprechende Bedarfssituation. So formulieren die Autoren (44): „Vielleicht kann die Offenlegung der vielen methodischen Probleme bei der Erhebung von laut- und gebärdensprachlichen Daten bei gehörlosen Kindern im deutschsprachigen Raum zu einer verstärkten Arbeit an der Entwicklung empirischer Erhebungsmethoden motivieren.“ Prillwitz / Wudtke erkennen im Rahmen ihres Projekts „die methodischen Grenzen, [...] wollen [aber] durch Vorlage der Daten und durch Offenlegung [...] [des] Vorgehens gerade alle Interessierten ermuntern, daran mitzuarbeiten, durch Schaffung einer empirischen Grundlage die Gehörlosenpädagogik aus unfruchtbaren bloß normativen Diskursen herauszuführen.“ (ebd., 70) Sie weisen außerdem auf Forschungsbedarf in Anschlussgebieten wie der Entwicklung des schriftsprachlichen Könnens (20), dem Sprachkönnen zum Schulabschluss und dessen Weiterentwicklung (21), Bezügen zwischen Gebärden und der Schriftsprache (21), Verwendung von Gebärden und ihre Auswirkungen (27) hin. Die Entwicklung des vorliegenden Wortschatztests versteht sich insofern auch als Basisarbeit zur Realisierung dieser u. ä. weiterführender Forschungsprojekte.

2.4.2 Testinventar

Wesentlicher Ausgangspunkt der vorliegenden Arbeit sind die Forschungsergebnisse Günthers zum Problemfeld der Diagnostik lexikalisch-semantischer Störungen. Die linguistischen sowie sprach- und entwicklungstheoretischen Grundlagen und deren Anwendung auf bestimmte Probleme in der kindlichen Sprachentwicklung sind in einer Anzahl von Veröffentlichungen zwischen 1980 und 1991 dargestellt (Günther 1980, 1987, 1989 a, 1990, 1991). Die vorliegende Untersuchung nimmt wesentlich auf den Aufsatz zum oben genannten Thema von 1991 Bezug. Hierin leitet Günther den dringenden Bedarf eines "diagnostischen Instrumentes zur Erfassung des Niveaus lexikalisch-semantischer Fähigkeiten im Bereich der referenzsemantischen Wortarten" (1991, 187) ab. Darüber hinaus werden grundsätzliche Merkmale zur Konzeption eines solchen Instrumentes aufgestellt.

Günther nimmt bei der Erarbeitung der linguistischen und sprach- und entwicklungs-theoretischen Theorie maßgeblich Bezug auf die Tradition Piagets (1969), auf die von Furth (1972) und Szagun (1983; 1986) vorgelegten Arbeiten, weiterhin auf die Wort-listen von Augst (1984, 1989), Pregel / Rickheit (1987) und Ruoff (1981) und ferner auf Texte zum Thema Wortschatz (z. B. Stern/Stern 1971; 1975). Diese theoretische Basis mündet in die Vorlage konkreter Handlungsanweisungen für die diagnostische und damit pädagogische Praxis.

Die vorliegende Arbeit knüpft an diesen Punkt des wissenschaftlichen Erkenntnisstandes an, indem sie das angezeigte Testinstrument entwickelt und darstellt. Dabei wird der in der Basisarbeit Günthers aufgestellte Merkmalskatalog im Detail erweitert, z. T. modifiziert und differenziert. Die Autorinnen führten die Testentwicklung und die dazu durchgeführten Untersuchungen durch auf der Grundlage der einschlägigen Fachliteratur zu den Themenbereichen Testkonstruktion, Testdiagnostik und pädagogische Diagnostik (Lienert 1994; Rauchfleisch 1994; Mietzel 1993; Kleber 1976; Ingenkamp 1972 a) sowie Statistik (Bortz 1993; Ingenkamp 1972 b, Lienert 1994; SPSS 1996).

Wesentliches und vollständig neues Element des vorliegenden Testinstruments ist seine Konstruktion in den drei Sprachmodalitäten Lautsprache, Schriftsprache und Gebärdensprache. Diese Auswahl der Sprachmodalitäten stellt das zentral unterscheidende Merkmal des vorliegenden Wortschatztests im Vergleich zu den für Hörende entwickelten Instrumenten dar, welche ausschließlich mit der Lautsprache operieren. Zu dem Kommunikationsspektrum Gehörloser gehört jedoch neben der (meist nur fragmentarisch verfügbaren) Lautsprache vorwiegend die (ihnen relativ leicht zugängliche) DGS. (Dabei sei die Gewichtung der Modalitäten in den erzieherischen und privaten Bezügen des Einzelnen an dieser Stelle dahingestellt.) Zudem nimmt die Schriftsprache für sie besondere Bedeutung ein, da sie geeignet ist, als Brücke zur kommunikativen Welt Hörender zu fungieren und im Sprachlernprozess eine vermittelnde Funktion zur Lautsprache einzunehmen. Die Analyse lexikalischer Fähigkeiten

Gehörloser muss daher alle drei Modalitäten, Laut-, Schrift- und Gebärdensprache einbeziehen, um den Betroffenen gerecht zu werden. Demnach müssen einem Wortschatztest für Gehörlose die drei Modalitäten konzeptionell zu Grunde liegen.

Ein sich ausschließlich auf die deutsche Lautsprache beziehendes Testergebnis kann nicht für sich in Anspruch nehmen, die lexikalisch-semanticen Fähigkeiten der gehörlosen Testperson¹¹ abzubilden, sondern könnte allenfalls eine Aussage zulassen über deren Fähigkeiten im Bereich der Lautsprache. Selbst dies ist jedoch stark zu bezweifeln, da weitere konzeptionelle Elemente von für Hörende konzipierten Tests dem im Wege stehen. So wurde bei der Itemauswahl selbstverständlich nicht auf (z. T. gravierende) Schwierigkeiten beim Ablesen von den Lippen geachtet. Bei Nichtberücksichtigung dieses Elementes jedoch werden betreffende Items von gehörlosen Testpersonen häufig auf Grund solcher Ableseschwierigkeiten fehlerhaft beantwortet und nicht auf Grund fehlender Begriffskennntnis. Dieser Umstand verstellt in erheblichem Maße den Blick auf den zu untersuchenden Wortschatz. Es muss demnach angenommen werden, dass für hörende Testpersonen entwickelte Wortschatztests zunächst für Gehörlose nur partielle Informationen bereitstellen, dass darüber hinaus jedoch die so bereitgestellten Informa-tionen nur begrenzt valide sind, weil die genannten spezifischen Bedingungen Gehörloser unberücksichtigt bleiben.

Durch die Einbeziehung aller drei Modalitäten bekommen Gehörlose somit erstmals die Möglichkeit, ihre Kompetenzen überhaupt zu zeigen. Zusätzlich wird durch die Betrachtung aller drei Segmente eine differenzierte Einschätzung der lexikalischen Kompetenzen des Einzelnen ermöglicht. Auf Grund der so bereitgestellten Ergebnisse kann in einem folgenden Schritt untersucht werden, ob Rückschlüsse auf bestimmte methodische Rahmenbedingungen gemacht werden können. So werden in Teil C zunächst die Ergebnisse der Pilotstudie und in einem weiteren Schritt die Ergebnisse der Hauptuntersuchung unter pädagogischen Fragestellungen ausgewertet, die für die Einschätzung der Kompetenzen gehörloser Schüler sowie für die Bewertung gehörlosenpädagogischer Methodik relevant sind. Das vorliegende Testinstrument lässt zudem Vergleiche mit den lexikalisch-semanticen Leistungen nicht-behinderter Kinder und von Kindern mit anderen Behinderungen zu. Diese Vielfalt setzt die Berücksichtigung der jeweiligen Bedingungen der einzelnen Sprachmodalitäten voraus, was die Testentwicklung entscheidend beeinflusst. Die Autorinnen konzentrieren sich bei der vorliegenden Testversion auf die spezifischen Umstände Gehörloser, was sich insbesondere mit dem angezeigten dringenden Handlungsbedarf auf diesem Gebiet begründet. Die Modi-

¹¹ In Abgrenzung gegen die ehemals übliche Verwendung des Terminus ‚Proband‘, die auf der Annahme des Menschen als einem eher passiven Subjekt fußt, wird in vorliegender Arbeit durchgängig der Terminus ‚Testperson‘ verwendet. Damit wird dem durch die Aktions- bzw. Handlungsforschung neu entwickelten Wahrheitsbegriff und dessen Folgen Rechnung getragen. Vertiefende Ausführungen zu diesem Themenkomplex finden sich unter A 2.5 Testdiagnostik.

fizierung oder Erweiterung des Instruments für Testpersonen anderer Behinderungen steht noch aus und sollte in der Zukunft als Anschlussprojekt bearbeitet werden.

Günthers eingehende Betrachtung von Wortschatzuntersuchungen mit Kindern im Vorschulalter (Günther 1991) zeigt Differenzen zwischen unterschiedlichen Untersuchungen vergleichbaren Anspruchs auf. Diese "... verweisen [...] auf bislang kaum reflektierte Probleme bei der Wortschatzerfassung [...] [und indizieren] eine stärkere Beachtung der Datenerfassungsmethoden und der individuellen Entwicklungsbedingungen ..." (ebd. S. 170). Günther erachtet es daher als erforderlich, die Wortschatzbildung von Vorschulkindern neu zu überprüfen. Dabei stellt er als zentral relevant heraus, nunmehr nicht ausschließlich quantitative, sondern zusätzlich qualitative Aspekte zu berücksichtigen. Dies wäre jedoch nur anhand neuer Korpora möglich, welche entsprechend differenzielle Angaben enthielten.

"Im Gegensatz zu der Datenlage bei den Ein- bis Fünfjährigen verfügen wir inzwischen für die älteren Vorschulkinder, Grundschulkinder, Jugendlichen und Erwachsenen über eine Reihe von detaillierten Korpora mit wichtigen, für diagnostische Zwecke relevanten Daten über Entwicklung und Ausdifferenzierung des Lexikons." Diese verdeutlichen, dass bei lexikalischen Entwicklungsstörungen die Messung des Wortschatzumfangs "...nicht mehr mit einfachen, direkten Überprüfungsverfahren möglich ist. Vielmehr sind über reine Ranghäufigkeitslisten hinaus differenzielle morphologische und soziolinguistische Angaben für die Auswahl der lexikalischen Prüfitems notwendig." (alles Günther 1991, 171) Unter Berücksichtigung dieser Notwendigkeit stellt Günther entsprechend umfangreiche und bezüglich sozial-personaler und linguistischer Merkmale differenzierte Korpora zu einer Wortliste zusammen. Dabei finden die Erhebungen von Ruoff (1981) für Erwachsene, von Augst (1984) für sechsjährige Vorschulkinder und von Pregel/Rickheit (1987) für Grundschulkinder Eingang.¹² Diese Wortliste bildet die Datenbasis für die Auswahl, Anzahl und Wortartengewichtung der lexikalischen Prüfitems des hier vorliegenden konzeptionell neuen diagnostischen Instrumentes. Die Wortliste stellt eine Synopse der in den angegebenen Sprachkorpora vorfindlichen Rang- und Vorkommenshäufigkeiten dar. Dabei betont Günther, dass die Ranghäufigkeitsorientierung für die Auswahl der Items eben nicht isoliert als Begründungshintergrund stehen darf, sondern notwendig der Ergänzung der von ihm vorgenommenen sprachtheoretisch und entwicklungspsychologisch begründeten Kriterien bedarf.

Bezüglich der Auswahl der Wortarten stellt Günther fest, dass eine Beschränkung auf die drei referenzsemantischen Hauptwortarten - Nomen, Verben, Adjektive (+ Inhalts-Adver-

¹² Wie dem Durchführungsteil vorliegender Arbeit zu entnehmen ist, wurde das Testinstrument mit Schülern der dritten, vierten und Anfang der fünften Klassenstufe angewendet. Unter Bezugnahme auf die Ableitung der Items aus der benannten Wortliste, wird die Eignung vorliegenden Tests für das Grundschulalter, d. h. die Einbeziehung auch der ersten und zweiten Klasse, angegeben, da für die Untersuchungen zur Erstellung der Wortliste ja mit breitem Hintergrund Grundschüler aller Klassenstufen einbezogen wurden.

bien) – sachgerecht ist. Als Gründe dafür führt er an, dass einerseits die Funktionswortarten – Artikel, Pronomen, Präpositionen, Konjunktionen – zusammengekommen nur etwa 300 im Deutschen Wortschatz insgesamt ausmachen, dass andererseits die bisherigen Untersuchungen und Arbeiten zur Semantik und deren Entwicklung primär von den kognitiven Konstruktionsprozessen und Strukturen ausgehen und die eigentliche sprachliche Seite nur sehr allgemein behandeln. Dem wirkt Günther entgegen, indem er die Wortarten differenziert unter linguistischen Aspekten aufeinander bezieht. Das Ergebnis einer derartigen Betrachtung der Wortarten auf dem Hintergrund seiner semantischen Theorie führt für das neu zu erarbeitende Testinstrument zu der Auswahl der drei referenzsemantischen Wortarten.

Die Erfassung lexikalisch-semantischer Störungen mit deutschsprachigen Wortschatz- und Bedeutungstests ist auf Grund gravierender Mängel der Testinstrumente nach Günther nicht oder nur bedingt möglich. Günther schlussfolgert dies aus seiner Untersuchung der wichtigsten, im deutschsprachigen Raum verwendeten Tests. Dies sind der „Frankfurter Test für Fünfjährige“ (FTF-W), der „Wortschatztest für Schulanfänger“ (WSS1), der „Aktive Wortschatztest für drei- bis sechsjährige Kinder“ (AWST) sowie der „Peabody Picture Vocabulary Test“ (PPVT). Auf die Ergebnisse dieser Untersuchung soll hier kurz eingegangen werden.

So gilt für alle der untersuchten Prüfverfahren, dass „... im Vergleich zum tatsächlichen Sprachgebrauch und zum deutschen Wortschatz Nomen unangemessen über-, Verben und/oder Adjektive in der Regel unterrepräsentiert, z. T. überhaupt nicht berücksichtigt sind“ (Günther 1991, 180). Günther zeigt auf, dass Nomen (1991, 175) „im Gegensatz zur gängigen diagnostischen Praxis [...] besonders ungeeignet zur Erfassung lexikalisch-semantischer Fähigkeiten und eventueller Störungen sind, [...]“. Er arbeitet zudem den Nutzen der Verwendung von Verben und Adjektiven zu diagnostischen Zwecken heraus, und dies besonders unter Berücksichtigung der Informationsgewinnung im semantischen Bereich. Aus diesen Überlegungen ergibt sich für das vorliegende Instrument die Vorgabe, die drei referenzsemantischen Wortarten in jeweils gleich großen Anteilen in den Itemkorpus aufzunehmen.

Für den FTF-W und den WSS 1 wird neben der Reduktion auf Nomen kritisch bemerkt, dass die Items nicht repräsentativ und die Hinweise auf Fördermaßnahmen, die sich aus den Testergebnissen ableiten lassen, mangelhaft sind. Diese Ansicht wird ergänzt durch die Beobachtung, dass die Autoren der im deutschen Sprachraum gängigen Tests zur Erfassung des (aktiven oder passiven) Wortschatzes sowie des begrifflich/bedeutungsmäßigen Fähigkeitsniveaus weder die Auswahl noch die Reihenfolge der Items mit strukturellen und/oder entwicklungstheoretischen Daten und Erfahrungen aus der Linguistik oder auch der Kognitionsforschung begründen. Die Ökonomie und Praxisfreundlichkeit der Tests wird von Günther angesichts der nachgewiesenen Mängel als völlig irrelevant bezeichnet. Günther bevorzugt

den AWST 3-6 gegenüber den beiden zuvor genannten Verfahren zwar, widerlegt jedoch dessen differenzialdiagnostischen Anspruch auf Grund seiner gravierenden Mängel.

Der Peabody Picture Vocabulary Test (PPVT) kann als Klassiker und Vater aller passiven Wortschatztests bezeichnet werden. Da er bereits 1936 entwickelt wurde ist seine Bewertung als stark veraltetes Instrument aus heutiger Sicht unvermeidbar. Zudem stammt er aus dem Amerikanischen, was zusätzliche Probleme aufwirft. Die Methode der Vierfeldertafeln des PPVT wird als gut geeignet bewertet und auch für die vorgeschlagene Konzeption des neuen Testinstrumentes übernommen. Die fehlende Begründung der Itemauswahl sowie der Zusammenstellung der Vierfeldertafeln jedoch disqualifizieren das Verfahren. Günther hält zudem den Anspruch, mit einem methodisch gleich bleibenden Verfahren das Wortschatzniveau von Zweieinhalb- bis Siebzehnjährigen zu erfassen, für besonders fragwürdig.

Das Verfahren direkter Befragung nach der Bedeutung der Testwörter des HAWIK-Wortschatztests hält Günther für eine metalinguistische Anforderung. Eine solche Anforderung ist vor allem für Kinder, die sprachlich-kognitive Probleme mitbringen, als Überforderung zu bewerten. In Folge einer solchen Überforderung bleibt ihr tatsächliches Wort- und Bedeutungswissen verborgen. In Abgrenzung dagegen wird vorliegend einem passiven Verfahren der Vorzug gegeben. Dabei folgt auf die kurze Vorgabe z. B. eines Objektes die Wahl einer der Darstellungen auf der vorliegenden Tafel durch die Testperson. Angeführte Bewertungsbeispiele des HAWIK-R verdeutlichen "nicht nur, dass die in den Aufgaben enthaltenen sprachlichen Informationen ignoriert werden, sondern auch, dass sie fast ausnahmslos auf eine elaborierte Sprachkompetenz der Untersuchungspersonen hinweisen, sodass die Fehlerbewertung geradezu im Widerspruch zu den diagnostischen Zielen des Tests stehen" (184). Zudem bleibt auch bei dem HAWIK-R die Itemauswahl und -reihenfolge unbegründet, sodass das Verfahren von Günther als völlig ungeeignet zur Erfassung des kindlichen Wortschatzes bewertet wird.

Neben den benannten inhaltlichen Mängeln bestehen nach Ansicht der Autorinnen des vorliegenden Tests gravierende Mängel in der Gestaltung der drei angeführten Tests. Die Testbilder sind zum einen nicht ansprechend oder gar anregend, zum anderen nicht mehr zeitgemäß. Beides steht in Beziehung zueinander, da unzeitgemäße Darstellungen den Erlebnishintergrund der Testpersonen nicht widerspiegeln und somit nicht ansprechend sein können.

Die Darstellungen sind schwarz-weiß gehaltene Strichzeichnungen. Besonders die Bilder des WSS1 transportieren keinerlei Plastizität oder Lebendigkeit. Die veraltete Gestaltung wird z. B. deutlich anhand des Testitems Nr. 23 des PPVT: 'Projektionsapparat'. Ein Gerät des dargestellten Modells findet sich heute nicht mehr auf dem Markt. Es muss daher davon ausge-

gangen werden, dass es Testpersonen heute nicht bekannt ist. Auch der Begriff ist überholt; der Apparat wird heute als Diaprojektor bezeichnet.

Nach Ansicht der Autorinnen muss Testmaterial Kinder dazu anregen, motiviert mitzuarbeiten. Seine zeichnerische Gestaltung muss daher motivierende Elemente transportieren. Denn eine negativ beeinflusste Einstellung der Testperson zum Testmaterial ist wahrscheinlich, wenn es nicht kindgerecht auf diejenigen zugeschnitten ist, die es bearbeiten. Zudem steht außer Frage, dass die Testpersonen bei der Identifizierung des Dargestellten die Möglichkeit haben müssen, sich auf Bekanntes aus ihrer Erlebniswelt zu beziehen.

Im Gegensatz zu dieser Forderung spiegeln die Zeichnungen der untersuchten Tests z. T. regionale Besonderheiten wider, deren Kenntnis nicht bei jedem deutschen Kind vorausgesetzt werden kann. Dabei handelt es sich zum einen um Besonderheiten deutscher Dialekte, wie z. B. dem Item 'Gasse' (WSS1 Nr. 23), welches vornehmlich in Süddeutschland verwendet wird und dessen Kenntnis sich einem norddeutschen Kind möglicherweise entzieht. Zum anderen finden sich spezifisch ausländische Darstellungen, wie beispielsweise die des Distraktors 'Thermometer' in dem amerikanischen PPVT (zu Item Nr. 23), dessen Skala Fahrenheit anzeigt. Die Auswahl sowohl der Testitems als auch der Distraktoren muss nach Ansicht der Autorinnen regionale Besonderheiten ausschließen. Soll ein aus dem Ausland stammender Test verwendet werden, wäre dies nur vertretbar, wenn die Darstellungen einzelner Items ggf. angeglichen würden.

Der FTF-W weist Mängel im gestalterischen Aufbau auf. Auf einer Seite im Testheft, die der Testperson vollständig vorliegt, befinden sich sieben Bilderreihen mit je vier Zeichnungen, sodass sich bei der Lösung einer Aufgabe noch sechs weitere Aufgaben im Blickfeld des Kindes befinden. Dabei sind die einzelnen Bilder sehr klein und direkt hintereinander aufgereiht, ohne Zwischenräume einzuhalten. Der optische Eindruck, der sich so ergibt, ist annähernd der einer einzigen Zeichnung. Das Material ist dadurch unübersichtlich und erschwert der Testperson die Konzentration. Nach Ansicht der Autorinnen ist es hingegen erforderlich, dass ein Testheft übersichtlich gestaltet ist, sodass eine gezielte Bearbeitung jeder einzelnen Aufgabe ohne optische Ablenkung ermöglicht wird.

2.4.3 Folgerungen

Bei der dem vorliegenden Projekt der Entwicklung eines Wortschatztests für gehörlose Grundschulkinder zu Grunde liegenden Forderung, die Ergebnisse der dargestellten Untersuchungen konzeptionell zu berücksichtigen, handelt es sich demnach nicht um den Ruf nach marginalen Verbesserungen, sondern nach grundsätzlicher Eignung. Die Analyse der Mängel bestehender Tests macht deutlich, dass bei der Beurteilung der Fähigkeiten Gehörloser im

Bereich lexikalisch-semantischer Fähigkeiten aus den genannten und weiteren Gründen (s. u.) mit Testergebnissen operiert wird, die den Testpersonen nicht gerecht werden. Dabei muss nach den dargestellten Zusammenhängen davon ausgegangen werden, dass weiterreichende Fähigkeiten vorhanden sein könnten, als bisher diagnostizierbar war.

Gegenwärtig werden zur Diagnostik lexikalisch-semantischer Fähigkeiten Gehörloser Wortschatztests eingesetzt, die für Hörende entwickelt wurden. Der Einsatz verfügbarer Prüfverfahren bei gehörlosen Testpersonen weist neben den aufgezeigten Mängeln die grundsätzliche Problematik auf, dass gehörlosenspezifische Bedingungen konzeptionell unberücksichtigt gelassen werden und so die Instrumente den betroffenen Testpersonen nicht gerecht werden können. Da gegenwärtig kein den spezifischen Bedingungen Gehörloser angepasstes Testinstrument zur Erhebung der erwünschten Daten zur Verfügung steht, macht sich die vorliegende Untersuchung dessen Entwicklung zur Aufgabe.

Das Testinstrument ist gekennzeichnet durch eine Reihe sich aus der Mängelanalyse bestehender Tests auf der einen und der Berücksichtigung gehörlosenspezifischer Bedingungsgefüge auf der anderen Seite ergebender konzeptioneller Säulen. Dazu zählen die Auswahl und gleichgewichtete Einbeziehung der drei Sprachmodalitäten Laut-, Schrift- und Gebärdensprache, die Auswahl und gleichgewichtete Einbeziehung der drei Wortarten Nomen, Verben und Adjektive, eine auf verschiedenen theoretischen Hintergründen reflektierte Auswahl der Testitems, die Wahl des Aufgabentyps, gestalterische Aspekte sowie die Berücksichtigung gehörlosenspezifischer Bedingungen der Sprachrezeption. Auf jede dieser inhaltlich-konzeptionellen Säulen, die in diesem Kapitel grundsätzlich hergeleitet wurden, wird im Rahmen der Darstellung der Grundkonzeption des Tests¹³ vertieft eingegangen werden.

2.5 Testdiagnostik

Einer pädagogischen Diagnostik liegt immer ein bestimmtes Menschenbild und liegen bestimmte Annahmen darüber zu Grunde, welche Art von Informationsgewinn durch sie ermöglicht wird. Bis in die 60er Jahre wurde empirische Forschung mit dem Mittel klassischer Testdiagnostik als dem traditionellen Handwerkszeug der Sozialwissenschaften praktiziert. Dieser Praxis lag die Annahme einer eher passiven Natur des Menschen zu Grunde, dessen individuelle Persönlichkeitsstrukturen sein Verhalten in vorhersagbarer Weise beeinflussen. Dementsprechend wurde davon ausgegangen, mit Hilfe von Testinstrumenten 'objektive' Aussagen treffen zu können, die als Grundlage weiterführender, z. T. existenziell bedeutsamer Entscheidungen dienten. Dieser 'Objektivismus' der konventionellen empirischen For-

¹³ Unter B 1.

schung wurde jedoch zunehmend kritisch hinterfragt. Titel wie die 1978 veröffentlichte detaillierte Auseinandersetzung von Grubitzsch und Rexilius mit der klassischen Testtheorie und ihren Schwächen dokumentieren den Höhepunkt einer fundamentalen Ablehnung dieser Methoden.

In Hinblick auf das den Sozialwissenschaften zu Grunde liegende Menschenbild fand so eine grundlegende Wende statt. In der Folgezeit nahmen die Ideen der sog. 'Handlungspsychologie' eine führende Position ein. Das Individuum wurde nicht länger deterministisch und passiv, sondern aktiv, konstruktiv und eigenverantwortlich in Hinblick auf seine Realität gesehen. Diese Sichtweise hatte Konsequenzen für die Diagnostik. So brachten beispielsweise (wie oben bereits erwähnt und für vorliegendes Projekt assimiliert) die herkömmlich verwendeten Termini 'Proband' und 'Testleiter' überholte Grundeinstellungen zum Ausdruck. Dem 'Probanden' wurde nun vielmehr als kompetenten und autonom handelnden Kooperationspartner gegenübergetreten. Da der Diagnostiker dabei versucht, sich in die Bedingungen seines Gegenübers hineinzusetzen und zu verstehen, statt objektive Fakten zu rekonstruieren, begibt er sich bewusst auf eine Ebene mit ihm. Diese neue handlungspsychologisch geprägte Orientierung maß der subjektiven Einschätzung der Testanwender größere Bedeutung bei. Jedoch mussten sie sich nun ihrerseits die kritische Frage nach dem Wahrheitsgehalt ihrer Einschätzung gefallen lassen.

Der überdeutliche Pendelausschlag fundamentaler Kritik an der sog. 'klassischen Testdiagnostik' wurde nach erreichtem Vollzug eines grundsätzlichen Umdenkens wieder in die entgegengesetzte Richtung korrigiert. Dies bedeutet, dass dem Erhalt sog. 'objektiver' Informationen über klassische Verfahren heute wieder Bedeutung beigemessen wird. Dies jedoch nur unter folgenden Bedingungen: Die Verfahren müssen differenziert und auf der Basis begründeter Annahmen entwickelt sein und die erhaltenen Informationen müssen in ihrer Tragweite realistisch eingeschätzt werden (d. h. sie sollten beispielsweise nicht zur alleinigen Grundlage schwerwiegender Entscheidungen herangezogen werden).

Entscheidende Folge dieses Umdenkens ist (u. a.) die Forderung nach qualitativer Bewertung quantitativ vorliegender Informationen. Dies bedeutet die Umsetzung eines differenzialdiagnostischen Ansatzes im Sinne einer persönlichen Einschätzung jeder einzelnen Testperson. Dies geschieht im Sinne der Standardisierung von Tests unter Beibehaltung klassischer Methodik. So zeichnet Schuck (1997, 16) in Bezug auf sein Forschungsprojekt über Sprachentwicklungsstörungen den Weg vor: „Wir folgen [mit einem Altersvergleich A. d. V.] im ersten Schritt des Projektes den herrschenden 'praktischen' Gepflogenheiten und der klassischen Logik der normorientierten Diagnostik, um in der Folge den Blick auf die zu Grunde liegenden individuellen Konstruktionsprozesse zu weiten.“

In diesem Zusammenhang beeinflussten die Ideen der sog. ‚strukturorientierten Diagnostik‘ nach Probst (1981, 1982, 1992) das vorliegende Projekt. Dieser bereits in den siebziger Jahren entwickelte und Anfang der achtziger Jahre differenziert ausgeführte Ansatz wird von Probst noch in den neunziger Jahren als „neuer methodischer Versuch“ (1992, 169) bezeichnet. Projekte, die die theoretischen Elemente des Ansatzes praktisch umzusetzen versuchen, bewegen sich demnach noch immer auf Neuland. Unter den bislang zur Verfügung stehenden Möglichkeiten pädagogischer Diagnostik ist die strukturorientierte Diagnostik neben ihrer relativen Neuartigkeit und Aktualität insbesondere für den sonderpädagogischen Bereich wegweisend.

Als einen der Ausgangspunkte für die Entwicklung der strukturorientierten Diagnostik schildert Probst den „Bedarf nach Diagnoseverfahren, die die individuellen Ausgangslagen für zu planende Lernprozesse bei einem heterogenen Schülerpublikum anzuzeigen vermögen“ (1992, 176). Die Heterogenität unter Schülern innerhalb einer Lerngruppe gilt heute auch im Regelschulbereich in ausgeprägtem Maße, im Sonderschulbereich jedoch in besonderer Weise. Insofern ist dies für das vorliegende Projekt der Entwicklung eines diagnostischen Instruments für Gehörlose zutreffend.

Probst sieht die Herausforderung in der Entwicklung qualitativer Strukturprinzipien bei standardisierten Verfahren. Dabei betont er die Notwendigkeit, sich bei der Konstruktion von Testinstrumenten stärker der Struktur der Inhalte zu widmen und diese konzeptionell zu berücksichtigen.

Ohne im vorliegenden Fall dem Modell nach Probst in den Details der konstruktiven Merkmale zu entsprechen, die ein qualitatives Testinstrument seines Erachtens aufweisen sollte, werden wesentliche Kernideen umzusetzen versucht. Grundsätzlich wird die Methodenfolge Schucks praktiziert, indem das Testinstrument zunächst nach den Regeln der klassischen Testdiagnostik entwickelt und anschließend versucht wird, dies durch qualitative Bewertungen der quantitativ vorliegenden Ergebnisse zu ergänzen. Dabei werden gleichwohl wesentliche Merkmale der strukturorientierten qualitativen Diagnostik konzeptionell, d. h. bereits in der Testkonstruktionsphase, eingearbeitet. Dies bildet den Hintergrund für eine Erweiterung und damit Verbesserung einseitig quantitativer Verfahren sowie für differenzierte Auswertungsansätze im Sinne qualitativer Orientierung. Demnach wird festgestellt, dass sich der vorliegende Test als quantitatives Verfahren mit wesentlichen strukturellen Merkmalen qualitativer Ansätze einordnen lässt.

Diese Einordnung stimmt mit der Positionsbeschreibung Wolfs (1995) überein, der herausarbeitet, dass in der Diskussion qualitative versus quantitative Forschung der Umstand wesentlicher Überschneidungen zwischen beiden beachtet werden sollte. Wolf stellt fest, dass

insbesondere im Rahmen empirischer sozialwissenschaftlicher Forschung i. d. R. Entscheidungspotenzial bzgl. der Methoden besteht. Auf diesem Hintergrund ergibt sich die Notwendigkeit, nach getroffener Entscheidung für den Leser oder Zuhörer Verfahrensdurchsichtigkeit zu schaffen. Dies soll hiermit geschehen.

In Bezug auf diesen Ansatz qualitativer Bewertung von Testergebnissen versteht sich der vorliegende Test als Pilotprojekt. Die Maßstäbe einer entsprechenden Ergebnisauswertung liegen nicht vor, sondern müssen entwickelt werden. In der vorliegenden Arbeit wird als Kernbereich qualitativer Bemühungen eine zweigliedrige Analyse der Daten vorgenommen. Zum einen wurde ein Bewertungssystem erstellt, in dem der klassischen Testtheorie gemäß ein Item richtig oder falsch beantwortet und entsprechend berechnet wurde. Dies ermöglicht die Lokalisation der Schwächen der einzelnen Testperson in den unterschiedlichen Sprachmodalitäten und Wortarten. Zum anderen wurde ein Distraktorenprofil entwickelt, welches der inhaltlichen Betrachtung vorkommender Fehler zu Grunde gelegt wird (vergl. B 3.10). Die differenzierte Betrachtung der fehlerhaft beantworteten Aufgaben soll durch eine Konzeption qualitativer Abstufung der Fehlermöglichkeiten gewährleistet werden.

Indem neben dem in einer Normierungs-Tabelle abgelesenen Wert auch die Struktur der fehlerhaften Antworten einer Testperson Beachtung findet, soll ein differenzierterer und mehrschichtiger Blick auf die Leistung einer Testperson ermöglicht werden. Dies soll letztlich Aufschluss geben über das Leistungs- bzw. Fehlerprofil des einzelnen Kindes und somit gleichzeitig den gezielten Einsatz von Fördermaßnahmen ermöglichen.

Insofern realisiert das Projekt verschiedene Kerngedanken Probsts. So fußt der Test (wie unten näher erläutert werden wird) auf einer empirisch gesicherten theoretischen Basis zum Thema Begriffsbildung und Wortschatz und es wird in der Itemauswahl Bezug genommen auf die Ergebnisse empirischer Wortschatz-Untersuchungen (Wortlisten). Damit ist der Forderung Probsts, sich der Struktur der Testinhalte eingehend zu widmen, statt Konstrukt und Itemauswahl unreflektiert-emotional zustande kommen zu lassen, nachgekommen worden. Zudem wird mit der Entwicklung des Distraktorenprofils der konzeptionelle Grundstein dafür gelegt, die Testergebnisse schließlich auch auf die ihnen zugrunde liegenden Konstruktionsprozesse hin zu untersuchen.

3. Zielsetzung und Zielentwicklung

Ziel der Untersuchung ist die Entwicklung und Erprobung eines Wortschatztests für gehörlose Kinder im dritten und vierten Schuljahr, dessen Ergebnisse Eingang finden sollen sowohl in die pädagogische Praxis als auch in die wissenschaftliche Diskussion.

Wenn Sprachvermittlung beim gehörlosen Kind im Kern Hilfe zur Lebensbewältigung sein soll und Vertreter unterschiedlicher methodischer Ansätze übereinstimmend davon ausgehen, dass dieses Ziel allein erreicht werden kann, wenn den Betroffenen ein Kommunikationssystem zur Verfügung steht, dann kristallisiert sich zur Bewertung und Weiterentwicklung der Ansätze als vorrangige Aufgabe die Untersuchung sprachlicher Fähigkeiten heraus.

Das Testinstrument ist konzeptionell darauf angelegt, ein differenziertes Bild der einzelnen Testperson unter dem Aspekt Wortschatz bzw. semantisch-lexikalischer Kompetenz zu erhalten. Dies erfolgt unter anderem, indem drei Sprachmodalitäten, drei Wortarten und qualitative Auswertungsmaßstäbe Berücksichtigung finden. Dadurch kann ein Bild gewonnen werden von den Leistungen der jeweiligen Testperson auf den verschiedenen sprachlichen Sektoren.

Auf diese Weise erfüllt der Test verschiedene Funktionen: Die Testergebnisse lassen ein Urteil zu über die Wirksamkeit der bei den betreffenden Testpersonen angewandten Methode der Sprachvermittlung. Die Testergebnisse stellen eine probate Grundlage für die Weiterentwicklung der einzelnen Ansätze sowie für die wissenschaftliche Bearbeitung angrenzender Fragestellungen dar. Der Test ermöglicht den gezielten Einsatz pädagogischer Fördermittel durch den differenzierten Blick darauf, wie die Kompetenzen der jeweiligen Testpersonen gelagert sind.

Das Erreichen dieser Zielmarken wird (im Auswertungsteil C) anhand verschiedener Annahmen, die unterschiedliche Teilbereiche beleuchten, überprüft. Diese Überprüfungen ermöglichen sowohl eine pädagogische als auch eine testtheoretische Auswertung bzw. Bewertung. So wird angenommen, dass ein Schwierigkeitsgefälle unter den Modalitäten bestehen wird. Die Gebärdensprache wird, so die Annahme, am leistungsstärksten ausfallen, gefolgt von Schrift- und schließlich der Lautsprache. Dabei werden im Subtest Gebärdensprache die Ergebnisse sämtlicher Schüler, unabhängig von der im schulischen Rahmen angewendeten Sprachvermittlungsmethode, am besten ausfallen. Darüber hinaus wird angenommen, dass die bilingual erzogenen Schüler in allen Subtests am leistungsstärksten abschneiden. Zudem wird überprüft werden, ob Korrelationen mit bestimmten Sozialisationsbedingungen festgestellt werden können. Anhand der Ergebnisse können die Objektivität der Testergebnisse sowie deren Aussagekraft überprüft werden. Schließlich wird anhand umfang-

reicher Untersuchungen der Distraktoren bzw. möglicher Wahlstrategien der Testpersonen überprüft werden, ob die konzeptionell angelegte qualitative Abstufung der Distraktoren Bestand hat.

Diese zu überprüfenden Annahmen seien damit kurz umrissen. Detailliertere Ausführungen sollen an dieser Stelle jedoch nicht erfolgen, da die Zusammenhänge teilweise erst in Teil B erarbeitet und in Teil C im Einzelnen jeweils eingehend dargestellt werden. Der Grund dafür, dass die Fragestellungen nicht zu Beginn präzise gefasst, daraufhin die entsprechende Untersuchung durchgeführt und schließlich die Ergebnisse zur Beantwortung der Fragestellungen ausgewertet werden, liegt in der Sache des Projekts einer Testentwicklung. So bestehen, wie in Teil A grundlegend hergeleitet, zentrale Eckpfeiler des Instrumentes von vornherein. Im Verlauf der Erarbeitung des Instrumentes werden nun Elemente präzisiert oder erst hinzugefügt. Durch diesen Prozess entstehen immer neue und partiell veränderte Szenarien, innerhalb derer sich angedachte mit dem Instrument zu untersuchende Fragestellungen konkretisieren oder erst ergeben. Dies gilt in besonderem Maße für den Bereich der qualitativen Ergebnisauswertung. Die Konzeption der qualitativ abgestuften Distraktoren bspw. wird erst im Verlauf der Arbeit vollständig entfaltet, obgleich sie bereits in frühen Abschnitten grundsätzlich behandelt wird. In diesem Sinne folgt die Niederschrift den realen Durchführungsbedingungen des Forschungsprojekts.

B. Testaufbau und Durchführung der Untersuchung

1. Grundkonzeption des Tests

1.1 Auswahl der Sprachmodalitäten

Prillwitz unternimmt bereits 1975 den Versuch einer Funktionsbestimmung der beiden Sprachmodalitäten Laut- und Gebärdensprache (in: Claussen u. a. 1975). Ausgangspunkt seiner Überlegungen bildet die Annahme, dass ein unvollständiger, z. T. gar veränderter (Laut-)spracherwerb Gehörloser, wie er vielfach festzustellen ist, negative Auswirkungen auf deren soziale, kognitive und affektive Entwicklung zur Folge hat. Dies könnte nach Prillwitz größtenteils vermieden werden, wenn die Gehörlosen zugängliche Gebärdensprache zumindest während der primären Sozialisationsphase in stärkerem Maße eingesetzt würde. Somit würde, so die These Prillwitz', eine Übernahme wesentlicher Funktionen der Lautsprache durch die Gebärdensprache ermöglicht werden, was der Gesamtentwicklung Gehörloser förderlich wäre.

Zur Überprüfung dieser Annahme legt Prillwitz zunächst eine Untersuchung der Funktionen der Lautsprache vor. Weiterhin geht er der Frage nach, inwieweit die Gebärdensprache diese Funktionen übernehmen kann bzw. Besonderheiten gegenüber der Lautsprache aufweist. Dies setzt wiederum die eingehende Untersuchung der Gebärdensprache in Hinblick auf die Qualität ihrer Zeichen, ihre Struktur und Entwicklungsgeschichte etc. voraus, die Prillwitz ebenfalls vorlegt.

Auf die Einzelheiten dieser Funktionsbestimmung und die spezifischen Charakteristika der Modalitäten soll an dieser Stelle nicht eingegangen werden, da diese in angegebener Quelle nachvollzogen werden können und zudem dem vorliegenden Zusammenhang lediglich als Ausgangspunkt theoretischer Grundlegung für die Auswahl der in den Wortschatztest aufgenommenen Sprachmodalitäten dient. Ziel der folgenden Ausführungen ist demnach eine Bestimmung der Komponenten des sprachlichen Umfeldes Gehörloser, auf welche der Wortschatztest abgestimmt sein muss.

Für gehörlose Kinder gelten die von Prillwitz aufgezeigten spezifischen Funktionen der Lautsprache nur bedingt, da sie die Lautsprache nur bedingt verstehen und erlernen können. Sie ermöglicht Gehörlosen (mehr oder weniger, je nach dem, wie vollständig sie erlernt wurde,) die Verständigung in ihrer hörenden Umwelt und ist daher wichtiger Bestandteil der schulischen Erziehung Gehörloser. So ist dies im Lehrplan für die Hamburger Gehörlosen-

schule (1975) verankert, wenn als Ziel des Artikulationsunterrichts (Sprecherziehung) die Verständigung des Gehörlosen in seiner hörenden Umwelt angegeben wird. Dem Erlernen der Bewegungsabläufe im Bereich der Sprachwerkzeuge wird daher im Sprachunterricht der Früherziehung und der ersten Grundschuljahre eine bedeutsame Stellung eingeräumt, da der dort erzielte Erfolg mit zur Minderung der Isolation des Gehörlosen beiträgt. Ebenso wie der aktive Gebrauch der Lautsprache ist auch der passive, das Ablesen sprachlicher Äußerungen vom Mund anderer, für Gehörlose in der Kommunikation mit Hörenden von großer Bedeutung. Trotz unsicherer Informationsübertragung durch die undeutlichen, geringen Differenzierungsgestalten des Absehbildes wird es zum Sinnträger. Die Lautsprache nimmt für Gehörlose als Mittel zur Kommunikation mit der hörenden Mehrheit demzufolge eine grundlegend andere Position ein als für Hörende, die die Lautsprache auf natürlichem Wege als Muttersprache erlernen, sie hat jedoch insbesondere unter Sozialisationsaspekten Gehörloser als (gebärdensprachlicher) Minderheit entscheidende Bedeutung.

Prillwitz' Untersuchung der Gebärdensprache verdeutlicht, dass diese für Gehörlose vergleichbare Funktionen übernimmt, wie die Lautsprache für Hörende. Es zeigt sich, dass die Gebärdensprache, zumindest in den ersten sechs Lebensjahren, viele der lautsprachlichen Funktionen erfüllen kann. Sie erscheint somit als Medium zur Sprachanbahnung und ersten Sprachverwendung bei gehörlosen Kindern geeignet. Folglich wäre der Stellenwert der Lautsprache in der frühen Bildung Gehörloser vornehmlich durch ihre Funktion als Medium zur Kommunikation mit der hörenden Umwelt und weniger als zentrales und möglicherweise einziges Kommunikationsmittel überhaupt bestimmt.

Die Bedeutung der Schriftsprache für jedes Mitglied der Gesellschaft - hörend oder gehörlos- ist evident, wenn man sich die Notwendigkeit der Orientierung in Lehrbüchern im Schulunterricht und Fachbüchern in der Berufsausbildung sowie im Bereich der Politik, Kultur, Wirtschaft etc. über Printmedien, Internet u.ä. verdeutlicht. Hinzu kommt der Wunsch der Aufnahmefähigkeit sog. 'schöngestiger' Literatur. Ähnlich betont Claussen (1975) die Bedeutung des Lesens als möglicherweise wichtigstes Mittel für den Hörsprachgeschädigten, sich beruflich, politisch und kulturell zu informieren. Nach Ansicht Claussens würde eine über das Lesen erreichte bessere Orientierung dessen Chancen maßgeblich verbessern, an Gesprächen Hörender zu partizipieren. Hinzu kommen technische Hilfsmittel für Gehörlose, die mit der Schrift operieren, wie das Schreibtelefon oder Untertitel im Fernsehen.

Die Schriftsprache kann als Brücke zwischen Gehörlosen und Hörenden fungieren, da Gehörlosen mit ihrer Hilfe nicht nur Informationen, sondern auch die Form und Grammatik der Lautsprache ihrer hörenden Umwelt vermittelt werden kann. Dementsprechend sieht der Lehrplan (1975) die Förderung des schriftlichen Ausdrucks als wichtiges Unterrichtsziel vor,

da die Schrift für Gehörlose das sicherste Mittel zur Fixierung ist und somit eine enge Kopplung zwischen dem Gesprochenen und Geschriebenen besteht. So wird die Bedeutung der Schriftsprache beim Sprachaufbau an der Gehörlosenschule bei Günther (1985) ausdrücklich hervorgehoben. Diese Hervorhebung geht bereits auf die Anfänge der Gehörlosenpädagogik in Spanien und später in Frankreich zurück und wird später bei Kern u. a. beibehalten. Günther stellt jedoch fest, dass die Schriftsprache in der jüngeren Vergangenheit unangemessen vernachlässigt wurde. Dabei ist es sein besonderes Anliegen, darauf hinzuweisen, dass der Schrift eine eigenständige und nicht nur eine Hilfsfunktion zukommt. Dieser Verweis der Schrift auf einen untergeordneten Rang sei eine auch in der allgemeinen Linguistik vorkommende Fehleinschätzung. Gleichzeitig leugnet er die Ansicht, dass die Schrift negative Auswirkungen auf die Sprechaktivität haben könnte. Der Einsatz der Schriftsprache wird bei Günther auch nicht als alternative, sondern als ergänzende Methode zur lautsprachlichen Erziehung betrachtet, ohne dabei jedoch der Schriftsprache lediglich die Funktion einer Sprechlernhilfe zukommen zu lassen. Der Erfolg des Einsatzes der Schriftsprache ist vielmehr maßgeblich davon abhängig, dass diese theoretisch, methodisch und didaktisch der Lautsprache entsprechend reflektiert wird. Auf Grund dieser Einsichten wird es als notwendig erachtet, der Schriftsprache beim Sprachaufbau an der Gehörlosenschule größeres Gewicht zu verleihen.

Die i. d. R. bestehende Annahme, dass der Schriftspracherwerb die gesprochene Sprache als notwendige Voraussetzung benötige, verstellt nach Günther (1983) den Blick auf die Einsicht, dass das Sprachsystem von der Schrift in autonomer Weise abgebildet wird und Sprachkompetenz demgemäß über die Schrift unabhängig vom erreichten lautsprachlichen Niveau erreicht werden kann. Der so herausgearbeitete eigenständige Charakter der Schriftsprache sowie die Einsicht in die Möglichkeiten, die sich daraus in Hinblick auf ihre vermittelnde und unterstützende Funktion auf dem Weg zur Sprachkompetenz ergeben, stellt die Grundlage für ihre gleichberechtigte und daher auch gleichgewichtete Berücksichtigung in dem vorliegenden Test dar.

Stellt man die Eigenartigkeit und damit Eigenständigkeit der beiden Modalitäten, der Schrift- und der Lautsprache, in den Mittelpunkt der Betrachtung, so wird, bezogen auf den vorliegenden Test, deutlich, dass sich die lexikalisch-semantische Kompetenz einer Testperson theoretisch nur in einer der Modalitäten unter Beweis stellen lassen kann. So kann ein Test, der ausschließlich in der lautsprachlichen Modalität operiert, den Blick auf lexikalisch-semantische Kompetenzen verstellen, die vorhanden sind, jedoch von der betreffenden Person nur in der schriftsprachlichen Modalität beherrscht werden.

Die Ausführungen Prillwitz' zur Funktionsbestimmung von Laut- und Gebärdensprache bilden als theoretische Grundlegung einen der frühen Ausgangspunkte für den Einsatz von Ge-

bärden in einigen gehörlosenpädagogischen Institutionen Deutschlands (insbesondere in den ersten Jahren der Vor- und Schulbildung). Diese Entwicklung der Einbeziehung der Deutschen Gebärdensprache neben der Lautsprachbahnung in den sprachlichen Anfangsunterricht mit Gehörlosen sowie die Betonung der Bedeutung der Schriftsprache und ihre konzeptionelle curriculare Berücksichtigung ist heute bereits deutlich fortgeschritten, wird fortlaufend wissenschaftlich untersucht (vgl. Ahrbeck 1992; Günther 1993; u. a.) und hat sich teilweise institutionell etabliert. Die Autorinnen des vorliegenden Wortschatztests berücksichtigen diese Entwicklung, wenn sie dem Testinstrument alle drei Sprachmodalitäten zu Grunde legen. Gehörlose Kinder bewegen sich heute häufig in allen drei, mindestens aber (bei oraler Erziehung) in zwei der Modalitäten. Um universal einsetzbar zu sein und dem neuesten Stand der Gehörlosenbildung zu entsprechen, muss ein Wortschatztest demnach modalitätenumfassend konzipiert sein.

Die Funktionsbestimmungen sowie die Beobachtung der Praxis im Umgang mit Gehörlosen legen die Annahme nahe, dass die Testergebnisse durchgängig, d. h. unabhängig von der jeweiligen Sprachbahnungsmethode, nach der die Testpersonen unterrichtet werden, in der Modalität Gebärdensprache am besten ausfallen werden, gefolgt von der Schriftsprache und schließlich der Lautsprache mit den vermutlich schwächsten Ergebnissen. Diese Annahme beruht auf den dargestellten modalitätenspezifischen Eigenarten und den mit diesen in Verbindung stehenden spezifischen Schwierigkeiten Gehörloser, sich insbesondere die Laut- aber auch die Schriftsprache anzueignen. Demgegenüber stehen deren vielfach gut ausgebildete Gebärdensprachkompetenzen.

Durch die Einbeziehung aller drei Sprachmodalitäten eignet sich der Test gleichermaßen für lautsprachlich und für gebärdensprachlich orientierte Anwender, die wahlweise auch nur einen oder zwei der jeweils eine der Modalitäten umfassenden Subtests zur Anwendung kommen lassen können. Bei vollständiger Durchführung jedoch bietet der Test die Möglichkeit, die lexikalisch-semantischen Fähigkeiten eines gehörlosen Kindes auf den verschiedenen sprachlich-modalen Ebenen zu untersuchen.

Durch die Einbeziehung der Gebärdensprache als Mutter- oder Basissprache¹⁴ Gehörloser bzw. als der für Gehörlose am leichtesten und daher potenziell am vollständigsten zugängli-

¹⁴Da die Gebärdensprache die einzige Sprachmodalität ist, die von Gehörlosen vergleichsweise leicht und damit potentiell vollständig erworben werden kann und strukturell auf die behinderungsspezifischen Bedingungen Gehörloser abgestimmt ist, ließe sie sich prinzipiell als ihre 'Muttersprache' bezeichnen. Im vorliegenden Zusammenhang soll der Begriff der Gebärdensprache als 'Muttersprache' jedoch enger gefaßt werden. Als 'muttersprachlich gebärdensprachlich' werden demnach (hörende oder gehörlose) Personen bezeichnet, die in ihrem häuslichen Umfeld durch bestehende Gehörlosigkeit mindestens einen Elternteils gebärdensprachlich aufwachsen. Für Gehörlose, die die Gebärdensprache im Rahmen von Kindergarten, Schule, Freizeiteinrichtungen o.ä. als Sekundärsprache erlernen, bleibt der Umstand der leichten Erlernbarkeit jedoch erhalten. Daher ergibt sich faktisch i. d. R. eine Vorliebe für die Gebärdensprache und im Vergleich zur Lautsprache – so wird im vorliegenden Zusammenhang angenommen – eine gute Gebärdensprachkompetenz. D.

chen Modalität eröffnet sich die Möglichkeit, deren Testergebnisse (im Subtest Gebärdensprache) denen Hörender, Schwerhöriger oder Sprachbehinderter (in ihrer Muttersprache, der Lautsprache) gegenüberzustellen. So wird es auch unter diesem Aspekt als zentrales konzeptionelles Element des vorliegenden Tests erachtet, dass dieser eben nicht, wie bei bisher gebräuchlichen Tests, nur lautsprachlich, sondern alternativ auch gebärdensprachlich und/oder schriftsprachlich eingesetzt werden kann, um ihn bei gehörlosen und anderen lautsprachlich behinderten Kindern verwenden und Vergleiche mit den lexikalisch-semanticen Leistungen nicht behinderter Kinder ziehen zu können.

Zudem ist ein solcher Vergleich mit den lexikalisch-semanticen Leistungen Hörender in Hinblick auf die Bezugsgröße ‘Grundwortschatz’ unter Aspekten der Instrumentenentwicklung interessant. Er dient der zusätzlichen Überprüfung der (durch die der Wortliste zu Grunde liegenden Untersuchungen bereits überprüften)¹⁵ Annahme, dass es sich bei den Items um Wörter handelt, die sich im Grundwortschatz hörender Grundschulkinder befinden und daher korrekt von ihnen beantwortet werden müssten.

1.2 Auswahl und Gewichtung der Wortarten

Die deutsche Sprache setzt sich hauptsächlich aus acht Wortarten zusammen, nämlich den Nomen, den Verben, den Adjektiven, den Adverbien, den Artikeln, den Pronomen, den Präpositionen und den Konjunktionen. Da die letztgenannten vier Wortarten insgesamt nur etwa 300 Wörter umfassen, liegt es nahe, sich bei der Auswahl der Wortarten für einen Test auf die Übrigen zu beschränken. Diese lassen sich nach Günther (1991, 172) in drei referenzsemanticen Hauptwortarten – Nomen, Verben und Adjektive (+Inhalts-Adverbien) gliedern, deren Funktionen und Besonderheiten in Hinblick auf ihre Verwendung in Wortschatztests im Folgenden erläutert werden.

Laut Brockhaus (multimedial 2000) ist ‘Nomen’ sprachwissenschaftlich definiert als „zusammenfassende Bezeichnung für Substantiv, Adjektiv und Numerale, im engeren Sinne nur für Substantiv.“ In diesem engeren Sinne wird es vorliegend verwendet, dabei wird der Begriff Nomen gewählt. Das Substantiv wird vom Brockhaus definiert als „(Nennwort,

h., das das Niveau der Sprachkompetenz von Muttersprachlern nicht nur bei muttersprachlich lautsprachlichen sowie bei muttersprachlich gebärdensprachlichen Personen, sondern annähernd auch bei sekundärsprachlich gebärdensprachlichen Gehörlosen vergleichbar ist. Dieser Umstand rechtfertigt das Vorgehen, sowohl muttersprachlich gebärdensprachliche als auch sekundärsprachlich gebärdensprachliche Gehörlose für den Vergleich mit muttersprachlich lautsprachlichen Hörenden heranzuziehen. Für die sekundärsprachlich gebärdensprachlichen Personen wird im Folgenden für die DGS der Begriff der ‘Basissprache’ verwendet.
¹⁵ Näheres s. B 1.3 Auswahl von Aufgabentyp, Testitems und Distraktoren

Hauptwort, Dingwort), eine Wortart, die Gegenständliches (Lebewesen, Dinge) und Nichtgegenständliches (Gedachtes, Begriffliches) bezeichnet.“ Nomen können gebraucht werden als Subjekt oder Objekt, als Umstandswort oder als Attribut. Die Wortart der Nomen macht im Wortschatz von Grundschulkindern 54,5% (3772 Nomen), in dem von Erwachsenen sogar 61,5% (9398 Nomen) aus, wobei es sich bei diesen wiederum um 52% bzw. 64% zusammengesetzte Nomen (Komposita) handelt (vgl. Pregel, Rickheit u. Ruoff in Günther 1991, 172).

Der Wortschatzzuwachs vom Grundschul- bis zum Erwachsenenalter ist zu zwei Dritteln auf die Nomen zurückzuführen, was bedeutet, dass gerade der Ausbau des nominalen Wortschatzes stark erfahrungs- und entwicklungsbedingt und somit weniger sprachtheoretisch verursacht ist. Der Umstand, dass die gebräuchlichen Wortschatztests überwiegend Items aus der Kategorie Nomen abfragen, kann demnach dazu führen, dass Testpersonen mit Wörtern konfrontiert werden, die sich in ihren persönlichen Erfahrungsfeldern nicht wiederfinden. Es wäre daher ungerechtfertigt, aus einem schlechten Testergebnis einen eingeschränkten Wortschatz abzuleiten, da hier u. U. Erfahrungswerte statt Wortschatzkenntnis unter Beweis gestellt werden. Entscheidend für einen Kinder- bzw. Grundwortschatztest ist daher die Verwendung kindgerechter Nomen, deren Verfügbarkeit erwartet werden kann.

Das ‚Verb‘ wird im Brockhaus (multimedial 2000) definiert als „konjungierbares Wort, das einen Zustand oder Vorgang, eine Tätigkeit oder Handlung bezeichnet und mit dem eine Tätigkeit oder Handlung bezeichnet wird.“ Eine Form der Verben, die Hilfsverben ‘haben’, ‘sein’ und ‘werden’, sollen im vorliegenden Zusammenhang außer Acht gelassen werden, da sie nur in Verbindung mit einem anderen Verb der ersten drei Gruppen auftreten und somit für einen Wortschatztest ungeeignet sind.

Im Lexikon von Grundschulkindern nehmen Nomen den Anteil von 54,5% ein, das entspricht einer Anzahl von 3772. Im Erwachsenenlexikon werden (mit insgesamt 61,5%) 7% mehr Nomen verwendet (9398). Demgegenüber steht das Gefälle im Vergleich der Anteile von Verben. Hier verwenden Grundschul Kinder mit 33,9% (2345) anteilig 5% mehr als Erwachsene mit 28,9% (4414). Verben nehmen also unterproportional zu (vgl. Pregel, Rickheit u. Ruoff in Günther 1991, 172).

Dabei machen nur etwa 100 Verben drei Viertel des gesamten Verbwortschatzes der Grundschul Kinder aus. Diese wenigen Verben können auch als ‘Grundverben’ bezeichnet werden, die die für den Alltag wichtigen Handlungsabläufe oder Zustände beschreiben und von denen sich eine Vielzahl der anderen Verben ableiten lassen. Diese Ausweitung des Verbwortschatzes kann durch direkte oder indirekte inhaltliche Ableitungen oder durch Komposita-Bildung erfolgen, letzteres jedoch weniger in der Verbindung Verb-Verb (häufigste Kompo-

sita-Bildung bei den Nomen: Nomen-Nomen), als durch Hinzufügung einer Präposition oder eines inhaltlich bedeutungslosen Präfixes, wie beispielsweise be-, ent-, ver- oder er-. In der Differenzierung der Grundverben liegt auch der Zuwachs des Verbwortschatzes bis zum Erwachsenenalter begründet. Da es nur diese relativ geringe Anzahl von Grundverben gibt, um alltägliche Vorgänge und Handlungen zu beschreiben, müssen sie relativ offen und verallgemeinerbar sein, um in den verschiedenen Situationen flexibel eingesetzt werden zu können. Die Beherrschung dieser Grundverben ist durch die Bedeutung des Verbs als Satzorganisator unbedingt erforderlich.

Das ‚Adjektiv‘ oder Eigenschaftswort ist laut Brockhaus (multimedial 2000) die „Wortart zur näheren Bestimmung eines Bezugsworts. Man unterscheidet **attributive** (z. B. die *rote* Rose), **prädikative** (z. B. die Rose ist *rot*) und **verbalappositive** Adjektive (z. B. er sieht *rot*).“ Wie auch die Nomen und die Verben können Adjektive in unterschiedlichen Situationen verschieden eingesetzt werden, nämlich als unreflektiertes, selbstständiges Satzglied, als reflektiertes Attribut beim Nomen oder als unreflektiertes Attribut beim Adverb oder Adjektiv. Adjektive sind im Gegensatz zu Nomen und Verben nicht unabdingbarer Bestandteil eines Satzes, sondern sind nach eigenem Ermessen und Sprachgefühl in ihn einzubauen. Sie nehmen somit eine primär semantisch-qualifizierende Funktion in der Sprachproduktion ein. So ist ihr quantitativer Anteil an der deutschen Sprache mit nur 11,6% (804 Adjektive) bei Grundschulkindern und 9,6% (1462 Adjektive) bei den Erwachsenen recht gering. Bemerkenswert ist dabei, dass von den 804 Adjektiven bei den Grundschulkindern nur zwölf Adjektive schon 50% und 100 Adjektive knapp 90% des gesamten aktiven Adjektivwortschatzes ausmachen (vgl. Pregel, Rickheit u. Ruoff in Günther 1991, 172 u. 177). Der ohnehin schon geringe adjektivische Wortschatz ist demnach auf sehr wenige gebräuchliche Adjektive begrenzt. Ein hoher adjektivischer Wortschatz qualifiziert somit das Sprachniveau einer Person.

Grundsätzlich muss zwischen absoluten und relativen Adjektiven unterschieden werden. Ein absolutes Adjektiv hat eine Merkmalsausprägung, die es von allen anderen situationsunabhängig unterscheidet. Dies sind z.B. alle Farb-, Form- und Substanzbezeichnungen. Es gilt: Entweder ein Kind beherrscht z.B. Grundfarbbezeichnungen oder eben noch nicht. Günther (1991, 178) stellt fest, dass die Absolutheit der Merkmalsausprägung dazu führt, dass die Bezeichnungen nicht systemhaft mit dem Erkennen der Dimension, sondern einzelheitlich sukzessiv angeeignet werden, sodass beispielsweise rot als alleinige Farbbezeichnung verfügbar sein kann. Beherrscht ein Kind die absoluten Adjektive erst einmal, so erweisen sie sich als weitaus unproblematischer als die relativen Adjektive, denen kein objektivierbarer Maßstab zu Grunde liegt. Relative Adjektive beschreiben immer einen Punkt zwischen zwei Extrempolen, so ist zum Beispiel die Größe einer Maus abhängig von ihrem Gegenpol: Die Maus ist groß im Vergleich zu einer Ameise, aber klein im Vergleich zu einem Elefanten.

Diese spezifische Struktur speziell der relativen Adjektive stellt an den Sprachlerner erhebliche sprachliche und kognitive Anforderungen. Dies beweist sich in dem Umstand, dass sie nach den Nomen und den Verben als letzte der referenzsemantischen Wortarten in der frühkindlichen Sprachproduktion auftauchen (Günther 1991, 180).

Die fast ausschließliche Verwendung von Nomen in bisher gebräuchlichen Wortschatztests mag insoweit sinnvoll erscheinen, als dass Hauptwörter über 50% des deutschen Wortschatzes ausmachen, zudem am einfachsten und deutlichsten bildlich darzustellen sind, und somit eine Eindeutigkeit des Items am ehesten gewährleistet ist. Ein solches Vorgehen vernachlässigt jedoch die Tatsache, dass sich Nomen als besonders ungeeignet zur Erfassung lexikalisch-semantischer Fähigkeiten und eventueller Störungen erweisen, weil die wenigsten Nomen unabhängig von Alter, sozial-kultureller Erfahrung und angeeignetem Wissen sind. Die Entwicklung des nominalen Wortschatzes weist also in hohem Maße erfahrungs- und entwicklungsbedingte Abhängigkeiten auf, was sich vor allem in der geringen Übereinstimmung des kindlichen und des erwachsenen Nomenwortschatzes, aber auch in Wortschätzen von Kindern unterschiedlicher Herkunft und unterschiedlichen Umfelds ausdrückt. Insofern lässt sich kein allgemeiner nominaler Grundwortschatz feststellen, der diese Abhängigkeiten nicht aufweist. Aus dem Blickwinkel der Linguistik muss zudem nach Günther (1991, 187) festgestellt werden, „daß sich aus Umfang und Niveau des nominalen Wortschatzes für sich genommen kaum etwas über die Fähigkeit schließen lässt, Sprache in semantisch und syntaktisch kohärenten Äußerungen zu gebrauchen. Verben als Organisationszentrum des Satzes sind diesbezüglich viel aussagekräftiger. Adjektive schließlich implizieren durch ihren relativ-dimensionalen Bedeutungscharakter und ihre primär semantisch-qualifizierende Funktion über den reinen Wortschatzaspekt hinaus Hinweise auf das Niveau der Sprachfähigkeit im semantischen Bereich.“

Diesen Überlegungen zu den Funktionen der Wortarten zufolge wurde die Itemauswahl des Testinstruments aus den drei Hauptwortarten Nomen, Verben und Adjektive rekrutiert. Wie gezeigt wurde, sind diese unter der Fragestellung lexikalisch-semantischer Fähigkeiten den übrigen Wortarten gegenüber mit Abstand am aussagekräftigsten. Dabei übernimmt jede der drei Wortarten spezifische Anteile dieses Fähigkeitsbereichs.

Neben der Auswahl der Wortarten muss eine Festlegung des Umfanges der jeweiligen Wortart im Itemkorpus vorgenommen werden. Betrachtet man dazu die vorangegangenen Ausführungen, so stellen sich zweierlei Faktoren einander entgegengesetzt dar. Die Aussagekraft der Wortarten in Hinblick auf die lexikalisch-semantischen Fähigkeiten lässt ein Gefälle von den Adjektiven über die Verben zu den am wenigsten aussagekräftigen Nomen erkennen. Demgegenüber steht das zahlenmäßige Vorkommen der Wortarten im Wortschatz. Hier stellt sich das Gefälle anders herum von den Nomen über die Verben zu den (unter diesen dreien)

seltensten Adjektiven dar. Aus der Zusammenschau dieser beiden ausschlaggebenden Faktoren ergibt sich die Entscheidung, die Wortarten in gleichgewichteten Dritteln in den Itemkorpus aufzunehmen.

Diese Zuteilungsvariante wird zusätzlich durch folgende pragmatische Notwendigkeiten gestützt. So mussten zunächst für jede Wortart (auch für die beiden weniger zahlreich vorkommenden) ausreichend Items zur Verfügung gestellt werden, um im Schwierigkeitsgrad differenzieren zu können. Wäre nun eine Angleichung des Umfanges der Wortarten an ihr Vorkommen in der Sprache vorgenommen worden, so hätte sich folgende Situation ergeben: Im Fall der Nomen wäre eine Auswahl, die dieser den Umfang betreffenden Forderung entspricht, ihrem zahlreichen Vorkommen in der Sprache entsprechend problemlos möglich gewesen. Die zahlenmäßig geringer vorkommenden Verben und Adjektive hätten jedoch bei einer Anpassung den Nomen gegenüber nicht gleichgewichtet berücksichtigt werden können. Hinzu kommt das Vorhaben, Aufschluss über die hypothetisch angenommene Schwierigkeitsabstufung der Wortarten zu erhalten. Auch in dieser Hinsicht ist ein ausreichendes Vorkommen jeder der Wortarten notwendig, um Rückschlüsse aus entsprechend repräsentativen Ergebnissen ziehen zu können.

In Verbindung mit den strukturellen Eigenarten der drei Wortarten stehen unterschiedliche Schwierigkeitsgrade ihrer Aneignung. So wird davon ausgegangen, dass den Testpersonen die Zuordnung der Nomen leichter fallen wird als die der Verben und diese wiederum leichter als die der Adjektive. Die Abfolge der Wortartenblöcke innerhalb der jeweiligen Subtests gestaltet sich gemäß dieser Reihenfolge.

1.3 Auswahl von Aufgabentyp, Testitems und Distraktoren

Der Aufgabentyp war im Rahmen der unter A 2.4.2 erläuterten grundsätzlichen Vorüberlegungen zur Testkonstruktion bereits auf die „Antwort-Auswahl-Aufgabe“ (Ingenkamp 1972a, 51) festgelegt worden. Prinzipiell kann für die Einteilung von Aufgabentypen eine Gruppierung nach Auswahlart und Konstruktionsart vorgenommen werden. Bei der Auswahlart, die im vorliegenden Fall interessiert, werden Antworten, Lösungen oder Konstruktionen vorgegeben und die Testperson soll daraus die richtige oder beste auswählen. Der gebräuchlichste Typ dieser Auswahlart ist die Auswahl-Antwort-Form, bei der nach einer Aufgabenstellung oder Frage vier oder fünf Lösungen angeboten werden, von denen eine ausgewählt werden muss. Ingenkamp verwendet den Terminus ‘Antwort-Auswahl-Aufgabe’, weil er den sonst oft verwendeten Begriff ‘Mehrfach-Wahl-Aufgabe’ für irreführend hält, da es bei der Auswahlart nur wenige und selten benutzte Aufgabentypen gibt, bei denen nicht

eine, sondern mehrere Lösungen gewählt werden müssen. Dieser Terminus wird im Folgenden beibehalten.

Der Aufgabentyp bietet vielfache Vorteile. So sind Antwort-Auswahl-Aufgaben nicht nur zeitsparend in der Antworttechnik, sondern auch objektiv und ökonomisch auswertbar. Zudem können sie in den inhaltlichen Anforderungen sehr variabel gestaltet werden. So ist beispielsweise der Schwierigkeitsgrad der Aufgaben durch die Wahl der Distraktoren sehr gut beeinflussbar und differenzierbar. Zudem verstehen die Testpersonen das einfache Prinzip der Beantwortung in der Regel schnell und der Vorgang des Antwortens geht rasch vor sich. Dies ist unter Berücksichtigung insbesondere der abnehmenden Konzentrationskurve einer Testperson im Verlauf der Testdurchführung wünschenswert und hat unter zeitökonomischen Erwägungen eine generelle Steigerung der Testökonomie zur Folge.

Als Grundlage für die Itemauswahl diente, wie bereits erwähnt, die von Günther zusammengestellte Wortliste (1989 b, s. Tab. 4-B 1.3, Anh. B 1.3a), bei der es sich um eine Zusammenfassung der Ergebnisse von vier verschiedenen Wortschatzstudien von Pregel, Rickheit und Ruoff und um deren Vergleich handelt. Die Wortliste stellt eine "Übersicht über die je 100 häufigst gebrauchten Nomen, Verben, Adjektive von Grundschulkindern (mündlich/schriftlich – Pregel & Rickheit 1987), von Viertklässlern (schriftlich – Augst 1989) sowie ihr Vorkommen bei Vorschülern (mündlich – Augst 1984), verglichen mit den je 100 häufigsten Inhaltsworten Erwachsener (mündlich – Ruoff 1981)" (Günther 1991, 1) dar. Pregel und Rickheit untersuchten 1987 sowohl den mündlichen als auch den schriftlichen Wortschatz von insgesamt 1046 Grundschulkindern aus verschiedenen soziografisch differenzierten Regionen der (damaligen) Bundesrepublik. Methode der Sprachdatenerfassung waren hierbei vier festgelegte Sprachproduktionsimpulse, nämlich eine lustige Geschichte, eine Tiergeschichte, ein Unfallbericht sowie die Beschreibung einer Person bzw. einer Kasperlepuppe. Zusätzlich waren die Kinder zu einer freien Erlebniserzählung aufgefordert. Die von Augst herausgegebene Untersuchung von 1984 beschäftigt sich mit dem mündlichen Wortschatz von zehn Vorschulkindern aus der mittleren bis oberen Mittelschicht. Es wurden hierfür 200 Stunden der sprachlichen Äußerungen der Kinder ausgewertet. Augst untersuchte weiterhin den Schriftwortschatz von zehn Viertklässlern, indem er sämtliche Schulhefte der Kinder bis zum vierten Schuljahr auswertete. Als Vergleich zum Wortschatz von Vorschülern bis Viertklässlern beinhaltet die Wortliste den mündlichen Wortschatz Erwachsener, 1981 von Ruoff untersucht, der 350 Stunden unvorbereiteter, freier Gespräche von 2500 Erwachsenen aus verschiedenen Orten des südwestdeutschen Sprachraums auswertete.

Die Itemauswahl des vorliegenden Tests wurde aus dieser Wortliste zusammengestellt. Auf Grund dieser Bezugnahme sind die Ergebnisse der mit dem Test durchgeführten Erhebungen stets auf der Grundlage bzw. in Abhängigkeit von diesem Material zu betrachten. Da es sich

dabei um empirisch erhobenes Material handelt, wird so eine willkürliche Wortauswahl vermieden. Die Wortliste bildet den Grundwortschatz hörender deutscher Grundschüler ab, der für einen Vergleich mit der Wortschatzkenntnis gleichaltriger gehörloser Grundschüler als normative Bezugsgröße fungiert.

Aus der 435 Wörter umfassenden Liste wurde eine Wortauswahl von 54 Testitems (je 18 pro Wortart) getroffen. Diese Anzahl bildet einerseits ein umfangreiches Bild über den Wortschatz des zu testenden Kindes ab, ohne es andererseits durch eine zu lange Dauer eines Testdurchganges zu überfordern. Um eine möglichst hohe Objektivität in der Wortauswahl zu gewährleisten, wurden die Items nicht beliebig, sondern in regelmäßigen Abständen aus der Vorlage ausgewählt. Dabei steigert sich, gemäß dem Aufbau der Wortliste, der Schwierigkeitsgrad (bzw. Seltenheitsgrad im Gebrauch) der Wörter innerhalb jedes Wortartenblocks stetig.

Die so erzielte Itemauswahl lautet:

Tab. 5-B 1.3: Itemkorpus Pilotversion		
Nomen	Verben	Adjektive
1. Mutter	1. sagen	1. ganz
2. Auto	2. gehen	2. klein
3. Haus	3. fahren	3. blau
4. Kopf	4. essen	4. voll
5. Straße	5. gucken	5. blond
6. Bett	6. fliegen	6. komisch
7. Angst	7. schlafen	7. krumm
8. Stein	8. denken	8. hellblau
9. Tier	9. hinsetzen	9. orange
10. Bart	10. warten	10. traurig
11. Zahn	11. verstecken	11. verletzt
12. Leute	12. weinen	12. warm
13. Zahl	13. schreiben	13. glatt
14. Geld	14. lesen	14. links
15. Kasten	15. singen	15. frei
16. Holz	16. waschen	16. schmutzig
17. Arbeit	17. brennen	17. weich
18. Schloß ¹⁶	18. schieben	18. leise

¹⁶ Da zum Zeitpunkt der Testkonstruktion und der Durchführung der Pilotstudie sowie der Hauptuntersuchung die ‚alte Rechtschreibung‘ Gültigkeit hatte, wurde diese (betreffend die Verwendung des ‚ß‘) vorliegend bis zum Abschluss der Darstellung von Pilotstudie und Hauptuntersuchung beibehalten. Mit dem Gültigkeitsbeginn der

Jedem Testitem wurden gemäß dem Aufgaben-Aufbauschema drei Distraktoren zugeordnet. Ein Distraktor sollte auf qualitativ unterschiedliche Weise, durch Ähnlichkeiten im Hinblick auf Gebärde, Schriftbild, Mundbild oder auf inhaltlicher Ebene, vom Testitem ablenken. (Auf dieses zentrale Element der qualitativen Konzeptionsmerkmale wird im Rahmen der Hauptuntersuchung unter B 3.4.5 ‚Distraktorenanalyse‘ und unter B 3.10 ‚Distraktorenprofil‘ vertieft eingegangen.) Zusätzlich musste der ausgewählte Begriff bildlich darstellbar sein. Im günstigsten Fall wurde einem Item je ein Distraktor mit Gebärdenähnlichkeit, einer mit Mund- bzw. Schriftbildähnlichkeit und einer mit Inhaltsähnlichkeit zugeordnet, (wie z. B. bei dem Item ‘Mutter’ mit ‘Bauer’ als Ähnlichkeit in der Gebärde, ‘Motte’ in Anlehnung an das Schrift- und Mundbild und ‘Mädchen’ als inhaltliche Ähnlichkeit). Auf Grund der vorgegebenen begrenzten Auswahlmöglichkeit, der Notwendigkeit zeichnerischer Darstellbarkeit und mangels Erfahrung mit dem Testmaterial wies diese Systematik in der Pilotversion noch Lücken auf, wie die im Anhang (Anh. 1: B 1.3b) vorfindlichen Tabellen zu den Distraktorenähnlichkeiten (Tab. 6a/b/c-B 1.3) und Erläuterungen unter B 2.4.4 verdeutlichen.

Der Aufbau der Tabellen, die diese Zuordnung der Distraktoren darstellen, stellt sich wie folgt dar: Auf drei Seiten sind jeweils die Nomen, die Verben und die Adjektive mit den dazugehörigen Distraktoren aufgelistet, wobei in der ersten Spalte die Items, in der zweiten Spalte die entsprechenden Distraktoren eines jeden Items abgebildet sind. Die Items sind gemäß ihres Vorkommens im Testheft aufgelistet, die Distraktoren jeweils alphabetisch geordnet. Dahinter können in drei Spalten die jeweiligen Ähnlichkeiten der Distraktoren mit ihrem Item abgelesen werden. Die erste Spalte steht für die Ähnlichkeit in der Gebärde, welche hauptsächlich für den ersten Subtest Gebärdensprache entscheidend ist. Hierbei kann es sich um Ähnlichkeiten in der Handform, der Handstellung, in der Ausführungsstelle oder im Bewegungsablauf der Gebärden handeln, wobei häufig nur eine oder zwei dieser Ähnlichkeiten ausreichen, um die Gebärde als Distraktor auszuwählen. Die zweite Spalte enthält Schrift- und Mundbildähnlichkeit in einem, wobei die Schriftbildähnlichkeit im zweiten Subtest, dem Schriftsprachttest, zum Tragen kommt, die Mundbildähnlichkeit hauptsächlich im dritten, dem Lautsprach-, aber auch im ersten, dem Gebärdensprachttest. Wenn die Ähnlichkeit, vor allem im Mundbild, nicht immer unmittelbar augenscheinlich ist, muss bedacht werden, dass beim Ablesen vom Mund ohne akustische Unterstützung viele der Buchstaben, die im hinteren Teil des Mundraums und im Rachenraum gebildet werden, nicht sichtbar sind und somit für das gehörlose Kind zur Wiedererkennung des Wortes praktisch ausfallen. Insofern eignet sich beispielsweise für ein Item wie ‘gucken’ nicht nur ein Distraktor wie ‘ducken’, der sich ausschließlich im Anlaut unterscheidet, sondern auch ‘küssen’ birgt ausreichend Ähnlichkeit, um als Distraktor in Betracht zu kommen. Die letzte Spalte, die Inhaltsähnlichkeit, ist testteilübergreifend von Bedeutung, denn inhaltliche Ähnlichkeit kann eine

‚neuen deutschen Rechtschreibung‘ und somit in der Endversion des Tests wird die Schreibung mit ‚ss‘ vorgenommen.

Testperson in jedem Subtest, unabhängig von der Sprachmodalität, ablenken. Die Auswahl der in diese Kategorie fallenden Distraktoren ist häufig nur im Zusammenhang mit den Zeichnungen zu verstehen, da sich einige Zeichnungen aufeinander beziehen und erst dadurch eine Ähnlichkeit zwischen Distraktor und Item ergibt, die dennoch inhaltlicher Art ist. Ein Beispiel hierfür sind die Begriffe ‘Arbeit’ als gesuchtes Item, dargestellt als Arbeitssituation am Schreibtisch, und ‘Eimer’ als Distraktor, dargestellt als Putzeimer mit Feudel. Auch der Eimer symbolisiert Arbeit, nämlich die Hausarbeit. Insofern besteht ein inhaltlicher Zusammenhang dieser beiden Begriffe durch die zeichnerische Darstellungsweise.

Im Nachhinein wurden einige Ungenauigkeiten in der Distraktorenauswahl erkennbar. Diese Ungenauigkeiten liegen in der Besonderheit der ‘inkorporierten’ Gebärden begründet. Von Inkorporation (Einverleibung) ist in der Gebärdensprache die Rede, wenn in eine Grundgebärde durch Veränderung von Handform, Handstelle, Ausführungsstelle oder Bewegung zusätzliche Informationen aufgenommen werden. So hat beispielsweise die Gebärde ‘halten’ unterschiedliche Ausführungen, je nach dem was gehalten wird, denn ein Buch wird bspw. anders gehalten als ein Baby. Da bei der Auswahl der Distraktoren die Darstellungsform der einzelnen Begriffe noch nicht klar war, traten im Nachhinein einige Fälle auf, die durch ihre zeichnerische Darstellung ihre Ähnlichkeit zum Item einbüßten. Ursprünglich war von der Grundform der Gebärden ausgegangen worden, ohne mögliche Veränderungen durch Inkorporation einzukalkulieren. So wurde das Verb ‘fahren’ als Fahren eines Fahrrads dargestellt, wodurch keine Gebärdenähnlichkeit zum Distraktor ‘schenken’ mehr zu erkennen ist. Die Grundgebärde ‘fahren’ hätte Ähnlichkeiten in Ausführungsstelle und Bewegung aufgewiesen. Der Begriffe ‘klein’ wird je nach bezeichnetem Gegenstand unterschiedlich gebärdet. Entweder deuten Daumen und Zeigefinger die Größe des Gegenstandes an oder die Hand zeigt die Höhe des Gegenstands vom Fußboden. Das Item wurde als Maus im Gegensatz(-paar) zu einem Elefanten dargestellt, wobei die Kleinheit der Maus mit Daumen und Zeigefinger angedeutet wird. Daher weist nur der Distraktor ‘klug’ eine Gebärdenähnlichkeit zum Item auf, während der ursprünglich in diesem Sinne eingeplante Distraktor ‘groß’ durch die Darstellung als Elefant (Hand zeigt die Höhe des Gegenstandes vom Fußboden) keine Ähnlichkeit aufweist. Da das Item ‘voll’ durch ein volles Glas symbolisiert ist, und nicht in dem ursprünglich anvisierten Zusammenhang ‘ich bin voll, ich habe genug’ gebraucht wird, entfällt die Gebärdenähnlichkeit zu dem Distraktor ‘viel’. Das Item ‘komisch’ wurde als witzig mit lachenden Personen dargestellt und nicht im Sinne von ‘eigenartig’, womit eine Gebärdenähnlichkeit zu ‘neu’ gewährleistet gewesen wäre.

1.4 Zeichnerische Gestaltung

Die zeichnerische Darstellung der Nomen gestaltete sich prinzipiell unproblematisch, da mit zwei Ausnahmen ('Angst' und 'Arbeit') Gegenständliches abzubilden war. Gegenstände lassen sich in aller Regel sowohl einfach als auch eindeutig darstellen. Eine solche Einfachheit der Darstellung ist wünschenswert, da auf diese Weise nichts (Zeichnungsimmanentes) von der Wahrnehmung des gefragten Gegenstandes ablenken kann. Wurden in einigen Fällen dennoch ausschmückende Details eingefügt, so geschah dies zum einen zur Unterstützung der Verständlichkeit. Allzu puristische Darstellungen können das Verständnis erschweren, da die Gegenstände auch in der Natur bzw. im Alltag der Testpersonen nicht in dieser Weise vorkommen, sondern i. d. R. Verzierungen, modische Elemente o.ä. an sich haben. Zum Zweiten sollte das Testinstrument ansprechend gestaltet sein, was ebenfalls für eine gewisse Ausgestaltung spricht. Zum Teil bestand zudem die Notwendigkeit, durch Randgegenstände den dargestellten Begriff überhaupt erkennbar zu machen. So ist bspw. ein 'Stein' durch seine Form nicht zweifelsfrei erkennbar, da er in der Natur in unterschiedlichsten Formen, Größen und mit verschiedenen Oberflächen vorkommt. Erst durch die Gestaltung der Bildumgebung mit Grashalmen und Kieselsteinen wird ein Verständniskontext geschaffen, der die korrekte Einordnung ermöglicht.

Die Darstellung der Verben war z. T. schwierig, da diese einen prozessualen Ablauf, eine Handlung oder einen Zustand beschreiben. Abläufe wurden daher vielfach durch angedeutete Bewegungslinien verdeutlicht. Abstrakta, wie z. B. Denkvorgänge, wurden z. T. unter Hinzunahme von Schriftzeichen und Symbolen dargestellt, was sich im weiteren Verlauf der Testentwicklung, wie unten erläutert werden wird, als ungünstig erwies.

Die zeichnerische Darstellung der Adjektive gestaltete sich erwartungsgemäß am schwierigsten. Zunächst stellte sich die Frage nach unterschiedlichen Gestaltungsformen der absoluten und der relativen Adjektive. Erstere, in vorliegendem Test die Farbadjektive, sind situationsunabhängig und lassen sich daher eindeutig darstellen. Die relativen Adjektive wurden meist als Gegensatzpaare dargestellt und dabei wurde der jeweils nicht gefragte Gegensatz durchgestrichen. Dieses Prinzip wurde jedoch nur in den Fällen angewandt, in denen die Darstellung als Gegensatzpaar unerlässlich war. Wäre jedes relative Adjektiv auf diese Weise dargestellt worden, hätte die Testperson pro Seite statt vier Bilder fast immer acht betrachten müssen, was unübersichtlich geworden wäre.

Die vier Bilder, die auf einer Seite des Testheftes abgebildet sind, wurden - unter zeichnerischen Aspekten - zunächst unabhängig voneinander betrachtet. Dieses Vorgehen trifft jedoch nicht in jedem Fall zu. Für das Item 'schmutzig' sind beispielsweise die Distraktoren 'sauber', 'zerknüllt' und 'zerrissen' ausgewählt und alle anhand einer Hose dargestellt worden.

So sind also die erste und zweite Hose im Vergleich zu den anderen und zueinander relativ schmutzig bzw. sauber, die dritte und vierte im Vergleich augenfällig zerknüllt bzw. zerrissen. Das Prinzip entspricht also nicht dem der Darstellung einiger der relativen Adjektive als Gegensatzpaare innerhalb einer Zeichnung, sondern die vier Bilder einer Seite - prinzipiell unabhängig voneinander – relativieren einander. Im Zusammenhang mit den Bezugspunkten der Bilder untereinander wurde zudem überprüft, ob bestimmte Zusammenstellungen von Darstellungen möglicherweise Missverständnisse provozieren könnten. So war bspw. das Item ‘Bart’ von der Zeichnerin zunächst als Kinn mit Vollbart und der Distraktor ‘Kinn’ als Kinn mit angedeuteten Bartstoppeln dargestellt worden. Diese hätten ebenfalls als Bart gelten können und lenkten überdies unnötig von dem gemeinten Gegenstand ‘Kinn’ ab. Die Darstellung musste als missverständlich gelten, woraufhin die Zeichnung des Kinns ohne Bartstoppeln ausgeführt wurde.

Fragwürdig war zudem die Relation der Größenverhältnisse des Dargestellten untereinander. Es stellte sich die Frage, ob die Nichteinhaltung einer einigermaßen maßstabsgetreuen Darstellung möglicherweise verwirrend für die Testpersonen sein könnte. Am Beispiel des Items ‘Haus’ mit den Distraktoren ‘Kirche’, ‘Hand’ und ‘Maus’ lässt sich dies verdeutlichen: Ein Haus, welches ebenso groß ist wie eine Kirche, müsste mindestens eine Villa, wenn nicht ein Schloß sein; oder umgekehrt, eine Kirche, die ebenso klein ist, wie ein Haus, kann bestenfalls als Kapelle bezeichnet werden. Hinzu kommen die anderen beiden Distraktoren ‘Hand’ und ‘Maus’, welche sich selbstverständlich nicht maßstabsgetreu im Verhältnis zu den Gebäuden abbilden lassen. Eine Abstimmung der Größenverhältnisse aufeinander ließe sich also nicht einheitlich einhalten. Selbst eine Angleichung größenmäßig ähnlicherer Gegenstände, wie ein Haus und eine Kirche, lässt sich praktisch nicht realisieren, denn der Raum, den eine kleine Abbildung zur Verfügung stellt, reicht für deutliche Unterschiede nicht aus. Daher wurde eine Angleichung verworfen.

Das Vorhaben, die abgebildeten Gegenstände gestalterisch der Erlebniswelt der Kinder entsprechen zu lassen, zog Änderungen einzelner Zeichnungen nach sich. Es sollte ausgeschlossen werden, dass die Entscheidung der Testperson für eine der Darstellungen dadurch beeinflusst werden könnte, dass sie mit dem Abgebildeten nichts verbinden kann, was von der Intention des Tests ablenken würde. Das Wort ‘anspitzen’ war beispielsweise als ein am Tisch festgeschraubter Anspitzer mit einer Kurbel dargestellt worden. Es kann nicht davon ausgegangen werden, dass jedem Kind ein solches Gerät geläufig ist. Stattdessen erschien es sinnvoller, einen kleinen Handanspitzer darzustellen, wie ihn jedes (deutsche) Kind in seiner Federtasche hat. Ein anderes Beispiel in diesem Zusammenhang ist das Wort ‘bohren’, das zunächst mit einem altmodischen Handbohrer dargestellt worden war, dessen Vorkommen in jedem Haushalt nicht vorausgesetzt werden kann. Die Wahrscheinlichkeit, dass einem Kind heute eine Elektrobohrmaschine geläufig ist, ist hingegen höher einzuschätzen. Die Zeich-

nung wurde entsprechend umgestaltet. Ein weiteres Beispiel, welches allerdings nicht in Verbindung mit einer zeitgemäßen Darstellung, sondern mit bestimmten Kenntnissen der Testpersonen zu sehen ist, betrifft das mit Thermometern dargestellte Item ‘warm’ und dessen Distraktoren ‘heiß’ und ‘kalt’. Es war fraglich, ob den Kindern die Interpretation der Zeichnungen, deren Voraussetzung die Beherrschung der Funktionsweise eines Wärmethermometers war, überhaupt möglich sein oder eine Überforderung darstellen würde.

Bei der Betrachtung der Seiten im Ganzen fiel auf, dass die Hintergründe der einzelnen Bilder teilweise unterschiedlich gestaltet waren. Die Variationen reichten von weißem, also praktisch nicht ausgestaltetem Hintergrund bis zu relativ dunklen Schraffierungen. Im Gesamteindruck wirkten die dunkel unterlegten Zeichnungen hervorgehoben. Es wurde jedoch als notwendig erachtet, die Zeichnungen gleichrangig nebeneinander stehen zu haben, um die Entscheidungen der Testpersonen, besonders im Fall von Unsicherheiten, nicht zu beeinflussen, indem einzelne Bilder möglicherweise ins Auge stechen und sich für die Wahl aufdrängen. Die Hintergründe wurden entsprechend aufeinander abgestimmt.

Entscheidend für die zeichnerische Darstellung der Testitems ist, wie bereits erwähnt, die Eindeutigkeit des Bildes, d. h. das Bild muss tatsächlich das zeigen, was der Begriff besagt, denn nur dann kann die Antwort einer Testperson objektiv bewertet werden. Zweideutigkeiten können beispielsweise auftauchen, wenn Item und Darstellung in Hinblick auf die Verwendung von Singular bzw. Plural nicht übereinstimmen. Das Item ‘Zahn’ wurde beispielsweise zunächst durch zwei gezeichnete Zähne dargestellt, von denen im Sinne der Eindeutigkeit einer aus der Zeichnung entfernt wurde. Bedenkenswert ist unter der Fragestellung der Eindeutigkeit der Interpretation auch, ob Inhalte von Kindern im Grundschulalter womöglich anders empfunden werden als von Erwachsenen. So wurde das Item ‘frei’ zunächst als Situation dargestellt, in der ein Hund sich von der Leine losreißt und seinem Herrchen, einem Kind, davon läuft. Fraglich ist, ob ein Kind diese Situation mit dem grundsätzlich positiven Begriff ‘frei’ belegen oder nicht vielmehr mit dem zurückbleibenden Kind mitfühlen würde. Die Zeichnung wurde umgestaltet, ohne ein Kind als mögliche Identifikationsfigur vorkommen zu lassen.

Feinheiten im Gebrauch der deutschen Sprache müssen bei der bildlichen Darstellung von Begriffen Berücksichtigung finden, da es sonst zu Unklarheiten kommen kann. So darf die Darstellung des Begriffes ‘krank’ nicht durch einen gebrochenen Arm erfolgen, da dies mit dem Terminus ‘verletzt’ beschrieben werden müsste, wohingegen jemand, der eine Verletzung erlitten hat, nicht im Bett liegend dargestellt werden sollte, da es sonst zu Verwechslungen mit ‘krank’ kommen könnte. Entscheidend bei der zeichnerischen Darstellung ist demnach die Abgrenzung zu anderen, im Sprachgebrauch ähnlich verwendeten Begriffen,

um unbeabsichtigten Ähnlichkeiten, die nicht unter die Kategorie inhaltlicher Distraktorähnlichkeit fallen, vorzubeugen.

1.5 Testheft und Ergebnisprotokoll

Der Wortschatztest setzt sich demnach aus drei Subtests zusammen, welche die drei Sprachmodalitäten Gebärdensprache, Schriftsprache und Lautsprache testen. Jeder dieser Subtests beinhaltet 54 Items, die die drei Wortarten Nomen, Verben, Adjektive in dieser Reihenfolge in gleichgewichteten Blöcken von je 18 Items berücksichtigen. Diese drei durch die Wortartenblöcke entstehenden Abschnitte eines jeden Subtests werden im Folgenden als ‚Subtestteile‘ bezeichnet. Die Subtests unterscheiden sich somit in den Sprachmodalitäten, nicht aber in der Auswahl der Items. Hinzu kommt ein leicht lösbares Übungsbeispiel zu Anfang des Tests (Nomen ‘Auge’). Dieses geht nicht in die Wertung ein, sondern dient ausschließlich der Einführung in das Verfahren und der Absicherung, dass die Vorgehensweise von der Testperson verstanden wurde.

Der Test besteht aus zwei Testheften (s. Beispielseiten im Anhang), wobei Testheft I für die Subtests Gebärden- und Lautsprache, Testheft II für den Subtest Schriftsprache verwendet wird. In Testheft I befinden sich vier Schwarzweiß-Zeichnungen (mit Ausnahme der in Farbe dargestellten Farbwörter) pro Seite (DinA5/Querformat), von denen eine das gefragte Item und die drei übrigen die Distraktoren abbilden. In Testheft II befinden sich auf der linken Seite eine Zeichnung, die das gefragte Testitem abbildet, und auf der rechten Seite vier in Druckschrift geschriebene Wörter, von denen eines das Testitem bezeichnet und die drei übrigen die Distraktoren. Für die Anordnung der Bilder bzw. Wörter auf der einzelnen Seite wird das System der formal bewährten Vierfelder-Wahltafeln des PPVT (vgl. Günther 1991, 192) übernommen. Die vier Felder sind dabei durch ein Kreuz voneinander abgegrenzt. Die Abb. 5-B 1.5 stellt dieses Aufbauschema dar.

Abb. 5 B-1.5: Aufbauschema Testheft; Vierfelder-Wahltafel

I. Gebärdenzeichen*	Abbildung 1	Abbildung 2*
	Abbildung 3	Abbildung 4
II. Abbildung*	Wortkarte 1	Wortkarte 2*
	Wortkarte 3	Wortkarte 4
III. Gesprochenes Wort*	Abbildung 1	Abbildung 2*
	Abbildung 3	Abbildung 4

Die Position, die das jeweilige Testitem innerhalb der vier möglichen Positionen einer Seite einnimmt, ergibt sich nach folgender Systematik: Jede Position wird innerhalb einer Wortart gleich häufig von dem Testitem belegt. D. h., das Item belegt zwei der vier Positionen je fünfmal und die übrigen zwei Positionen je viermal. Die Verteilung unterliegt dabei keiner Regelmäßigkeit, nach der sich ein aufmerksames Kind womöglich richten könnte, sondern dem Zufallsprinzip. Demnach ergibt sich eine Systematik gleichgewichteter Verteilung bei unregelmäßiger Reihenfolge. Dieses Prinzip gilt sowohl für das Testheft I als auch für das Testheft II, wobei sich die Anordnung in den Testheften nochmals voneinander unterscheidet, indem sie erneut dem Zufallsprinzip unterzogen wurde. Dies beugt bei Anwendung beider Hefte zusätzlich einem Erinnerungseffekt vor.

Die Antworten der Testpersonen werden in Ergebnisprotokolle eingetragen. Die folgende Abbildung verdeutlicht deren Aufbau:

Abb. 6-B 1.5: Schema des Ergebnisprotokolls Pilotversion

Testperson:

Alter:

Datum:

Wortschatztest Teil:

Item		Distraktor			weiß nicht	K. wiederholt			T. wiederholt	Bemerkungen
Nr.	Wort	I	II	III		G.	L.	M.		
1	Mutter	Bauer	Mädchen	Motte						
2	Auto	Bus	Laster	Zug						
3	Haus	Hand	Kirche	Maus						
4	Kopf	Junge	Topf	Zopf						
5	Straße	Blase	Strauß	Weg						
6	Bett	Brett	Decke	Sofa						
7	Angst	Freude	Trauer	Wut						
8	Stein	Kaffee	Stange	Stirn						
9	Tier	Bier	Tor	Verkehr						
10	Bart	Bett	Kinn	Post						
11	Zahn	Hahn	Mund	Zaun						
12	Leute	Herde	Mensch	Zahn						
13	Zahl	Buchstabe	Form	Zeichen						
14	Geld	Feld	Gold	Salz						
15	Kasten	Kasse	Kissen	Schrank						
16	Holz	Brot	Gold	Höhle						
17	Arbeit	Abend	Eimer	Kaffee						
18	Schloß	Haus	Kirche	Schlüssel						

Auf einem Protokoll-Bogen befinden sich jeweils 18 Items mit den dazugehörigen Distraktoren, sodass pro Subtest und Kind drei Bögen, für jede Wortart einer, benötigt werden. In der Kopfzeile der Protokoll-Bögen werden Name, Alter (auf den Monat genau; aufgerundet), Testdatum und die Angabe, um welchen Subtest es sich handelt (Bsp.: Gebärdensprache/Nomen) eingetragen. Neben den vier Items pro Zeile (Testitem und vier Distraktoren) können folgende Kategorien angekreuzt werden: 'weiß nicht'; 'Kind wiederholt' (differenziert in Gebärde, Lautsprache, Mundbild); 'Testleiter wiederholt'; 'Bemerkungen'.

Die Auswahl der neben der Wortwahl in das Ergebnisprotokoll aufgenommenen Kategorien erfolgte nach dem angenommenen Wert der Informationen in der Auswertungsphase der Daten. Die Darstellung des weiteren Projektverlaufs wird verdeutlichen, dass aus unterschiedlichen Gründen nicht alle erhobenen Kategorien in die Auswertung eingeflossen sind. Dennoch werden an dieser Stelle gemäß des Verlaufs der Testentwicklung auch die Aspekte erwähnt, welche letztlich keinen Einfluss auf die Weiterbearbeitung des Testinstruments genommen haben.

Von den vier Wörtern - einem Item und drei Distraktoren - wird dasjenige, das von der Testperson ausgewählt wurde, angekreuzt. Gibt das Kind an, keine Antwort zu wissen, oder fragt es nach, sodass der Testleiter wiederholen muss, wird ein Kreuz in die entsprechenden Kästchen gesetzt. Muss der Testleiter mehrmals wiederholen, wird in Klammern hinter das Kreuz in dieser Spalte '2x' geschrieben. Mehr als zwei Wiederholungen, also dreimaliges Anbieten insgesamt, sollen nicht stattfinden, da erstens nicht anzunehmen ist, dass darüber hinaus ge-

hende Wiederholungen dem Kind helfen könnten, und zweitens die Durchführungsdauer dadurch weiter erhöht werden würde, was die Konzentration der Testperson unnötig beanspruchen würde.

Die Spalte 'Kind wiederholt' teilt sich in drei Kategorien auf, die die Sprachmodalität bezeichnen, in der wiederholt wurde, also Gebärde (G.), Lautsprache (L.) und Mundbild (M.). Es kann hier vorkommen, dass mehr als eine der Kategorien für ein Item angekreuzt werden, wenn nämlich das Kind mehrmals oder gleichzeitig auf unterschiedliche Art wiederholt. Diese Aufteilung wurde in der Annahme vorgenommen, dass möglicherweise Hinweise auf den Zugang der jeweiligen Testperson zu den Sprachmodalitäten bzw. zur Lösung der Aufgaben entnommen werden können, der sich zudem bei den einzelnen Subtests unterscheiden kann.

Für alles zusätzlich Erwähnenswerte bzw. Außergewöhnliche wurde die Kategorie 'Bemerkungen' eingerichtet. Innerhalb des ersten Testdurchlaufs sind diverse hier verzeichnete Anmerkungen in Hinblick auf die Testentwicklung zu verstehen, insbesondere auf den Umgang der Testpersonen mit dem Material sowie auf die Distraktorenanalyse. Es wurde beispielsweise verzeichnet, wenn ein Kind das Item gebärdend wiederholte, dabei jedoch nicht das geltende Gebärdenzeichen verwendete, sondern seine Gebärde der jeweiligen Zeichnung anpasste und beispielsweise den Holzstapel manuell beschrieb. Es wurde angenommen, dass dies u. U. Hinweise auf die Qualität der Zeichnungen oder auf den Zugang der Testpersonen zu den Lösungen geben könnte. Insofern wurde diese erste Version des Ergebnisprotokolls zu Zwecken der Testentwicklung konzipiert und war daher vorläufiger Art.

Grundsätzlich wird jede falsch gegebene Antwort, also die Wahl eines Distraktors statt des Testitems, als Fehlerpunkt verzeichnet. Darüber hinaus gibt es in Zusammenhang mit der Fehlerbewertung bestimmte Sonderfälle, die hier erläutert werden:

Muss das erste Testitem auf Grund methodischer Verständnisschwierigkeiten der Testperson als zusätzliche Übungsaufgabe hinzugezogen werden, geht diese als Fehler in die Bewertung ein. Dies erschien im Rahmen des ersten Testdurchlaufs insofern unumgänglich, als die Verwendung der Aufgabe der Testperson als zusätzliche Übung angekündigt werden muss, und die Antwort, die das Kind gibt, noch nicht gewertet werden kann. Das Vorkommen von Verständnisschwierigkeiten, die ein zweites Beispiel notwendig machten, war angesichts der simplen Methode, mit der die Aufgaben gelöst werden, überraschend. Dementsprechend war nur eins von elf Kindern betroffen. Es wurde dennoch bereits zu diesem Zeitpunkt die Ergänzung durch ein zweites Übungsbeispiel erwogen, was Bewertungsschwierigkeiten ausräumen würde.

Das Prinzip der Auswahl-Antwort-Aufgabe birgt die grundsätzliche Möglichkeit von Zufallstreffern. Im vorliegenden Fall, der Wahl zwischen vier Bildern, beträgt die Wahrscheinlichkeit, zufällig richtig zu tippen, 25%. Diese Komponente warf dort Schwierigkeiten auf, wo Testpersonen angaben, die Antwort nicht zu wissen, was ihnen in der Testinstruktion der ersten Version zugestanden wurde. In die Wertung ging eine solche Antwort nicht als ganzer Fehler ein, sondern wurde mit 0,75 Fehlerpunkten bewertet. Dem zu Grunde lag die pädagogische Erwägung, den wahrheitsgemäß antwortenden Testpersonen die Zufallstrefferquote von 25% in Rechnung zu stellen, um sie den ratenden Testpersonen gegenüber nicht zu benachteiligen.

Diese pädagogisch motivierte Methode richtet sich nicht nach den Vorgaben der klassischen Testdiagnostik, ist aber schlüssig und grundsätzlich vertretbar. Dennoch wurde geplant, für eine Folgeversion des Tests zu überprüfen, ob die Möglichkeit, Nichtwissen einer Aufgabe anzugeben, beibehalten werden sollte. Es zeichnete sich bereits ab, dass die Instruktion im Sinne einer eindeutigen Bewertbarkeit u. U. würde dahingehend verändert werden müssen, den Testpersonen in jedem Fall eine Wahl abzuverlangen, selbst wenn Unsicherheit bestünde. Andernfalls hätte ein zusätzliches Testinstrument zur Anwendung kommen müssen, welches identifiziert, an welcher Stelle eine Testperson wählt, ohne sich der Antwort sicher zu sein, oder ob sie nur aus Lustlosigkeit angibt, die Antwort nicht zu wissen, welche Motivation also hinter den Antworten steht. Dies musste allein aus Gründen der Testökonomie ausgeschlossen werden, denn der Ertrag der Testdurchführung des Wortschatztests stünde dann in keinem angemessenen Verhältnis mehr zu dem zu seiner Entwicklung betriebenen Aufwand.

1.6 Testsituation und -durchführung¹⁷

Testleiter und Testperson setzen sich wahlweise einander gegenüber oder über Eck an einen Tisch. Dem Kind liegt das Testheft vor, dem Testleiter die Ergebnisprotokolle und ein Stift. Die Protokolle sollen vom Testleiter so ausgefüllt werden, dass das Kind es nicht sieht und verfolgen kann. Das Kind kann sich dadurch ablenken oder verunsichern lassen. Es empfiehlt sich, eine Unterlage bereit zu legen, auf der geschrieben werden kann, sodass die Protokoll-Bögen nicht auf dem Tisch liegen und die Testperson sie nicht sieht.

¹⁷ Dem Leser mag auffallen, dass dieses Kapitel B 1.6 im Präsens statt wie die vorangehenden Kapitel des Teil B in der Vergangenheitsform verfasst wurde. Diesem Wechsel liegt der Umstand zu Grunde, dass die voranstehenden Kapitel Abschnitte des Projektverlaufs bzw. Konzeptionen schildern, die z. T. vorläufiger Natur sind, während die Testanweisung bis zum Abschluss des Projekts beibehalten wurde und sich ihre Schilderung in der Art eines Anwenderhandbuches in der Sache anbot.

Die Wahl des Raumes, in dem getestet wird, soll sich danach richten, Ablenkungen so weit wie möglich auszuschließen. Liegen die Fenster eines Raumes beispielsweise zu einem Schulhof oder einem Fußweg gewandt, oder befinden sich in dem Raum attraktives Spielzeug, Bilder o. ä., kann die Aufmerksamkeit der Testperson leicht abgelenkt und damit beeinträchtigt werden. Gut geeignet ist ein möglichst neutral eingerichteter, wenn möglich etwas abgelegener Raum, in dem keine äußeren Bedingungen es der Testperson erschweren, seine Konzentration ausschließlich auf das Testgeschehen zu richten. (Sorgsamkeit bei der Auswahl des Raumes stellt sich als besonders bedeutsam heraus bei Kindern, die ohnehin Konzentrationsschwierigkeiten mitbringen, und denen ablenkende Impulse im Raum es zusätzlich erschweren, mit der Aufmerksamkeit beim Testgeschehen zu bleiben. Auch für lustlose Kinder mit mangelnder Motivation bieten äußere Reize eine willkommene Ablenkung, die mit der richtigen Wahl der Umgebung, in der der Test stattfindet, von vornherein im Interesse eines objektiven Ergebnisses ausgeschlossen werden müssen.) Die Lichtverhältnisse schließlich sollen derart beschaffen sein, dass Testleiter und Testheft für das Kind einwandfrei sichtbar sind, sodass unter diesem Aspekt keine Beeinflussung des Ergebnisses auftreten kann. Auch für den Testleiter ist es wichtig, mögliche Gebärden oder Mundbilder der Testperson genau erkennen zu können.

Ein weiterer Aspekt, der reflektiert werden muss, ist die Tageszeit, zu der der Test durchgeführt wird. Hier wird in den meisten Fällen der Testpraxis kaum von der 'Wahl' der Zeit gesprochen werden können, da sowohl Testleiter als auch Kind, je nach institutionellem Rahmen, in dem der Test stattfindet, eingebunden sind in Schulalltag, Raumverfügbarkeit, Freizeit- und sonstige Termine, nach denen sich der Testtermin als außerordentlicher Termin richten muss. Es sollte dennoch darauf geachtet werden, die Testpersonen nicht aus Situationen heraus zu holen, denen ihr besonderes Interesse gilt, denn dies wird sich mit einiger Wahrscheinlichkeit negativ auf Aufmerksamkeit und Motivation auswirken.

Das Testheft I mit vier Abbildungen auf jeder Seite kommt in den Subtests Gebärden- und Lautsprache zum Einsatz. Um einen Wiedererkennungseffekt der Zeichnungen auszuschließen, sollen diese beiden Subtests in der Anwendung nicht aufeinander folgen, sondern der Subtest Schriftsprache mit dem Testheft II dazwischen eingesetzt werden. Zu Anfang soll mit dem Gebärdensprachtest gearbeitet werden, da dieser spielerischen Charakter hat und somit positiv motiviert. Zudem wurde angenommen, dass hier für Gehörlose auf Grund ihres relativ leichten und positiv motivierten Zugangs zur Gebärdensprache die geringsten Schwierigkeiten auftreten würden, was ebenfalls die 'Eisbrecherfunktion' des Testeinstiegs unterstützt. Die Anwendung der Subtests in der Reihenfolge Gebärden-, Schrift- und Lautsprache entspricht der hypothetischen Annahme einer Steigerung des Schwierigkeitsgrades.

Sind die äußeren Umstände der Testdurchführung geklärt bzw. bereitgestellt, beginnt der Test mit einer Instruktion seitens des Testleiters an die Testperson. Ziel dieser Anweisung ist es, der Testperson vor Beginn der Testdurchführung Informationen über den Test bzw. über die Beantwortung der Aufgaben zu geben. In dieser Anweisung sollten auch Übungsbeispiele enthalten sein, durch deren Lösung die Testperson zeigen kann, dass die Instruktion verstanden wurde und befolgt werden kann. Für die Formulierung dieser Anweisung gelten folgende Richtlinien: Mehrdeutige oder den Testpersonen evtl. unbekannte Begriffe sollen vermieden werden. Anweisungen, Aussagen und Fragen sollen möglichst nur positiv verwendet werden. Die Instruktion soll nicht zu lang und nicht zu kurz ausfallen.

Diese methodische Erklärung wird jedem einzelnen Subtest des Wortschatztests voran gestellt. Sie wird beim ersten Testdurchgang etwas ausführlicher ausfallen als beim zweiten oder dritten, denn die Methode ändert sich prinzipiell nicht und das Kind erinnert sich i. d. R. an den vorangegangenen Testdurchlauf. Den Testaufgaben geht ein Übungsbeispiel voran, anhand dessen das Kind die Vorgehensweise einmal erproben und der Testleiter feststellen kann, ob das Prinzip verstanden wurde. Bevor jedoch die erste Seite aufgeschlagen wird und der Blick des Kindes auf die vier Zeichnungen bzw. vier Wörter und ein Bild des Übungsbeispiels gezogen werden kann, erfolgt die Erklärung wie folgt (hier beispielhaft für den Subtest Gebärdensprache):

"Hier hast du ein Heft, in dem wirst du auf jeder Seite vier Zeichnungen sehen. Ich werde eine Gebärde machen, du musst sie dir gut anschauen und überlegen, welches Bild dazu passt. Zuerst machen wir ein Beispiel zum Üben." (Testleiter schlägt die erste Seite des Testheftes auf. Testleiter gebärdet 'Auge'.)

Zeigt das Kind auf das richtige Bild, kann die Erklärung direkt fortgesetzt werden. Sollte jedoch deutlich werden, dass der Testperson die Vorgehensweise noch nicht ganz klar ist, muss die Erklärung noch einmal wiederholt werden. Hat das Kind dann noch Verständnisschwierigkeiten, wird das erste Testitem 'Mutter' als zusätzliches Übungsbeispiel herangezogen. Es geht in diesem Falle als Fehler in die Auswertung ein.

Die Erklärung sollte um folgende Anmerkungen ergänzt werden: "Es geht nicht darum, besonders schnell darin zu sein, die Aufgaben zu lösen. Du kannst dir die Zeit lassen, die du brauchst. Wenn du einmal keine Lösung weißt, dann sagst du, 'ich weiß es nicht'. Das ist nicht schlimm, niemand kann alles wissen. Hast du alles verstanden?"

Gibt das Kind an, alles verstanden zu haben, kann begonnen werden. Es sollte vorher allerdings noch ausdrücklich auf Folgendes hingewiesen werden: "Wenn du noch Fragen hast,

musst du sie jetzt stellen, bevor wir angefangen haben. Während des Testens kann ich sie dir nicht beantworten. Hast du noch Fragen?"

Verneint das Kind, kann begonnen werden mit den Worten: "O.K. - Fangen wir an!"

Der Testleiter trägt die Anfangszeit in den ersten Protokollbogen ein und der Test beginnt: Das Kind wählt nun nach dem Prinzip der Antwort-Auswahl-Aufgabe eine von vier Zeichnungen aus und dies 54 mal. Der Übergang der Wortarten geschieht im Prinzip fließend und wird der Testperson nicht extra angezeigt. (Selbstverständlich kann das Kind den Wechsel theoretisch selbstständig feststellen. Einen Hinweis bekommt es möglicherweise dadurch, dass der Testleiter die Auswertungsprotokolle – ein Bogen pro Wortart – nach jeweils 18 Items umblättert.) Am Schluss wird die Uhrzeit nach der letzten gegebenen Antwort eingetragen, sodass die Dauer des Tests errechnet werden kann.

Kommt nach 14 Tagen der Subtest Schriftsprache zur Anwendung, so muss die Erklärung zum Verfahren erfahrungsgemäß lediglich teilweise wiederholt werden. Der Testleiter beginnt nun mit der Frage: „Erinnerst du dich noch an das letzte Mal?“

Wenn das Kind verneinen sollte, kann die Erklärung vom ersten Mal mit den spezifischen Modifikationen für den Subtest Schriftsprache wiederholt werden. Bejaht die Testperson, wird die Einführung fortgesetzt mit den Worten: „Dieses Mal geht es ein bisschen anders.“ Die erste Seite des Testhefts mit dem Übungsbeispiel wird aufgeschlagen. „Hier siehst du ein Bild (Testleiter deutet auf die linke Seite) und hier vier Wörter. Du musst überlegen, welches Wort passt zu dem Bild.“

In der Regel deutet das Kind nun auf die von ihm angenommene Lösung. Andernfalls fordert der Testleiter dazu auf, das Übungsitem zu bearbeiten. Wird die richtige Lösung gezeigt, kann davon ausgegangen werden, dass das Verfahren verstanden wurde. Sollte jedoch deutlich werden, dass der Testperson die Vorgehensweise noch nicht klar ist, muss die Erklärung noch einmal wiederholt werden. Hat das Kind dann noch Verständnisschwierigkeiten, wird das erste Testitem ‘Mutter’ als zusätzliches Übungsbeispiel herangezogen. Es geht in diesem Falle als Fehler in die Auswertung ein.

Die Erklärung sollte auch beim zweiten Testdurchlauf nochmals um folgende Anmerkungen ergänzt werden: "Es geht auch heute nicht darum, besonders schnell darin zu sein, die Aufgaben zu lösen. Du kannst dir die Zeit lassen, die du brauchst. Wenn du einmal keine Lösung weißt, dann sagst du, ‘ich weiß es nicht’. Das ist nicht schlimm, niemand kann alles wissen. Hast du alles verstanden?"

Bestehen keine weiteren Verständnisschwierigkeiten, wird begonnen mit den Worten: „O.K., fangen wir an.“

Bei Anwendung des dritten Subtests Lautsprache kann die Kenntnis des Verfahrens sicher vorausgesetzt werden. Es wird ausschließlich auf die subtestspezifische Vorgehensweise hingewiesen mit den Worten: „Ich sage ein Wort, du musst genau ablesen und auswählen, welches Bild dazu passt.“

Die in B 1.5 erwähnten Kategorien auf dem Protokoll-Bogen werden, mit Ausnahme der Kategorie ‘Bemerkungen’, angekreuzt. Es empfiehlt sich, die Eintragungen direkt nach den einzelnen Aufgaben vorzunehmen, da man sich im Nachhinein kaum an jedes Detail erinnern kann. Besonders die Modalität, in der die Testperson bestimmte Items ggf. wiederholt, musste genau beobachtet und sofort eingetragen werden, da es unmöglich ist, sich solche Details durch das Testgeschehen hindurch bis zum Schluss zu merken. Die ‘Bemerkungen’ betreffend würde eine ausformulierte Eintragung allerdings zu viel Zeit beanspruchen und den Testablauf stören. Es reicht hier aus, sich ein Zeichen oder ggf. ein Stichwort zu vermerken, an welcher Stelle bzw. was für eine Bemerkung zu machen ist, und diese nachträglich einzutragen.

Das Umblättern der Seiten des Testheftes kann – je nachdem, wie es die Situation ergibt – vom Testleiter oder von der Testperson selbst ausgeführt werden. Beide Varianten haben Vor- und Nachteile: Blättert das Kind selbst, so ist es in Form dieser Bewegung am Testgeschehen beteiligt, sozusagen aktiv. Das reine Schauen, Überlegen, Tippen und dabei Stillsitzen kann – je nach Temperament des Kindes und Gesamtsituation – u. U. langweilig sein und Ungeduld fördern. Ein weiterer Vorteil für den Testleiter liegt darin, die Hände frei zu haben zum Schreiben und nicht mit dem Stift in der Hand blättern zu müssen. Die Erfahrung der Testleiterinnen zeigt allerdings, dass es notwendig sein kann, selbst zu blättern, denn so wird Seite für Seite aufgeschlagen, während die Kinder u. U. zu weit vor blättern und abgelenkt werden, indem sie sich die Bilder ansehen, die erst später an der Reihe sind. Dies kann aus motorischer Ungeschicklichkeit geschehen oder weil es dem Kind als willkommene Ablenkung dient.

Schweift ein Kind hin und wieder mit seiner Aufmerksamkeit ab, stellt Fragen oder macht Bemerkungen über die Bilder, bleibt aber im Prinzip bei der Sache, sollte der Testleiter dies geschehen lassen, ohne näher darauf einzugehen, und dann fortfahren. Sollte ein Kind sehr lange bei einzelnen Aufgaben verweilen und sich schwer dabei tun, eine Entscheidung zu treffen, kann als indirekte Aufforderung zum Weitermachen das Item noch einmal wiederholt werden. Nimmt sich das Kind jedoch mehrmals zu viel Zeit, muss es zur Entscheidung gedrängt werden. Dies ist besonders in Hinblick darauf wichtig, dass die Dauer der Konzen-

tration Grenzen hat, die durch Unterbrechungen nicht erweitert werden. Ein hinausgezögerter Testablauf wird insgesamt länger, was den letzten Abschnitt für das Kind zusätzlich erschwert.

Sollten zwei oder alle drei Subtests zur Anwendung kommen, muss eine Unterbrechung von jeweils 14 Tagen zwischendurch eingehalten werden. Dieser Zeitraum ist unter zwei Aspekten notwendig: Zum einen muss, da der Itemkorpus in den Subtests nicht variiert, sich also die Items bei mehrfacher Anwendung wiederholen, eine reine Wiedererkennung bei dem Kind vermieden werden. Dies fällt insbesondere bei der zweimaligen Verwendung des Testheftes I (Subtests Gebärden- und Lautsprache) ins Gewicht. Denn die Testperson soll ja ihre Wortschatzkenntnis der angebotenen Sprachmodalität unter Beweis stellen und nicht die Wiedererkennung der Zeichnungen. Zum Zweiten sollte die Motivation der Testpersonen auch über die Anwendung mehrerer Subtests erhalten bleiben und nicht durch eine schnelle Folge von Testsituationen beeinträchtigt werden.

2. Pilotstudie

Die Durchführung der Pilotstudie, die Darstellung und Analyse ihrer Ergebnisse sowie die Folgerungen daraus werden vorliegend detailliert dokumentiert. Diese Form detailgetreuer Dokumentation begründet sich in dem Umstand, dass die hier angestellten Überlegungen die Grundlage für die folgende Hauptuntersuchung bilden. Insofern kommt der Pilotstudie weniger ein lediglich vorläufiger als vielmehr ein grundlegender Charakter zu. Eine zusammenfassende Darstellung des Untersuchungsherganges würde die Transparenz des Zustandekommens bestimmter konstruktiver oder modifikatorischer Maßnahmen beeinträchtigen. Diese sind konstitutiv für die Hauptuntersuchung.

2.1 Stichprobe und Durchführung der Untersuchung

Mit der Erstversion des Wortschatztests wurde von November 1995 bis Januar 1996 eine Pilotstudie durchgeführt. Ziel der Untersuchung war die Gewinnung von Datenmaterial und Erfahrungswerten als Grundlage einer Modifikation und Weiterentwicklung des Testinstruments.

Die Stichprobe bildeten die beiden Versuchsklassen des Hamburger Bilingualen Schulversuchs an der Gehörlosenschule Hammer Straße. Dabei handelte es sich um eine dritte Klasse

mit sieben Kindern und eine vierte Klasse mit fünf Kindern. Da einer der sieben Drittklässler aus psychischen Gründen nicht an dem Testdurchlauf teilnehmen konnte, belief sich die Zahl der Testpersonen auf elf. Es soll an dieser Stelle darauf verzichtet werden, die biografischen Daten der Testpersonen im Einzelnen aufzuführen. Diese finden sich im Zwischenbericht der wissenschaftlichen Begleitung des Schulversuchs (Günther 1999 a). Erwähnt werden soll hier jedoch der Umstand, dass die Kinder im schulischen Umfeld durch die Zielsetzung des bilingualen Schulversuchs alle mit DGS konfrontiert und mehr oder weniger gebärdensprachkompetent sind. In vielen Fällen besteht die bilinguale Förderung bereits seit der Phase der Frühförderung in Vorschuleinrichtungen. Zusätzlich wird in den Familien mit unterschiedlichen Manualsyste¹⁸ gebärdensprachlich kommuniziert und die Eltern haben nicht an LBG¹⁹, sondern an DGS-Kursen teilgenommen. In der Zusammenschau bedeutet dies, dass die sprachliche Basis der Gruppe in besonderer Weise einen einheitlichen Hintergrund hat. Insofern hebt sie sich aus den üblicherweise vorfindlichen heterogen strukturierten Gruppierungen gleichaltriger Gehörloser hervor und stellt sich für die vorliegenden Zwecke als besonders geeignet dar. So konnte bei den bilingual erzogenen Kindern (unabhängig von Varianzen) mit Sicherheit davon ausgegangen werden, qualifizierte Aussagen über gebärdensprachliche Kompetenzen zu erhalten.

Der Test wurde gemäß den oben geschilderten Durchführungsmodalitäten in einem abgelegenen Raum der Hamburger Gehörlosenschule mit guten Lichtverhältnissen durchgeführt. Jedes der Kinder erschien im Abstand von jeweils ca. zwei Wochen zu allen drei Subtests. Auf Grund des geringen Zeitaufwands für die Durchführung eines Tests konnten die Schüler nacheinander den Unterricht kurz verlassen, um zum Testen zu erscheinen.

Jeder Testdurchlauf im Rahmen der Pilotstudie wurde mit einer Videokamera aufgezeichnet. Mit Hilfe der Filmaufnahmen konnten Informationen zu den Leistungen einzelner Testpersonen bzw. deren Verhalten sowie zum Testablauf im Nachhinein detailliert nachvollzogen werden. Aus den Folgekapiteln zur Modifikation der Erst- und späteren Zweitversion des Wortschatztests wird hervorgehen, in welcher Weise diese Informationen Einfluss genommen haben.

¹⁸ Einige Elternpaare bzw. weitere Angehörige und Freunde der Kinder, die in den Schulversuch eingebunden sind, besuchen LBG-Kurse (lautsprachbegleitende Gebärde).

¹⁹ Lautsprachbegleitende Gebärden

2.2 Testdurchführung mit hörenden Testpersonen

Im Anschluss an die Testerprobung mit den gehörlosen Kindern der Schulversuchsklassen wurden Testdurchläufe an fünf Hamburger Grundschulen mit hörenden Kindern durchgeführt. Da hörende Testpersonen nicht über Gebärdenkenntnisse verfügen, kamen nur die Subtests Laut- und Schriftsprache zum Einsatz. Die Schulen liegen in unterschiedlichen Stadtteilen, um die Erfassung einseitiger sozialer Strukturen zu vermeiden. Die Kinder besuchten gemäß der gehörlosen Versuchsgruppe und des Testkonzepts die dritte und vierte Klasse.

Diesem Testdurchlauf lagen folgende Zielsetzungen zu Grunde: Zunächst sollte der Datenpool, der die Grundlage testmodifikatorischer Maßnahmen zur Testversion I bildete, erweitert werden. Da, wie bereits erwähnt, die Grundgesamtheit gehörloser Grundschüler an deutschen Gehörlosenschulen gering ist und mit den elf Testpersonen der Pilotstudie bereits ein Anteil davon in Anspruch genommen worden war, sollten in dieser frühen Phase der Testentwicklung keine weiteren Schüler mit dem Instrument konfrontiert werden. Denn diese hätten für eine zukünftige Erhebung zur Überprüfung der Zweitversion, die bereits geplant war, nicht mehr herangezogen werden können, ohne das Testmaterial bereits zu kennen. Es konnte davon ausgegangen werden, dass sich überarbeitungsbedürftige Schwächen des Tests auch in einem Testdurchlauf mit hörenden Kindern herauskristallisieren würden. Da es zu diesem Zeitpunkt um die Identifizierung derartiger Schwächen anhand der Ergebnisse der Testpersonen und nicht um die Beurteilung der Testpersonen selbst ging, wurde die Inanspruchnahme hörender Schüler als geeignet erachtet.

Weiterhin war die Erhebung der Leistungen hörender Testpersonen notwendig, um diese als Vergleichsmaßstab für die Bewertung der Leistungen Gehörloser anlegen zu können. Zudem sollte der Ergebnispool hörender Testpersonen als Vergleichsmaterial dem gehörloser Testpersonen gegenübergestellt werden, um zu der Bezugnahme auf die Wortliste, der die Items entnommen wurden, eine Überprüfungsmöglichkeit des Konstrukts ‚Grundwortschatz‘ zu erhalten. Die zu Grunde liegende Fragestellung dabei war, ob sich bestätigen würde, dass hörende Grundschüler der dritten und vierten Klassenstufe über die angebotenen Wörter der Wortliste verfügen.

2.3 Ergebnisse der Pilotstudie

Die unter B 2.3.1 bis B 2.3.4 erläuterten Tabellen (Tab. 7a/b-B 2.3.1, Tab. 8a/b-B 2.3.2, Tab. 9a/b-B 2.3.3 und Tab. 10a/b/c-B 2.3.4), die im Anhang (Anh. 1: B 2.3.1, Anh. 1: B 2.3.2, Anh. 1: B 2.3.3 und Anh. 1: B 2.3.4) zu finden sind, dokumentieren die Leistungen der Testpersonen. Für die Erhebung mit Gehörlosen wird zusätzlich die für die jeweiligen Subtests benötigte Zeit aufgelistet. Dabei werden mit den unterschiedlichen Darstellungen verschiedene Auswertungsschwerpunkte fokussiert. Die Darstellungen bleiben mit Ausnahme der Dokumentation der Zeiten zunächst unkommentiert.

Diese Daten werden unter B 2.4 in unterschiedlichen Darstellungen unter unterschiedlichen Fragestellungen aufbereitet, um so erste Bewertungsmaßstäbe für die Testrohversion zu erhalten und auf diesem Hintergrund in einem weiteren Schritt Modifikationen vornehmen zu können. Dabei ist folgende Unterscheidung zu beachten: Die Darstellung der Fehlerverteilung Gehörloser unter B 2.4.1 und die des Ratings unter B 2.4.2 beantworten die Fragestellungen, unter denen sie erstellt wurden, direkt, d. h. sie fließen nicht weiter in die Itemdiskussionen ein. Diesen Zweck erfüllen die folgenden Darstellungen unter B 2.4.3 bis B 2.4.7.

Aus den Werten hervorgehende Beschlüsse über Testmodifikationen werden in dem Auswertungsteil B 2.5 dargestellt. Dazu wird für jede Testaufgabe einzeln eine Zusammenschau sämtlicher zur Verfügung stehender Informationen aus dem Pool der aufbereiteten Daten vorgenommen, um davon ausgehend zu den Beschlüssen für die testmodifikatorischen Maßnahmen zu gelangen. Diese Maßnahmen führen zu der zweiten Version des vorliegenden Wortschatztests.

2.3.1 Klasseninterner Vergleich der Leistungen Gehörloser

Die Tabellen Tab. 7a/b-B 2.3.1 im Anhang (Anh. 1: B 2.3.1) führen die Ergebnisse der Kinder der dritten und vierten Klasse in allen drei Subtests sowie deren Gesamtergebnisse auf. Dabei sind die drei Wortarten einzeln angegeben. Die Ergebnisse sind jeweils in der erreichten Punktzahl und in Prozent aufgeführt. Maximal konnten pro Wortart jeweils 18 Punkte, pro Subtest entsprechend 54 Punkte erreicht werden. Die Prozentangaben sind stets auf ganze Zahlen gerundet. Die einzelnen Testpersonen erscheinen in den tabellarischen Darstellungen mit dem Anfangsbuchstaben ihrer Codenamen. Diese sind für Klasse 3 Anna (A), Florian (F), Michael (M), Orhan (O), Thomas (T) und Vanessa (V) und für Klasse 4 Erik (E), Farid (F), Katharina (K), Lars (L) und Simon (S).

In der ersten Spalte lassen sich die Ergebnisse in den Subtestteilen Nomen, Verben, Adjektive sowie das Gesamtergebnis des Subtests Gebärdensprache ablesen. Die Ergebnisse sind

für jeden Subtestteil in eine Rangfolge gebracht, sodass jeweils die beste Leistung an oberster Stelle steht. Erreichten zwei oder mehr Testpersonen die gleiche Punktzahl, so wird dasjenige Kind zuerst erwähnt, das dieses Ergebnis in kürzerer Zeit erreichte (vergl. B 2.3.2). Damit wird nicht impliziert, dass das schneller arbeitende Kind eine bessere Leistung erzielte, sondern lediglich ein Kriterium zur Bildung der Rangfolge festgelegt. Identische Ergebnisse werden also unabhängig von der Zeit, und damit auch unabhängig vom Rangfolgenplatz, als gleich gut bewertet. Entsprechendes gilt für die zweite und dritte Spalte, in der sich jeweils die Leistungen in den Subtests Schrift- und Lautsprache ablesen lassen.

2.3.2 Klasseninterner Vergleich der Zeiten Gehörloser

Die Tabellen Tab. 8a/b-B 2.3.2 im Anhang (Anh. 1: B 2.3.2) führen die benötigten Zeiten jeder einzelnen Testperson für jeden der neun Subtestteile und die jeweiligen Gesamtsummen in Minuten und Sekunden auf. Da es sich im vorliegenden Fall um keinen Speedtest handelt, wurden die Testpersonen nicht zur schnellen Lösung der Aufgaben angehalten. Es wurde demgemäß auch kein zeitliches Abbruchkriterium festgelegt. Die jeweils benötigten Zeiten können also von Kind zu Kind variieren. Die Zeiten für jeden Subtestteil wurden in eine Rangfolge gebracht, wobei die jeweils schnellste Subtestteilbearbeitung an erster Stelle steht, ohne dass dies eine Wertung impliziert. Weiterhin entspricht die Auflistung dem zur Tabelle des Leistungsvergleichs erläuterten Schema.

Mit dieser Dokumentation der Dauer, die jeweils für die Bearbeitung der Subtests bzw. Subtestteile benötigt wurde, sollte zunächst eine Orientierung über den zeitlichen Aufwand, den die Durchführung des Wortschatztests bedeuten würde, erreicht werden. Es stellte sich heraus, dass für die Durchführung eines Subtests fünf bis zehn Minuten einkalkuliert werden müssen. Die einzige Ausnahme bildet das Ergebnis der Testperson Simon mit 13 Min. 06 Sek. im Lautsprachtest. Dieses Ergebnis muss auf Grund von Verhaltensauffälligkeiten der Testperson jedoch als ‚Ausreißer‘ bewertet werden.

Darüber hinaus wurde angenommen, dass über die Auflistung der im Einzelnen benötigten Zeiten zusätzliche Hinweise auf den Schwierigkeitsgrad der Sprachmodalitäten und der Wortarten gewonnen werden könnten. Es wurde davon ausgegangen, dass erhöhter Zeitbedarf auf einen erhöhten Schwierigkeitsgrad hinweisen würde. Diese Annahme konnte durch die Daten nicht bestätigt werden. So stellen sich die von den elf Testpersonen benötigten Zeiträume durch alle Subtests und alle Wortarten hindurch heterogen dar.

2.3.3 Gesamtergebnisse Gehörloser

Die folgenden Tabellen Tab. 9a/b-B 2.3.3 im Anhang (Anh. 1: B 2.3.3) stellen eine Zusammenfassung der Tabellen zum klasseninternen Vergleich der Leistungen und der Zeiten dar.

Dabei werden die Gesamtergebnisse und Zeiten der Subtests jeder Testperson sowie deren Gesamtpunktzahl des vollständigen Tests und die insgesamt benötigte Zeit dargestellt. Die Gesamtergebnisse der Subtests sind bereits in der Tabelle zum klasseninternen Vergleich der Leistungen dargestellt. Aus diesen ergeben sich die Gesamtergebnisse des vollständigen Tests.

Die erste Spalte zeigt die Ergebnisse der in den einzelnen Subtests erreichten sowie die insgesamt erreichten Punktzahlen. Die Ergebnisse sind auch hier stets zusätzlich in gerundeten Prozentzahlen angegeben. Die maximal zu erreichende Punktzahl in den Subtests liegt bei jeweils 54 Punkten, die Gesamtpunktzahl aller Subtests bei 162 Punkten. Die Ergebnisse sind in eine Rangfolge gebracht; die jeweils beste Leistung eines Subtests bzw. die beste Gesamtleistung stehen an oberster Stelle. Auch hier befinden sich die Codenamen der Testpersonen als Anfangsbuchstaben hinter den entsprechenden Ergebnissen. Entsprechendes gilt für die Aufführung der Zeiten in der zweiten Spalte.

2.3.4 Ergebnisse Hörender

Die Ergebnisse der Erhebung mit hörenden Testpersonen wurden in drei Tabellen (Tab. 10a/b/c-B 2.3.4) im Anhang (Anh. 1: B 2.3.4) dargestellt. Dabei sind die jeweiligen Werte für die Wortarten sowie die beiden Testdurchläufe Lautsprache und Schriftsprache getrennt aufgeführt.

2.4 Datenaufbereitung unter verschiedenen Fragestellungen der Testentwicklung

2.4.1 Fehlerverteilung Gehörlose

Die Tabelle Tab. 11-B 2.4.1 im Anhang (Anh. 1: B 2.4.1) zur Fehlerverteilung stellt eine Gesamtübersicht über die Anzahl der von den elf Testpersonen gemachten Fehler ohne Differenzierung nach Klassenzugehörigkeit dar. Die Darstellung wurde erstellt unter der Fragestellung, ob die hypothetisch angenommenen Schwierigkeitsabstufungen der Subtests sowie der Wortarten durch das Datenmaterial bestätigt würden.

Die oberste Zeile zeigt die Fehlerverteilung der Nomen, unterteilt in die drei Sprachmodalitäten und das Gesamtergebnis. Die zweite Zeile dokumentiert entsprechend die Fehlerverteilung der Verben, die dritte Zeile die der Adjektive. Gleichzeitig lassen sich vertikal die Fehlerverteilungen innerhalb der Sprachmodalitäten ablesen. Es werden sowohl die Fehleranzahl als auch der entsprechende Prozentsatz angegeben.

Die Übersicht verdeutlicht, dass sich die angenommene Steigerung des Schwierigkeitsgrades sowohl der Subtests als auch der Wortarten anhand der Ergebnisse bestätigt. So steigert sich die Schwierigkeit vom Gebärdensprachtest über den Schriftsprachtest bis zum schwierigsten Subtest, dem Lautsprachtest, und dies nicht nur im Gesamtergebnis, sondern in jeder der Wortarten. Ebenso steigert sich die Schwierigkeit von der Wortart Nomen über die Verben bis zu den Adjektiven, und zwar in jedem der drei Subtests.

Dabei wird sichtbar, dass die Adjektive in allen Subtests mit einigem Abstand hinter den Verben liegen, während die Schwierigkeitsabstufung zwischen Nomen und Verben geringer ausfällt. Im Vergleich der Subtests untereinander wird ein größerer Abstand des Gebärdensprachtests zum Schriftsprachtest als von diesem zum Lautsprachtest offensichtlich. So fiel den Testpersonen der Pilotstudie die Modalität Gebärdensprache mit Abstand am leichtesten und die Wortart der Adjektive mit Abstand am schwersten.

2.4.2 Rating zur Arbeitshaltung

Unter der Fragestellung, ob mit dem Testinstrument isoliert das Merkmal Wortschatz erhoben wird oder ob sich Persönlichkeitsmerkmale, vor allem die Arbeitshaltung der Testpersonen, in deren Ergebnissen widerspiegeln, wurde eine Ratingskala entwickelt. Ziel des Ratings war eine Einschätzung, ob in der Arbeitshaltung zum Ausdruck kommende Verhaltensweisen der Testpersonen regelmäßig Einfluss auf die Testergebnisse nehmen würden.

Die Vergabe von Ratingwerten wurde von den Testerinnen gemäß Beobachtungen während der Testdurchläufe vorgenommen. Diese Beobachtungen wurden vor Festlegung auf ein Niveau auf der Ratingskala jeweils anhand der Videoaufzeichnungen überprüft und zusätzlich mit der Einschätzung der bei den betreffenden Durchläufen nicht eingesetzten Testerin abgeglichen. Die Skala umfasst, wie der folgenden Abbildung entnommen werden kann, vier Kategorien und sechs Niveaupunkte:

Abb. 7-B 2.4.2: Ratingskala zur Arbeitshaltung

	1	2	3	4	5	6	
flüchtig							konzentriert
lustlos							motiviert
nervös							ruhig
spielerisch							ernst
	1	2	3	4	5	6	

Diese in das Rating aufgenommenen vier Gegensatzpaare wurden unter der Überlegung ausgewählt, den Aspekt ‚Arbeitshaltung‘ zu umreißen. Mit dem Ansatz, eine mögliche Einflussnahme von ‚Persönlichkeitsmerkmalen‘ bzw. ‚Verhaltensweisen‘ der Testpersonen zu untersuchen, wäre theoretisch die Einbeziehung diverser anderer Verhaltenskomponenten denkbar gewesen. Insofern beschränkt sich die Aussagekraft dieser Untersuchung auf die hier berücksichtigten vier Verhaltensmerkmale. Eine derartige Beschränkung muss jedoch stets vorgenommen werden, es kann niemals jede denkbare Verhaltenskomponente berücksichtigt werden. Im vorliegenden Zusammenhang einer Testsituation liegt nach Ansicht der Autorinnen insbesondere der Aspekt der Arbeitshaltung zur Untersuchung nahe.

Die Ergebnisse des Ratings, die sich unter Tab. 12a-k-B 2.4.2 im Anhang (Anh. 1: B 2.4.2) finden, wurden nicht statistisch ausgewertet, indem etwa Korrelationen zwischen Ratingwerten und Testpunktwerten errechnet worden wären. Der Umfang der zu Grunde liegenden Stichprobe reichte dazu nicht aus. Auf seiner Basis konnten die Ergebnisse jedoch für jede Testperson einzeln ausgewertet werden. So war eine Übersichtlichkeit gewährleistet, die es ermöglichte, Einschätzungen im genannten Sinne abzuleiten.

Die Ratings wurden gemäß den Tabellen zum klasseninternen Vergleich der Leistungen für die Klassen drei und vier getrennt betrachtet. Die Leistungsübersichten und die aus ihnen hervorgehenden Rangfolgen der Testpersonen bildeten die Bezugsgröße bzw. den Vergleichsmaßstab für die Auswertung der Ratings. Da es sich bei der Zuweisung eines Rangplatzes um eine leistungsbezogene Aussage handelt, ist die Bezugnahme nur klassenintern angemessen.

Diese Bezugnahme von Ratingwerten auf die Scores der Testpersonen legte eine Einteilung in vier Kategorien nahe. So wurde die erste Kategorie definiert als ‚+/-‘-Beziehung, d. h. dass sich die Einschätzung gemäß des Ratings in den Testergebnissen widerspiegelt, und zwar in jeweils positiver Ausprägung. Die Fälle, die unter diese Kategorie fielen, legten die Folgerung einer korrelativen Beziehung zwischen Ratingwert und Score nahe. Die zweite Kategorie wurde als ‚+/-‘-Beziehung definiert. In diesen Einteilungsbereich fallen die Ergebnisse der Testpersonen, die zwar eine mäßige bis schlechte Arbeitshaltung aufwiesen, dabei jedoch trotzdem relativ gute Testpunktwerte erzielten. Als Drittes wurde die Kategorie ‚-/+‘ definiert, unter der die Ergebnisse derjenigen Testpersonen zusammengefasst wurden, die trotz konstruktiver Arbeitshaltung lediglich mäßige bis sehr schlechte Scores erreichten. Die beiden letztgenannten Kategorien würden gleichermaßen die These unterstützen, dass sich keine regelmäßigen Korrelationen zwischen dem Ergebnis nach Punkten und der Einschätzung gemäß des Ratings ablesen lassen. In die vierte, die Kategorie ‚-/-‘ schließlich, wurden die Ergebnisse der Testpersonen aufgenommen, die bei mangelhafter Arbeitshaltung

entsprechend schwache Scores erzielen. Diese würden gemäß Kategorie eins die Folgerung einer korrelativen Beziehung nahe legen.

Dieser Kategorisierung gemäß teilen sich die Ergebnisse der elf Testpersonen wie folgt ein:

Abb. 8a-B 2.4.2: Kategorisierung Ergebnisse Gehörloser Klasse 3

Klasse 3: ++

Anna und Florian arbeiten sehr konstruktiv mit und erreichen dem entsprechende sehr gute Ergebnisse.

Klasse 3: -/+

Michael arbeitet sehr schlecht mit, erreicht dabei jedoch trotzdem mittelmäßige bis sehr gute Ergebnisse.

Klasse 3: +/-

Thomas, Vanessa und Orhan arbeiten relativ gut mit, erreichen dabei jedoch trotzdem nur mäßige bis sehr schlechte Ergebnisse.

Klasse 3: --

Kein Vorkommen.

Abb. 8b-B 2.4.2: Kategorisierung Ergebnisse Gehörloser Klasse 4

Klasse 4: ++

Katharina und Lars arbeiten sehr konstruktiv mit und erreichen dem entsprechende sehr gute Ergebnisse.

Klasse 4: -/+

Simon arbeitet im Schriftsprachtest sehr schlecht mit, erreicht dabei jedoch (durch die Wortarten) durchschnittlich nur relativ schlechte Ergebnisse. Im Lautsprachtest arbeitet er schlecht mit und erreicht dabei durchschnittlich gute Ergebnisse.

Klasse 4: +/-

Erik arbeitet insgesamt, Farid im Lautsprachtest mittelmäßig mit, wobei beide schlechte bis sehr schlechte Ergebnisse erreichen. Im Gebärden- sowie im Schriftsprachtest arbeitet Farid gut mit, erreicht dabei jedoch nur durchschnittlich mittelmäßige Ergebnisse. Simon arbeitet im Gebärdensprachtest gut mit, die Ergebnisse sind jedoch schlecht bis sehr schlecht.

Klasse 4: --

Kein Vorkommen.

Die Analyse zeigt zunächst, dass sich die Ergebnisse heterogen gestalten. Es zeichnet sich demnach kein einheitliches Muster korrelativer Zusammenhänge ab. Bei Berücksichtigung

des geringen Umfangs der Stichprobe lässt sich daher – in statistischer Begrifflichkeit gesprochen – sagen, dass keine Signifikanz erkennbar ist. Insoweit ergibt sich kein Anlass für eine weitere generelle Untersuchung der Fragestellung.

Im Einzelnen fallen bei der Analyse die Fälle unter der Kategorie ‚+/+‘ auf. Bei den vier unter diese Kategorie fallenden Testpersonen handelt es sich um die beiden besten Schüler der jeweiligen Klassen. Das heißt, dass ein Gleichklang von Arbeitshaltung und Score bei denjenigen Testpersonen erkennbar ist, die bei ausgezeichneter Arbeitshaltung ebenso gute Ergebnisse erzielen. Diese Testpersonen stellen jedoch unter der genannten Fragestellung keine kritischen Fälle dar. Denn der Fragestellung liegt die Überlegung zu Grunde, dass möglicherweise vorhandene (Wortschatz)Potenziale durch destruktive Verhaltensweisen der Arbeitshaltung überlagert werden könnten. Eine solche Überlagerung würde dem Ziel, das Merkmal Wortschatz isoliert, d. h. ohne Einfluss von Störfaktoren, zu erheben, entgegenstehen. Bei den ‚+/+‘-Fällen ist jedoch unzweifelhaft, dass ihre Leistungen ‚ungestört‘ zum Ausdruck gebracht werden konnten, da das Leistungsmaximum durch eine veränderte Arbeitshaltung nicht weiter steigerbar gewesen wäre.

Da die Kategorie 4 ‚-/-‘, welche eine entsprechend lineare Beziehung von Arbeitshaltung und Score widerspiegelt hätte wie Kategorie 1, von keiner der Testpersonen belegt wurde, können hier keine weiteren Hinweise auf korrelative Zusammenhänge entnommen werden.

Unter diesen Aspekten, der mangelnden Signifikanz, der Untersuchung der Fälle unter Kategorie ‚+/+‘ sowie des Nicht Vorkommens der Kategorie ‚-/-‘, wird somit davon abgesehen, die Untersuchung möglicher Wechselwirkungen zwischen Arbeitshaltung und Testergebnis weiter zu verfolgen. Die Ergebnisse der Analyse legen die Annahme nahe, dass das Testinstrument das Merkmal Wortschatz erfasst, ohne dass die Scores dabei von persönlichkeitsbedingten Störfaktoren (Arbeitshaltung) signifikant beeinflusst bzw. verfälscht werden.

2.4.3 Schwierigkeit und Trennschärfe

Entsprechend den fünf Testdurchläufen (alle drei Subtests bei Gehörlosen, zwei Subtests bei Hörenden) wurden pro Wortart fünf (insgesamt 15) Koordinatensysteme (Abb. 9a-i-B 2.4.3 und Abb. 10a-f-B 2.4.3) zur Kenntlichmachung der Position der einzelnen Testaufgaben in Bezug auf die Kriterien Schwierigkeit und Trennschärfe erstellt und im Anhang (Anh. 1: B 2.4.3a und Anh. 1: 2.4.3b) aufgeführt. Die X-Achse bildet jeweils die Schwierigkeit ‚p‘ und die Y-Achse die Trennschärfe ‚r_{it}‘ im positiven sowie unterhalb der X-Achse im negativen Bereich ab. Die Positionen der Items wurden mit Kreuzchen eingetragen und nebenstehend die zugehörige laufende Nummer des Items verzeichnet.

Die in den Koordinatensystemen veranschaulichten Trennschärfekoeffizienten (r_{it}) und Schwierigkeitsindizes (p) sind, wie unter B 4.1 näher erläutert werden wird, nicht nur jeder für sich aussagekräftig, sondern stehen in Beziehung zueinander. So besteht zwischen Schwierigkeit und Trennschärfe eine parabolische Abhängigkeit, d. h. dass die Trennschärfe bei mittlerer Schwierigkeit am größten sein kann (vergl. Lienert 1994, 31). Dies bedeutet, dass bei denjenigen Testaufgaben geringer oder hoher Schwierigkeit, die dabei trotzdem einen hohen Trennschärfekoeffizienten aufweisen, dieser höher zu bewerten ist als bei Aufgaben mittlerer Schwierigkeit.

2.4.4 Distraktoren-Ähnlichkeiten

Die bereits in B 1.3 beschriebene dreiseitige tabellarische Übersicht Tab. 6a/b/c-B 1.3 im Anhang (Anh. 1: B 1.3b) schlüsselt für jede Testaufgabe gesondert auf, in welcher Hinsicht die einzelnen Distraktoren dem Item ähneln. Die Tabellen führen in der ersten Spalte jeweils die Items einer Wortart, dahinter die zugeordneten Distraktoren und schließlich die jeweiligen Ähnlichkeiten, mit Kreuzchen gekennzeichnet, auf. Diese Übersicht kann zunächst Hinweise auf die Qualität der Testaufgabe insgesamt geben, denn je vollständiger die Ähnlichkeiten innerhalb einer Testaufgabe vertreten sind, desto höher wird die Qualität der Distraktoren und damit der gesamten Aufgabe eingeschätzt. Zudem werden Items oder einzelne Distraktoren, deren Erhalt in Frage gestellt wird oder die modifiziert werden sollen, auch anhand ihrer Ähnlichkeit zum Item beurteilt.

2.4.5 Fehleranalyse Gehörlose

Die Darstellungen Tab. 13a/b/c-B 2.4.5 im Anhang (Anh. 1: 2.4.5) schlüsseln für jedes Item und jeden Subtest auf, auf welche Distraktoren die Fehlantworten entfallen. Diese Übersicht verdeutlicht Wahlhäufungen von Items und Distraktoren. Diese können Hinweise auf Modifikationsbedarf der Testaufgaben geben. Dies können Hinweise sowohl auf die Auswahl der Wörter für den Test als auch auf deren Gestaltung sein. Dabei sind auch die Relationen der Werte untereinander aussagekräftig.

Die Tabellen, jeweils eine pro Wortart, listen die 18 Items auf, und zwar in drei Spalten pro Subtest. Jeweils unter dem in der ersten Zeile aufgeführten Item kann die gesamte Fehleranzahl durch alle drei Subtests abgelesen werden, d. h. wie oft eine Testperson bei der entsprechenden Testaufgabe einen der Distraktoren statt des Items gewählt hat. In den drei folgenden Spalten kann für die drei Subtests jeweils die Gesamtfehleranzahl abgelesen werden und dahinter, welcher der Distraktoren statt des Items gewählt wurde und wie oft.

2.4.6 Fehleranalyse Hörende

Zu dem gleichen Zweck der Verdeutlichung von Wahlhäufungen der Items und Distraktoren werden die unter B 2.3.4 angeführten Tabellen Tab. 10a/b/c-B 2.3.4 (im Anhang Anh. 1: B 2.3.4) betrachtet. In der Regel ist gemäß der insgesamt sehr guten Ergebnisse hörender Testpersonen insbesondere im Lautsprachtest die dem Item zugeordnete Wahlhäufigkeitsrate überproportional groß zu den restlichen.

Unter B 2.2 wurde als eine der Intentionen für die Erhebung mit Hörenden die Fragestellung genannt, ob sich den Daten zufolge bestätigen würde, was die Wortliste vorgab, dass nämlich hörende Grundschüler der dritten und vierten Klassenstufe über die angebotenen Wörter verfügen. Im Falle einer Bestätigung dieser Annahme kann davon ausgegangen werden, dass vorliegend mit dem Konstrukt ‚Grundwortschatz‘ operiert wird. Anhand der fast durchgängig überproportional großen Säule unter der Kennziffer ‚1‘ insbesondere im Lautsprach-, jedoch auch im Schriftsprachtest beweist sich nun, dass die Daten der Erhebung mit Hörenden die Annahme bestätigen, mit dem ausgewählten Itemkorpus im Bereich des Grundwortschatzes zu liegen.

2.4.7 Itemrangliste

Den Tabellen Tab. 14a/b/c-B 2.4.7 im Anhang (Anh. 1: B 2.4.7) liegt die Fragestellung zu Grunde, ob die ursprünglich gemäß der Wortliste angenommenen Schwierigkeitsgrade der Items sich gemäß der Ergebnisse der Pilotstudie bestätigen oder inwieweit sie davon abweichen. Auffällige Abweichungen wurden als Hinweise auf möglichen Änderungsbedarf der betreffenden Items gewertet.

Dazu wurde ausgezählt, wie oft die einzelnen Items von den elf gehörlosen Testpersonen in allen drei Testdurchläufen zusammengekommen fehlerhaft beantwortet wurden, sie also einen der Distraktoren wählten bzw. sich dafür entschieden, gar nicht auszuwählen. Daraus ergab sich gemäß der Fehleranzahl eine neue Rangliste, nämlich an erster als am leichtesten geltender Stelle das Item, welches am seltensten falsch beantwortet wurde, bis zur letzten Stelle, an der sich das am schwersten geltende, weil am häufigsten falsch beantwortete Item befindet. Alle Angaben wurden in drei Tabellen – nach Wortarten getrennt – aufgelistet. In der jeweils ersten Spalte befindet sich die – immer höher werdende – Fehleranzahl, aus der sich die in der zweiten Spalte befindlichen neuen Ränge ergeben. Dahinter steht das jeweilige Item und zum Vergleich sein ursprünglicher Rang.

Vergleicht man nun den alten und den neuen Rang, so ergibt sich eine mehr oder weniger prägnante Differenz nach oben oder nach unten. Diese befindet sich in der letzten Spalte pro Wortart. Verschiebt sich ein Item von der alten zur neuen Rangliste nach oben, so wurde die

Angabe der Rangdifferenz mit einem Pluszeichen versehen, variiert es nach unten, mit einem Minuszeichen.

Die hier dokumentierten Werte weisen, je höher desto prägnanter, zunächst darauf hin, dass die sich aus dem Datenmaterial ergebenden Schwierigkeitsgrade der Items von den gemäß der Wortliste angenommenen Werten abweichen. Die Ergebnisse im Einzelnen fließen in die Diskussion der Testaufgaben ein.

2.5 Vorgehen bei der Modifikation der Pilotversion

Zunächst wurden auf dem Hintergrund der Erfahrungen aus den Testdurchläufen im Rahmen der Pilotstudie folgende grundsätzliche zeichnerische Modifikationen beschlossen: Wie dem Bildmaterial der Pilotversion im Anhang zu entnehmen ist, wurden Items und Distraktoren mit zahlreichen Strichen, Schraffierungen und vielfach unterbrochener Linienführung ausgeführt. Die Darstellungen wirkten auf diese Weise lebendig und somit ansprechend für den Betrachter. Dies war zwar prinzipiell im Sinne des bereits erwähnten Anspruches, ein optisch motivierendes Material zu erstellen, stand jedoch einer erforderlichen Übersichtlichkeit und optischen Klarheit der Darstellungen entgegen. In diesem Sinne wurde entschieden, die Darstellungen grundsätzlich, d. h. vollständig und unabhängig von dem spezifischen Modifikationsbedarf der einzelnen Aufgabe, zu überarbeiten.

Die Linienführung wurde nun weitgehend durchgängig ausgeführt und an Stelle von Schraffierungen zur Darstellung von Plastizität oder Schattierungen wurden Grauf Flächen unterschiedlicher Helligkeit verwendet. Auch das Trennungskreuz der vier Zeichnungen wurde in durchgängiger Linienführung ausgeführt. Zudem wurden die Trennungsstriche gradliniger gehalten, im Falle des Schriftsprachtstheftes sogar mit Lineal gezogen. Außerdem wurde das Seitenformat von ursprünglich DinA5 verbreitert. Zum einen sind die Einzeldarstellungen auf diese Weise weniger dicht gedrängt auf einer Seite angeordnet. Zum Zweiten wird der Eindruck insbesondere der beiden linksseitig angeordneten Darstellungen nicht durch die Spiralbindung irritiert, da Raum gelassen wurde für unbeschriebene Randbereiche. Das Bildmaterial der Testversion I (ebenfalls im Anhang) lässt erkennen, dass diese Maßnahmen den Gesamteindruck einer Testheftseite deutlich beruhigen. Der erwünschte Effekt wurde erzielt, ohne weniger ansprechend und motivierend zu wirken als in der Erstversion. Durch die Großzügigkeit und optische Beruhigung der Darstellungen wurden die bereits in der ersten Version positiv empfundenen Elemente im Gegenteil unterstützt und verstärkt.

Bei einigen Testaufgaben hebt sich eine der vier Darstellungen auf einer Seite optisch hervor. Dies kann durch hell-dunkel-Kontraste, durch die Relation der Größenverhältnisse oder aber dadurch zu Stande kommen, dass eine der Zeichnungen szenischen Charakter hat, während die anderen Gegenständliches abbilden bzw. umgekehrt. Generell ist ein einheitlicher Eindruck einer Seite erstrebenswerter als eine – u. U. notwendige – Hervorhebung, da diese zwei negative Folgen haben kann. So kann sie durch den Umstand, dass sie dem Betrachter optisch auffällt, anziehend wirken, sodass ihre Wahl begünstigt wird. Umgekehrt kann ein schnelles Verwerfen eintreten, weil der Betrachter dadurch, dass die Abbildung sich abhebt, eine zu leichte und nahe liegende ‚Erfolgswahl‘ argwöhnt. Wenn daher die Werte einer Aufgabe dies zusätzlich nahe legten, wurde versucht, den Begriff durch einen Distraktor zu ersetzen, der sich in einheitlicher Weise mit den übrigen drei Zeichnungen darstellen lässt.

Modifikationen der zeichnerischen Darstellungen, die über diese grundlegenden, durchgängigen Maßnahmen hinausgehen, fließen in die Darstellung der Diskussion der einzelnen Testaufgaben ein.

Diese Diskussionen verfolgen jeweils ein einheitliches Vorgehensmuster. Zunächst werden die entscheidenden Werte der Trennschärfe und Schwierigkeit betrachtet. Darauf folgt die Distraktorenanalyse unter den Fragestellungen, ob alle Ablenkertypen vorhanden sind und ob aus der Analysetabelle der Erhebung mit Gehörlosen sowie aus den Säulendiagrammen der Erhebung mit Hörenden Wahlhäufungen abzulesen sind. Zudem werden die Ranglisten-differenz und die gestalterischen Elemente der Testaufgabe betrachtet. Bei der Schilderung dieser jeweiligen Zusammenschau werden die Werte nicht numerisch wiederholt, sondern ihre relative Bedeutung formuliert.

Die so zusammengetragenen Informationen werden zueinander in Bezug gesetzt, u. U. zusätzlich mit Erfahrungswerten aus den Testdurchläufen abgeglichen und schließlich interpretiert. Diese Interpretation stellt die Grundlage dar für die modifikatorischen Maßnahmen. Deren Schilderung bildet jeweils den Abschluss der Diskussion und führt zu Testversion I.

In einigen Fällen wurde, wie aus dem Text hervorgehen wird, eine Verschiebung der Positionen der Abbildungen auf einer Seite vorgenommen. Dazu sei erwähnt, dass die Systematik der gleichmäßigen Verteilung der jeweils gefragten Begriffe auf die vier möglichen Positionen einer Seite, wie sie unter B 1.5 geschildert wurde, uneingeschränkt erhalten blieb.

Grundsätzlich sind die Schwierigkeitswerte der Testdurchläufe mit hörenden Personen vergleichsweise wenig aussagekräftig, da die Testaufgaben den Kindern erwartungsgemäß leicht fielen, die Ergebnisse dementsprechend durchgängig gut ausfallen, und daher kaum Differenzierungsmöglichkeiten bestehen. Im Regelfall sind daher auch die Säulendiagramme

zu dem Aspekt der Distraktorenhäufung unauffällig und entsprechend wenig aussagekräftig. Einzelne Fälle von Distraktorenwahl verblissen auf dem Hintergrund der Stichprobengröße der Erhebung Hörender von 86 bzw. 81. (In diesen Fällen werden sie auch nicht jedes Mal gesondert erwähnt.) Sind jedoch deutliche Häufungen bei der Wahl der Distraktoren festzustellen, ist deren Aussagekraft auf dem Hintergrund des Stichprobenumfangs umso größer.

Ein entscheidendes Ergebnis der folgenden Itemdiskussionen soll an dieser Stelle bereits vorweggenommen werden. Insgesamt stellte sich anhand der Werte insbesondere des Gebärdensprachtests ein Überhang an (z. T. zu) leichten Testaufgaben heraus. Daher wurde der Itemkorpus für die überarbeitete Testversion I um weitere, relativ schwer eingeschätzte Items erweitert. So wurde neben der Intention, die Differenzierungsleistung des Tests unter den Testpersonen zu steigern, gleichzeitig Spielraum dafür geschaffen, nach einer weiteren Erhebung ggf. ungeeignete Items aus dieser zweiten Version eliminieren zu können, ohne den Umfang damit unerwünscht schmälern zu müssen.

Vorweggenommen sei zudem, dass der Test dahingehend modifiziert wurde, für Testversion II durchgängig für jedes Item einen Distraktor eines jeden Ähnlichkeitstyps bereitzustellen. Auf diesem Hintergrund werden in Einzelfällen Distraktoren ersetzt, ohne dass die Analyse der Werte dies nahe legt.

Diese Modalitäten und Vorbeschlüsse für die Modifikationen der Testaufgaben seien hiermit zu Grunde gelegt und werden, selbst bei ausdrücklicher Bezugnahme im Rahmen der einzelnen Itemdiskussionen, im Folgenden nicht wiederholt.

2.6 Analyse und Modifikation zur Testversion I

2.6.1 Diskussion der Testaufgaben/Nomen

Nomen 1: Mutter (Bauer/Mädchen/Motte)

Das Item ist im Gebärdensprachtest von allen Items das trennschärfste und schwierigste. Im Lautsprachtest mit gehörlosen Testpersonen ist es relativ gut trennscharf bei mittelmäßiger Schwierigkeit. Im Schriftsprachtest ist das Item gut trennscharf und mittelmäßig schwer. Im Lautsprachtest mit Hörenden schneidet das Item am trennschärfsten ab und ist dabei zwar leicht, jedoch von den vielen leichten Items eines der schwierigeren. Im Schriftsprachtest mit Hörenden ist es sehr leicht bei mäßiger Trennschärfe. Diese Ergebnisse überraschen angesichts des Umstands, dass das Wort an erster Stelle stehend als am leichtesten lösbar eingeschätzt wurde. In der neuen Rangliste erhält das Item einen deutlich nach unten verschobenen Rangplatz. Die Ähnlichkeiten der Distraktoren zum Item sind gleichmäßig verteilt, d. h. es ist jeweils ein Ablenkertyp vorhanden. Die Tabelle der Fehleranalyse weist eine geringe Häufung der Wahl des Distraktors 'Mädchen' durch alle drei Subtestdurchführungen mit Gehörlosen auf, was sich in den Werten der Wahlverteilung Hörender bestätigt. Diese Datenlage wird als Hinweis

darauf gewertet, dass darstellungsimmanente Elemente zur Erhöhung des Schwierigkeitsgrades geführt haben, da insbesondere bei dem Wort Mutter nicht anzunehmen ist, dass es den Testpersonen unbekannt ist.

Auf Grund der vergleichsweise sehr guten Ergebnisse wurde das Item beibehalten. Die Interpretation der Schwierigkeits- und Trennschärfewerte, die Häufung der Wahl des Distraktors 'Mädchen' und die der Itemrangliste entnommenen Hinweise führten zu einer Abänderung seiner Darstellung sowie der des Distraktors 'Mädchen'. Angenommen wurde, dass sich die beiden Zeichnungen in Bezug auf das Alter der dargestellten Personen zu wenig voneinander unterschieden. So konnte die sehr jung dargestellte Mutter möglicherweise für ein Mädchen gehalten werden bzw. umgekehrt, was Auslöser der fehlerhaften Wahl sein konnte. Daher wurde die Mutter älter und das Mädchen jünger dargestellt.

Nomen 2: Auto (Bus/Laster/Zug)

Im Gebärdensprachtest differenziert das Item, wie auch die meisten anderen in diesem Testdurchlauf, nicht, d. h. dass alle Testpersonen die Aufgabe korrekt beantwortet haben. Im Lautsprachtest mit Gehörlosen ist es das zweitrennschärfste und dabei eher leicht; im Lautsprachtest mit Hörenden ist es mittelmäßig trennscharf und dabei sehr leicht. Im Schriftsprachtest haben ebenfalls alle gehörlosen Testpersonen das Item korrekt beantwortet. Bei den hörenden Testpersonen stellt sich das Item sehr gut trennscharf und dabei sehr leicht dar. Die Ähnlichkeiten der Distraktoren zum Item bestehen in jedem der drei Fälle in Bezug auf Inhalt bzw. Zeichnung, im Falle des Distraktors 'Laster' zusätzlich in der Gebärde. Ein Distraktor mit Ähnlichkeit im Schrift- bzw. Mundbild ist nicht vorhanden. Bei der geringen Anzahl von zwei Fehlantworten insgesamt lassen sich aus der Fehleranalyse Gehörloser keine Informationen ableiten. Auch die Werte der Wahlhäufigkeiten Hörender sind unauffällig. Die Ränge der alten und der neuen Rangliste verschieben sich geringfügig.

Das Item stellt sich offenkundig als sehr leicht lösbar dar. Die Trennschärfewerte sind jedoch durchschnittlich mittelmäßig. Da das Item an zweiter Position steht und demnach planmäßig leicht lösbar sein soll, wurde es unter Einbeziehung der akzeptablen Trennschärfewerte, trotz fehlenden Distraktors mit Ähnlichkeit im Schriftbild zum Item, zunächst beibehalten.

Nomen 3: Haus (Hand/Kirche/Maus)

Das Item stellt sich in allen drei Darstellungen der Ergebnisse Gehörloser sowie in den beiden Testdurchläufen Hörender mit geringen Abweichungen voneinander gut bis sehr gut trennscharf dar und ist dabei im leichten bis mittelmäßig leichten Bereich angesiedelt. Die Distraktorenauswahl berücksichtigt alle drei Ähnlichkeiten zum Item. Die Wahlhäufigkeiten der Distraktoren sind sowohl bei den gehörlosen als auch bei den hörenden Testpersonen unauffällig und breit gestreut. Der Rangplatz der neuen Liste differiert im Vergleich zur alten lediglich um einen Punkt.

Das Item wird damit von allen Werten als gut bis sehr gut ausgewiesen und wird daher beibehalten. Zur zusätzlichen Verbesserung wurden zwei Korrekturen in der Darstellung vorgenommen. Das 'Haus' hatte zunächst Villencharakter und wurde schlichter und einfacher dargestellt, um der entsprechend schlichten Bezeichnung zu entsprechen. Zudem wurde die Anordnung der Zeichnungen auf der Heftseite verändert, da die oberen beiden Zeichnungen von Gebäuden mit Umgebungselementen szenischen Charakter haben und sich so deutlich von den unteren beiden Zeichnungen von Einzelobjekten abhoben. Der Blick des Betrachters wurde so auf die oberen Darstellungen gelenkt bzw. von den unteren, schlichteren Zeichnungen abgelenkt. Dies sollte als mögliche Hilfe, die zur Wahl des Items als einer der oben angesiedelten Zeichnungen führen könnte, ausgeschlossen werden.

Nomen 4: Kopf (Junge/Topf/Zopf)

Die Trennschärfe- und Schwierigkeitswerte differieren in den einzelnen Subtests deutlich. Im Gebärdensprachtest beantworteten alle Testpersonen das Item korrekt, was keine Aussage über Leistungsdifferenzen zulässt. Im Schriftsprachtest mit Gehörlosen hingegen ist es sehr gut trennscharf und dabei mittelmäßig schwierig. Im Lautsprachtest sind die beiden Werte demgegenüber umgekehrt angesiedelt; die Aufgabe ist die schwerste der positiv trennscharfen Items und dabei mittelmäßig trennscharf. Im Lautsprachtest mit Hörenden ist es sehr leicht, wie fast alle Items in diesem Testdurchlauf, und dabei vergleichsweise mäßig trennscharf. Im Lautsprachtest ist das Item zwar an ähnlicher Stelle im Koordinatensystem angesiedelt wie im Schriftsprachtest, schneidet jedoch im Vergleich zu den übrigen Aufgaben als zweitrennschärfstes und schwerstes Item unter den positiv trennscharfen sehr gut ab. Alle drei Ähnlichkeiten der Distraktoren zum Item sind vorhanden. Im Lautsprachtest mit Gehörlosen treten Häufungen in der Wahl der Distraktoren auf. So entfallen von acht Fehlern insgesamt fünf auf den Distraktor 'Topf' und drei auf den Distraktor 'Zopf'. Dies erklärt sich dadurch, dass sich beide vom Item ausschließlich im Anlaut unterscheiden und dabei der Anlaut des Items, das 'K', ein Verschlusslaut im Rachenbereich ist und die beiden Anlaute der Distraktoren, 'T' und 'Z', Verschlusslaute im Mundhöhlenbereich sind, so dass die Unterscheidung der jeweiligen Mundbilder voneinander außerordentlich schwierig ist. Da das Item jedoch in diesem Subtest positiv trennscharf ist, wurde dies nicht als Grundlage für Änderungsbedarf gewertet. Ähnliches gilt für die Bewertung der Wahlhäufigkeitswerte Hörender. Hier zeigen sich Häufungen der Wahl des Distraktors 'Zopf' in beiden Testdurchläufen, wobei sich diese Häufung im Schriftsprachtest mit neun Testpersonen sogar deutlich heraushebt. Die Testaufgabe scheint von daher selbst hörenden Testpersonen z. T. schwer zu fallen. Auch der Platz in der neuen Rangliste fällt den Hinweisen auf erhöhte Schwierigkeit des Items entsprechend deutlich tiefer aus als in der alten. Da jedoch die Trennschärfe- und Schwierigkeitswerte durchschnittlich eine gute Differenzierung der Leistungsunterschiede der Testpersonen untereinander bei mittelmäßiger Schwierigkeit anzeigen, wurde von Modifikationen des Items abgesehen und die Aufgabe unverändert in die Testversion I übernommen.

Nomen 5: Straße (Blase/Strauß/Weg)

Die Trennschärfewerte stellen sich durch alle fünf Subtests gut dar. Dabei liegt der Wert im Lautsprachtest mit gehörlosen Probanden etwas unterhalb des Durchschnitts (dieser fünf Werte), während er im Schriftsprachtest mit Hörenden gemeinsam mit vier weiteren Items den Spitzenwert der in diesem Durchlauf erzielten Trennschärfeergebnisse erreicht. Die Schwierigkeitswerte in den Testdurchläufen mit hörenden Testpersonen weisen das Item als sehr leicht aus, wenngleich es noch eine Anzahl leichterer Aufgaben gibt. Bei den gehörlosen Testpersonen stellt sich das Item in den Subtests Gebärdensprache und Lautsprache relativ leicht, jedoch im Vergleich zu den Hörenden etwas schwerer dar. Anders ergibt sich der Wert im Schriftsprachtest. Hier ist es im Mittelfeld angesiedelt und ist damit im Vergleich recht schwer, da nur zwei andere Items sich noch im Feld davor bewegen. Die Schwierigkeitswerte der Testdurchläufe mit Gehörlosen stehen damit, insbesondere im Schriftsprachtest, jedoch im Vergleich zu der relativ großen Anzahl sehr leichter Aufgaben auch im Gebärdensprachtest dem angenommenen Schwierigkeitsgrad etwas entgegen. Dies zeigt auch die Rangliste, auf der die Differenz von minus 5 Punkten genau die Mitte im Gesamtfeld der Differenzen darstellt.

Ein Distraktor jeden Typs ist vorhanden, wobei der Begriff ‚Weg‘ sogar Ähnlichkeiten in der Gebärde sowie inhaltlicher Art aufweist. So entfallen, wie die Distraktorenanalyse verdeutlicht, auch acht von insgesamt neun durch alle Subtests gemachten Fehler auf diesen Distraktor. Im Gegensatz dazu steht der Distraktor ‚Blase‘, der nie gewählt wurde, auch nicht von den hörenden Testpersonen. Die genauere Analyse dieses Distraktors ergibt, dass sich seine Wahl zunächst im Gebärdensprachtest überhaupt nicht anbietet, da die Hände bei der Ausführung der Gebärde ‚Weg‘ einen Streckenverlauf nachzeichnen, während eine Blase rund ist und demgemäß kaum Verwechslungsmöglichkeit besteht. Die Ähnlichkeit zum Item besteht zwar im Mund- sowie im Schriftbild des hinte-

ren Wortteils, jedoch differenziert der Anlaut des Distraktors offenbar zu deutlich vom Item, als dass er eine plausible Wahlmöglichkeit darstellen könnte. Unter gestalterischen Aspekten kommt hinzu, dass sich der Distraktor recht deutlich von den übrigen Abbildungen abhebt, was nicht in Verbindung zu den Werten zu sehen ist, jedoch grundsätzlich, gemäß der Richtlinie, optische Ruhe auf einer Seite anzustreben, nicht wünschenswert ist.

Auf diesem Hintergrund wurde beschlossen, das Item beizubehalten, da die Trennschärfewerte durchschnittlich gut und die Schwierigkeitswerte ausreichend sind. Der Distraktor ‚Blase‘ jedoch wurde eliminiert und durch den Begriff ‚Strand‘ ersetzt. Dieser passt sich zunächst der szenischen Darstellung der Begriffe ‚Straße‘ und ‚Weg‘ besser an und stört somit den optischen Eindruck der Seitenansicht nicht. Zudem ist nun der Anlaut dem des Items angepasst, sodass er seine Funktion als Ablenker vollständiger erfüllen kann.

Nomen 6: Bett (Brett/Decke/Sofa)

Das Item gehört durch alle fünf Testdurchläufe zu der jeweiligen Gruppe der leichtesten. Im Gebärdensprachtest differenziert es nicht zwischen den Testpersonen. Im Gegensatz dazu stehen die Trennschärfeergebnisse der anderen vier Testdurchläufe, in denen es durchgängig sehr gut abschneidet. Der Schwierigkeitsgrad differiert nur unauffällig von der angenommenen Position. Alle Ablenkertypen sind vorhanden, wobei der Distraktor ‚Decke‘ sogar doppelte Funktion hat. Das Item ist von den gehörlosen Testpersonen insgesamt dreimal fehlerhaft beantwortet worden. Dabei verteilen sich diese Fehlerpunkte auf jeweils einen der Distraktoren, sodass sich unter dem Aspekt möglicher Häufungen keiner von ihnen disqualifiziert. Die diesbezüglichen Werte der Erhebung mit Hörenden sind unauffällig.

Das Item kann damit übernommen werden. Die überwiegende Anzahl der Werte fällt ausgewogen und erwartungsgemäß aus. Die schlechten Trennschärfe- und Schwierigkeitswerte des Gebärdensprachtests bilden dabei die einzige Ausnahme. Die Bedeutung des Schwierigkeitswerts wurde jedoch in diesem Fall nicht hoch eingeschätzt, da das Item mit der Position sechs ja definitionsgemäß auf relativ leichter Position angesiedelt wurde.

Nomen 7: Angst (Freude/Trauer/Wut)

Im Gebärdensprachtest gehört das Item der Gruppe der nicht differenzierenden Testaufgaben an. Im Gegensatz dazu ist es im Schriftsprachtest mit Gehörlosen sehr gut trennscharf und vergleichsweise schwierig. Im Lautsprachtest ist es sogar am trennschärfsten von allen und dabei ebenfalls relativ schwierig. Die Trennschärfewerte in den Testdurchläufen mit Hörenden bewegen sich jeweils im Mittelfeld. Der Schwierigkeitsrang weicht nur geringfügig von der vorangehenden Annahme ab. Die Verteilung der Fehler auf die Distraktoren ist unauffällig. Auf dem Hintergrund dieser insgesamt sehr guten Ergebnisse wird das Item in dieser Form beibehalten.

Nomen 8: Stein (Kaffee/Stange/Stirn)

Während das Item im Subtest Gebärdensprache nicht differenziert, ist es in den beiden anderen Durchläufen mit Gehörlosen zwar sehr leicht, dabei jedoch mittelmäßig trennscharf. Diese mittelmäßige Trennschärfe zeigt sich gleichermaßen im Lautsprachtest mit Hörenden, während es sich im Schriftsprachtest in der Gruppe der Trennschärfsten befindet. Mit einer Differenz von plus sechs Rangplätzen weicht das Item von seinem angenommenen mittelmäßigen Schwierigkeitsgrad ab und stellt sich entsprechend leichter dar. Die Aufschlüsselung der Distraktorenwahl zeigt, dass in dem Testdurchlauf mit den Gehörlosen der Distraktor Stirn als Einziger nicht gewählt wurde. Die Ergebnisse der Wahlhäufigkeiten Hörender sind unauffällig. Der einzige einmalig gewählte Distraktor ist dieser Begriff ‚Stirn‘, was jedoch gemäß voranstehender Erläuterung mäßig aussagekräftig ist.

Das Item wurde prinzipiell auf Grund der guten Trennschärfewerte beibehalten. Um den Distraktor effektiver zu gestalten und gleichzeitig eine Steigerung des Schwierigkeitsgrades zu versuchen, wurde ersatzweise der Distraktor ‚Stirn‘ durch ‚Felsen‘ ersetzt.

Nomen 9: Tier (Bier/Tor/Verkehr)

Da das Item im Gebärdensprachtest sowie in den beiden Testdurchläufen mit Hörenden gar nicht differenzierte und darüber hinaus im Lautsprachtest mit Gehörlosen negativ trennscharf war, wurde das Item eliminiert.

Nomen 10: Bart (Bett/Kinn/Post)

In Bezug auf Trennschärfe und Schwierigkeit ergeben die Werte ein heterogenes Bild: So differenziert das Item im Gebärdensprachtest nicht, während es im Laut- und im Schriftsprachtest mit Gehörlosen gut trennscharf bei mittelmäßiger Schwierigkeit ist. Im Lautsprachtest mit Hörenden ist es mittelmäßig trennscharf; im Schriftsprachtest jedoch negativ trennscharf. Der Schwierigkeitsrang differiert unauffällig. Der Distraktor ‚Bett‘ wurde als Einziger weder von gehörlosen noch von hörenden Testpersonen gewählt, während sich die Wahl der anderen beiden Distraktoren gleichmäßig verteilt.

Sieht man sich in Hinblick auf den Trennschärfewert im Lautsprachtest mit Hörenden das Säulendiagramm zur Wahlhäufung an, so zeigt sich, dass 85 der insgesamt 86 Testpersonen das Item und nur einer den Distraktor ‚Kinn‘ gewählt hat. Durch diese eine, insgesamt leistungsstarke Testperson ergibt sich hier ein negativer Wert. Da dabei jedoch nicht von einer Häufung oder sonstigen Auffälligkeit gesprochen werden kann und die Relation von einer zu 86 Personen die Aussagekraft des Ergebnisses deutlich schwächt, beeinflusst es nicht die Entscheidung, das Testitem beizubehalten. Die zum überwiegenden Teil guten Ergebnisse legen dies nahe. Zur Verbesserung wurde jedoch der Distraktor ‚Bett‘ durch den Begriff ‚Haare‘ ersetzt. Dieser ähnelt dem Item in inhaltlicher Weise, womit der Testaufgabe das Vorkommen aller drei Ablenkertypen erhalten bleibt.

Nomen 11: Zahn (Hahn/Mund/Zaun)

Mit Ausnahme des Ergebnisses im Schriftsprachtest mit Gehörlosen differenziert das Item gar nicht oder fast gar nicht. Die Rangdifferenz von plus sechs legt zusätzlich die Annahme nahe, dass die Testaufgabe zu leicht konzipiert ist. Dabei fallen zwei gestalterische Aspekte ins Auge: Zum einen die Schraffierung links von der Darstellung des Distraktors ‚Mund‘; zum anderen die szenische Darstellung des Zauns, welche im Gegensatz zu den anderen drei gegenständlichen Zeichnungen steht.

Es wurde beschlossen, das Item zwar beizubehalten, jedoch zu modifizieren, um so eine Steigerung des Schwierigkeitsgrades zu erwirken. Dazu wurde die Schraffierung entfernt sowie der Mund selbst durch eine Grauschattierung noch etwas in den Vordergrund der Wahrnehmung gerückt, was ihn – in Angleichung an die anderen Zeichnungen – gegenständlicher erscheinen lässt. Der ‚Zaun‘ wurde ersetzt durch eine (gegenständliche) ‚Zange‘.

Nomen 12: Leute (Herde/Mensch/Paar)

Die Trennschärfewerte stellen sich durch alle fünf Testdurchläufe relativ bis sehr schlecht dar. Die Schwierigkeitswerte hingegen variieren deutlich. Während das Item im Gebärdensprachtest von allen Probanden richtig beantwortet wurde, ist es im Schriftsprachtest das zweitschwerste der unter dem Schwierigkeitsaspekt ohnehin breit gestreuten Items. Auch im Lautsprachtest mit Gehörlosen befindet es sich unter den schwierigeren. Im Lautsprachtest mit Hörenden befindet es sich im breiten Feld der sehr leichten; im Schriftsprachtest ist es das drittschwerste.

Ein vergleichbar heterogenes Bild zeigt sich anhand der Distraktorenanalyse. So wurden im Gebärdensprachtest null Fehler gezählt; im Lautsprachtest mit Gehörlosen wurden sieben und im Schriftsprachtest sechs Fehler gemacht. Verglichen mit sämtlichen Werten der Distraktorenanalyse/Nomen zählen beide Werte zu den wenigen hohen Fehlerbeträgen. Hinzu kommt eine ungleichgewichtige Verteilung der Wahlhäufigkeit der Distraktoren. So entfallen bspw. fünf der sieben Fehler im Schriftsprachtest auf den Distraktor ‚Mensch‘. Dieser Distraktor wurde auch im Schriftsprachtest mit Hörenden dreimal gewählt. Diese Anzahl ist innerhalb dieser Erhebung vergleichsweise herausragend hoch. Zudem hebt sich hier eine weitere Besonderheit von den bisherigen Erscheinungsbildern der Säulendiagramme zur Wahlhäufung ab: Im Subtest Schriftsprache haben drei, im Lautsprachtest eine der Testpersonen mit ‚weiß nicht‘ geantwortet.

Diese auffälligen und heterogenen Ergebnisse legten die Annahme nahe, dass hier Missverständnisse bzw. Unverständnisse seitens der Testpersonen bestanden. Dies wurde auf eine mangelhafte Gestaltung der Testaufgabe zurückgeführt, da die Werte eine Eliminierung einzelner Distraktoren oder der gesamten Aufgabe nicht nahe legten. Missverständnisse erklären sich möglicherweise u. a. damit, dass auf der Seite insgesamt sehr viele Personen und Tiere abgebildet sind, sodass ein relativ unruhiger und verwirrender Eindruck entsteht. Ein solcher Eindruck wird noch dadurch unterstützt, dass die beiden Abbildungen, auf denen viele Subjekte sichtbar sind, auf der unteren Hälfte der Seite platziert sind, während oben mit insgesamt drei dargestellten Personen größere optische Ruhe besteht. Hinzu kommt eine mangelhafte Ausführungsqualität insbesondere der Darstellung des gefragten Begriffs ‚Leute‘. Daher wurde die Platzierung der Distraktoren geändert und die Darstellungen unter der Prämisse größerer Eindeutigkeit und optischer Ruhe umgestaltet.

Nomen 13: Zahl (Buchstabe/Form/Zeichen)

In den beiden Durchläufen des Subtests Schriftsprache schneidet das Item gut bis sehr gut ab. Dabei ist es zwar jeweils relativ leicht, im Test mit Gehörlosen jedoch das Trennschärfste von allen. Im Gegensatz dazu stehen die Ergebnisse der anderen drei Testdurchläufe. Im Gebärdensprachtest differenziert es gar nicht, im Subtest Lautsprache mit Gehörlosen ist es sogar negativ trennscharf und im Lautsprachtest mit Hörenden ist es kaum trennscharf und sehr leicht. Die Analyse zeigt, dass sich die Fehlerpunkte auf die drei Distraktoren durch alle fünf Subtests verteilen. Dabei wird der Distraktor ‚Form‘ jedoch nur zweimal gewählt, gegenüber der vier- bzw. fünfmaligen Wahl der anderen beiden Distraktoren. Auffällig ist die Abweichung des Rangplatzes vom ursprünglichen Rang um plus sieben Punkte. Zudem ist dem gefragten Begriff dieser Testaufgabe kein Distraktor mit Ähnlichkeit in der Gebärde zugeordnet.

Um alle Distraktorentypen vertreten zu haben, wurde das ‚Zeichen‘ durch eine ‚Wolke‘ (mit Ähnlichkeit in der Gebärde) ersetzt. Dies war ohnehin geboten, da die Anforderung, aus einem bestimmten Zeichen (in diesem Fall ein Pfeil auf einem Schild), als das es sich ja nur darstellen lässt, den übergeordneten Begriff zu erkennen, eine missverständliche Aufgabenstellung und somit eine Überforderung darstellte. Da hierdurch ein optisches Ungleichgewicht auf der Seite entstanden wäre und zudem mit dem ‚Zeichen‘ der einzige mund- bzw. schriftbildähnliche Distraktor eliminiert worden war, wurde die ‚Form‘ gegen einen ‚Pfahl‘ ersetzt. Mit diesen Maßnahmen wurde intendiert, zum einen den Schwierigkeitsgrad der Aufgabe zu steigern, zum anderen missverständliche Elemente zu entfernen.

Nomen 14: Geld (Feld/Gold/Salz)

Das Item stellt sich in den Testdurchläufen Schriftsprache und Lautsprache mit Gehörlosen gut bis sehr gut trennscharf bei mittelmäßiger Schwierigkeit dar. Im Gegensatz dazu steht der Wert negativer Trennschärfe im Gebärdensprachtest, für den im vorliegenden Fall kaum gebärdensprachspezifische Zweideutigkeiten verantwortlich sein können. (Wäre dies der Fall gewesen, so hätten sich diese in erhöhter Wahlhäufigkeit des gebärdensprachähnlichen

Distraktors ‚Salz‘ zeigen müssen.) Im Lautsprachtest mit Hörenden schneidet das Item schwierigkeitsmäßig vergleichsweise gut ab, ist jedoch dafür sehr wenig trennscharf. Im Schriftsprachtest mit Hörenden befindet sich das Item in der Gruppe der trennschärfsten bei sehr geringer Schwierigkeit. Die Fehlerwahl in den Testdurchläufen mit Gehörlosen verteilt sich auf alle drei Distraktoren. Das Item differiert mit plus einem Punkt rangmäßig fast gar nicht von der angenommenen Schwierigkeit. Alle drei Ablenkertypen sind vorhanden.

Eine herausragende Besonderheit zeigt sich anhand des Säulendiagrammes zur Wahlhäufigkeit im Lautsprachtest mit Hörenden: Neunzehn der sechszwanzig Testpersonen wählten hier den Distraktor ‚Gold‘. Der Grund dafür wird in zweifacher Hinsicht in der Darstellungsweise des Items vermutet. Zum einen befinden sich auch in der Ansammlung von Kostbarkeiten, die die Goldbarren als Verständnishintergrund umgeben, Taler. Zum anderen geben die auf Scheinen liegenden Münzen bei der Darstellung des gefragten Begriffs ‚Geld‘ ein besonders unklares Erscheinungsbild auf Grund der (durchgängig in dieser Rohversion angewendeten und beabsichtigten) zeichnerischen Modalität ab.

Da die Daten auf der einen Seite größtenteils positiv oder unauffällig ausfallen und auf der anderen Seite vermutet wird, dass für die negativen Ergebnisse zeichnerische Aspekte verantwortlich sind, wurde das Item beibehalten und gemäß der oben geschilderten grundsätzlichen zeichnerischen Modifikationen überarbeitet. So sollte eine größere optische Ruhe und Eindeutigkeit die Verwechslungsrate zwischen ‚Geld‘ und ‚Gold‘ senken.

Nomen 15: Kasten (Kasse/Kissen/Schrank)

In den Testdurchläufen mit Gehörlosen fallen die Trennschärfe- und Schwierigkeitswerte auffällig heterogen aus. So stellt sich das Item im Gebärdensprachtest sehr leicht und dabei kaum trennscharf dar, differenziert also annähernd gar nicht. Im Schriftsprachtest ist es das schwerste von allen, dabei jedoch vergleichsweise mäßig trennscharf. Im Lautsprachtest fällt es aus dem Rahmen als hoch negativ trennscharf. Bei der Erhebung mit Hörenden liegen die Werte unauffällig im Mittelfeld.

Dabei hebt sich das Item in den Subtests Schrift- und Lautsprache mit Gehörlosen gleichzeitig durch sehr hohe Gesamtfehlerwerte hervor. Im Lautsprachtest wird mit neun Fehlern insgesamt der Spitzenwert unter den Nomen erreicht, im Schriftsprachtest wurde insgesamt sieben Mal fehlerhaft gewählt. Diese sieben Fehler entfallen alle auf den Distraktor ‚Schrank‘, der ausgerechnet der einzige der drei Ablenker ist, der keine Schriftbildähnlichkeit zum Item hat. Demgegenüber steht eine Wahlhäufung in dem Testdurchlauf Schriftsprache mit Hörenden: Hier haben 13 von insgesamt 81 Testpersonen den Distraktor ‚Kasse‘ mit Schriftbildähnlichkeit gewählt, keiner den Distraktor ‚Schrank‘. Diese Wahlhäufung erscheint entsprechend plausibler. Im Lautsprachtest mit Gehörlosen entfällt mit sechs Punkten die überwiegende Anzahl der Fehlantworten auf den Distraktor ‚Kissen‘. Da dieser Mundbildähnlichkeit zum gefragten Begriff hat, erscheint dies ebenfalls vergleichsweise plausibel. Wie der negative Trennschärfekoeffizient jedoch beweist, sind diese Fehler von den insgesamt guten Testpersonen begangen worden.

Zusammenfassend lässt sich sagen, dass erstens das schlechte Ergebnis des Gebärdensprachtests nicht überzeugt, zweitens die Distraktorenhäufung im Schriftsprachtest mit Gehörlosen unplausibel ist und drittens der negative Trennschärfewert im Lautsprachtest keine nachvollziehbare Erklärung findet und die Aufgabe damit disqualifiziert. Der unauffällige Ranglistenwert von minus drei sowie die plausiblen Ergebnisse aus den Testdurchläufen mit Hörenden fallen dagegen nicht ins Gewicht. Das Item wurde daher aus dem Test eliminiert.

Nomen 16: Holz (Brot/Gold/Höhle)

Die Koordinatensysteme zeigen ein uneinheitliches, jedoch insgesamt unauffälliges Bild. So fällt das Item im Gebärdensprachtest zwar leicht und mittelmäßig trennscharf aus, ist dabei jedoch im Vergleich das zweitschwerste und dritttrennschärfste. Im Schriftsprachtest ist es kaum trennscharf und dabei relativ leicht, während es im Subtest Lautsprache relativ schwierig, jedoch trotzdem nur wenig trennscharf ist. Im Testdurchlauf Lautsprache mit Hörenden ist es unter den vielen leichten Items eines der trennschärfsten, während es im Schriftsprachtest mit Hörenden gar nicht differenziert. Dem gesuchten Begriff ist kein inhaltsähnlicher Distraktor zugeordnet. Die Ranglistenverschiebung ist mit plus vier relativ unauffällig. Wahlhäufungen der Distraktoren zeichnen sich bei den Gehörlosen nicht ab, während im Schriftsprachtest mit Hörenden eine Auffälligkeit besteht: Dreizehn der 81 Testpersonen wählten hier den Distraktor ‚Höhle‘.

Der Distraktor ‚Höhle‘ fällt insofern aus dem Rahmen der Seite im Testheft, als er als einziger der vier Zeichnungen szenischen Charakter hat. Um diesen Umstand zu verändern und zum anderen die Testaufgabe um einen inhaltsähnlichen Distraktoren zu ergänzen, wurde er durch den Begriff ‚Baum‘ ersetzt. Diese Maßnahme versteht sich im Rahmen der allgemeinen modifikatorischen Beschlüsse. Die Datenlage gibt keine prägnanten Hinweise auf bestimmte modifikationsbedürftige Elemente. Da sie sich jedoch insgesamt gleichzeitig mittelmäßig darstellt, wurde das Item beibehalten.

Nomen 17: Arbeit (Abend/Eimer/Kaffee)

Die Datenlage legt die Annahme nahe, dass das Item zu leicht bzw. missverständlich ist. So differenziert es im Gebärdensprachtest sowie im Schriftsprachtest mit Hörenden gar nicht und im Lautsprachtest mit Hörenden fast gar nicht. Im Subtest Lautsprache mit Gehörlosen fällt es negativ trennscharf aus. Dem gegenüber steht allein das positive Ergebnis im Schriftsprachtest mit Gehörlosen. Hier ist das Item vergleichsweise schwierig und dabei recht gut trennscharf. Der Ranglistenplatz weicht um plus neun Punkte auffällig nach oben ab. Wahlhäufungen in den Testdurchläufen mit Gehörlosen zeigen sich im Schriftsprach- sowie im Lautsprachtest bei dem Distraktor ‚Abend‘. Dieser erscheint durch seine geringfügige Unterscheidung im Schrift- bzw. Mundbild vom Item plausibel. Optisch fällt er durch seine dunkle szenische Darstellung ins Auge. Zudem befinden sich die beiden Szenarien in der unteren Bildhälfte. Die hörenden Testpersonen jedoch haben sich davon nicht ablenken lassen.

Diese Ergebnisse legen eine Modifikation der Testaufgabe zur Steigerung des Schwierigkeitsgrades bzw. zur Eliminierung von missverständlichen Elementen nahe. So wurde der Distraktor ‚Kaffee‘, der in keinem der fünf Testdurchläufe gewählt wurde, gegen den ebenfalls gebärdenähnlichen und ebenfalls gegenständlich darstellbaren Distraktor ‚Stein‘ ausgewechselt. Der ‚Abend‘ wurde eliminiert und stattdessen der ebenfalls mund- bzw. schriftbildähnliche Distraktor ‚Ampel‘ eingesetzt. Zudem wurden die Positionen umverteilt.

Nomen 18: Schloß (Haus/Kirche/Schlüssel)

Die Schwierigkeits- und Trennschärfewerte ähneln in allen fünf Durchläufen denen des Items Nr. 17. Im Lautsprachtest mit Gehörlosen ist eine Wahlhäufung bei dem Distraktor ‚Schlüssel‘ zu erkennen, der als mund- bzw. schriftbildähnlicher Distraktor dafür auch plausibel ist. Bemerkenswert ist in diesem Zusammenhang, dass dieser alleine gegenständlich dargestellt wurde und daher ins Auge fällt. In den Testdurchläufen mit hörenden Testpersonen wurden vergleichsweise häufig Distraktoren statt des Items gewählt. Mit der geringfügigen Abweichung von plus eins bleibt der Schwierigkeitsgrad der Aufgabe weitgehend erhalten.

Der Distraktor ‚Haus‘ musste in Hinblick auf die Ähnlichkeitsverteilung zum Item mit seiner Inhaltsähnlichkeit nicht unbedingt erhalten bleiben, da diese durch die anderen beiden ebenfalls realisiert war. Zudem sollte eine der drei szenischen Zeichnungen einer gegenständlichen weichen. Daher wurde das ‚Haus‘ gegen ein ‚Schaf‘ ersetzt.

2.6.2 Diskussion der Testaufgaben/Verben

Item 1: sagen (fragen/schreien/schweigen)

Das Item ist in allen Testdurchläufen mit Ausnahme des Schriftsprachtests mit Hörenden negativ trennscharf. In diesem ist es mittelmäßig trennscharf. Die Schwierigkeitswerte stellen sich deutlich positiver dar. So ist das Item im Lautsprachtest mit Hörenden das schwierigste, im Gebärdensprachtest sowie im Schriftsprachtests mit Hörenden das zweitschwierigste Item. Im Schrift- sowie im Lautsprachtest mit Gehörlosen zählt es ebenfalls zu den schwierigeren Items. Setzt man diesen Umstand, dass die Lösung des Items den Testpersonen durchgehend relativ schwer fiel, in Beziehung zu den disqualifizierenden Trennschärfewerten, so liegt die Annahme nahe, dass die Aufgabe missverständlich gestaltet wurde. Diese Annahme wird gestützt durch den Schwierigkeitsrang eins, denn das Verb ‚sagen‘ kann als Basisvokabular eingestuft werden. Die drei Ablenkertypen sind jeweils vertreten. Im Gebärdensprachtest sowie im Lautsprachtest mit Gehörlosen häuft sich die Wahl des Distraktors schreien, was in beiden Fällen unplausibel ist und so die These einer verwirrenden Gestaltung der Testaufgabe stützt. Allein die Häufung der Wahl des Distraktors ‚fragen‘ in beiden Durchläufen des Schriftsprachtests ist stimmig, da dieser als einziger Mund- bzw. Schriftbildähnlichkeit aufweist. Der Umstand jedoch, dass im Lautsprachtest sieben der hörenden Testpersonen den Distraktor ‚fragen‘ und zweiundzwanzig den Distraktor ‚schreien‘ gewählt haben, weist ebenfalls daraufhin, dass die zeichnerische Darstellung missverständlich sein muss.

Aufgrund dieser Annahme der darstellerischen Missverständlichkeit wurde entschieden, das Item trotz teilweise disqualifizierender Werte mit überarbeiteten Zeichnungen in den neuen Itemkorpus aufzunehmen. Die Überarbeitung sollte folgende Punkte berücksichtigen: Alle vier Zeichnungen müssen mimisch deutlicher werden, so dass sie auch ohne Verständnis des Sprechblaseninhalts verständlich sind. Der Mund bei ‚sagen‘ muss deutlicher geöffnet sein (wie bei ‚fragen‘), um sich gegen ‚schweigen‘ abzuheben. Für den Schriftsprachtest muss eine eigene Zeichnung angefertigt werden, die das Gegensatzpaar ‚sagen‘ und ‚schweigen‘ darstellt.

Item 2: gehen (nagen/sehen/stehen)

Im Gebärdensprachtest sowie in beiden Testdurchläufen mit Hörenden differenziert das Item nicht. Im Schrift- sowie im Lautsprachtest mit Gehörlosen stellt es sich jedoch sehr trennscharf und relativ schwierig dar. Aufgrund dieser Ergebnisse verschiebt sich der Schwierigkeitsgrad auf der Rangliste um vier Plätze nach unten. Alle Ähnlichkeiten sind unter den Distraktoren vertreten. Der Distraktor ‚nagen‘ stellt sich sowohl in gestalterischer Hinsicht als auch in Hinblick auf die Fehleranalyse auffällig dar. So wählt bei insgesamt zwei vorkommenden Fehlern eine der gehörlosen Testpersonen diesen Distraktor im Schriftsprachtest, obgleich nur eine Ähnlichkeit in der Gebärde vorliegt. Im Lautsprachtest entfallen zwei von fünf vorkommenden Fehlern auf diesen Distraktor, was in Hinblick auf die Ähnlichkeiten zum gefragten Begriff ebenso wenig plausibel ist.

Plausibilität bekommen diese Ergebnisse jedoch bei der Ansicht der vier Zeichnungen in der Zusammenschau: Das nagende Tier hebt sich gestalterisch in jeglicher Hinsicht von den drei anderen einander ähnelnden Zeichnungen ab und drängt sich durch diese Hervorhebung zur Wahl auf.

Im Ergebnis legen diese Interpretationen die Beibehaltung des Items aufgrund der teilweise herausragend guten Ergebnisse nahe, wobei der Distraktor ‚nagen‘ durch das Verb ‚rennen‘ ersetzt wurde. Die Darstellung wurde damit den übrigen drei Zeichnungen angeglichen. Die Verteilung der Ähnlichkeiten wurde verschoben, da ‚rennen‘ dem gefragten Begriff nicht wie ‚nagen‘ in Hinblick auf die Gebärde, sondern im Mundbild sowie inhaltlich bzw. zeichnerisch angeglichen ist.

Item 3: fahren (fliegen/frieren/schenken)

Im Gebärdensprachtest sowie in den beiden Testdurchläufen mit Hörenden differenziert das Item nicht. Im Schriftsprachtest mit Gehörlosen ist es kaum trennscharf und relativ leicht. Im Lautsprachtest mit Gehörlosen ist es das leichteste der positiv trennscharfen Items und dabei das zweitleichteste. Auf der Rangliste verschiebt sich das Item nicht. Aufgrund der geringen Gesamtfehlerzahl ergeben sich keine auffälligen Häufungen.

Diese Ergebnisse machen deutlich, dass das Item zu leicht ist, um die Leistungen der Testpersonen zu differenzieren. Für die Testversion I wurden daher zwei der Distraktoren ersetzt. Statt ‚fliegen‘ wurde das Verb ‚feiern‘ aufgenommen, das dem gesuchten Begriff in Mund- und Schriftbild sowie in der Gebärde ähnelt. Das Verb ‚schenken‘ wurde durch ‚reiten‘ ersetzt, was dem gesuchten Begriff inhaltlich bzw. zeichnerisch sowie in Hinblick auf die Gebärde ähnelt. Auf diese Weise sind wieder alle drei Ähnlichkeiten vorhanden. Zeichnerisch wurde ein optisches Gleichgewicht auf der Seite im Testheft hergestellt, indem die beiden linken Begriffe ‚feiern‘ und ‚frieren‘ mit jeweils drei nebeneinander stehenden Personen dargestellt wurden und die beiden rechten Begriffe ‚reiten‘ und ‚fahren‘ mit jeweils einer sich auf dem entsprechenden Gefährt fortbewegenden Person. So erhält keine der Zeichnungen einen hervorstechenden Charakter, was die Wahl in sinnvoller Weise erschweren kann.

Item 4: essen (schmieren/trinken/Zähne putzen)

Im Gebärdensprachtest differenziert das Item nicht. In den übrigen vier Durchläufen jedoch sind die Ergebnisse sehr gut. Die Fehleranalyse bei den Gehörlosen ist unauffällig. In den Durchläufen mit Hörenden wird relativ häufig der Distraktor ‚Zähne putzen‘ gewählt. Dies kann in der Darstellung begründet sein, da die Zahnbürste nur undeutlich als solche zu erkennen ist und so mit etwas Essbarem verwechselt werden kann. Dieses mögliche Missverständnis sollte mit der grundsätzlich ausgereifteren und eindeutigeren Darstellungsweise der Testversion I ausgeräumt werden.

Insofern stellte sich das Item grundsätzlich gut geeignet dar, es musste jedoch einer der Distraktoren ersetzt werden, da kein mund- bzw. schriftbildähnlicher Ablenker vorhanden war. So wurde statt ‚schmieren‘ nun ‚messen‘ verwendet.

Item 5: gucken (backen/ducken/küssen)

Die Werte von Trennschärfe und Schwierigkeit zeigen ein sehr heterogenes Bild. So ist das Item bspw. im Schriftsprachtest mit Gehörlosen das zweitschwierigste, während es im Lautsprachtest (mit Gehörlosen) negativ trennscharf ist. Auf der Rangliste verschiebt sich das Item um 12 Plätze deutlich in den schwierigeren Bereich. Die Wahlhäufungen der Ergebnisse mit Hörenden ergeben keine plausiblen Hinweise auf Änderungsbedarf. Dieser ist jedoch aufgrund der einseitig vertretenen Ähnlichkeiten zum Item gegeben, denn alle drei Distraktoren ähneln ausschließlich im Mund- bzw. Schriftbild. Daher wurde der Distraktor ‚backen‘ durch ‚schminken‘ ersetzt (inhaltliche Ähnlichkeit) und der Distraktor ‚küssen‘ durch ‚denken‘ (Gebärde). Zudem wurde beschlossen, den sich Duckenden mit gesenktem Blick darzustellen, um eine Verwechslung mit gucken auszuschließen.

Item 6: fliegen (fragen/liegen/winken)

Die Trennschärfe/Schwierigkeits-Ergebnisse disqualifizieren das Item. So ist es im Lautsprachtest mit Gehörlosen negativ trennscharf und nur im Schriftsprachtest mit Gehörlosen relativ gut trennscharf, dabei aber sehr leicht. In den übrigen Durchläufen differenziert das Item nicht. Die Fehleranalysen ergeben keine verwertbaren Hinweise zur Überarbeitung. Zudem fehlt der Distraktor mit inhaltlicher Ähnlichkeit. Das Item wurde daher eliminiert.

Item 7: schlafen (liegen/schlagen/träumen)

Die Trennschärfe- und Schwierigkeitswerte zeigen ein heterogenes Bild. Insbesondere in den Durchläufen mit Hörenden sind die Ergebnisse jedoch relativ gut, sodass von einer Eliminierung trotz negativer Trennschärfe im Lautsprachtest mit Gehörlosen abgesehen werden sollte. Hinzu kommt, dass dieser Wert durch nur eine fehlerhafte Wahl zustande kommt und entsprechend mäßig aussagekräftig ist.

Auffällig ist eine Wahlhäufung im Schriftsprachtest mit Hörenden. Vier der Testpersonen haben den Distraktor ‚träumen‘ gewählt. Im Lautsprachtest mit Gehörlosen entfällt der einzige Fehler ebenfalls auf diesen Distraktor. Dies ist plausibel, da die Darstellung einer mit geschlossenen Augen sitzenden Person mit leerer Denkblase auch für ‚schlafen‘ gehalten werden kann. Daher wurde dieser Distraktor durch das Verb ‚streicheln‘ ersetzt. Dies erfüllte zusätzlich die Funktion, alle Ablenkertypen bereitzustellen, da ein gebärdensprachlicher Ablenker zuvor fehlte. Außerdem konnte die Heftseite auf diese Weise ausgewogen gestaltet werden. So wurden nun zwei liegende und zwei im Portrait dargestellte Personen gezeichnet.

Item 8: denken (danken/lenken/schenken)

Während die Trennschärfe/Schwierigkeitswerte in den Durchläufen mit Gehörlosen mit negativer Trennschärfe im Schrift- und Lautsprachtest und ohne Differenzierung im Gebärdensprachtest sehr schlecht ausfallen, stellen sich die Ergebnisse mit Hörenden umso positiver dar. Im Schriftsprachtest ist das Item das trennschärfste von allen und auch im Lautsprachtest gehört es in das obere Viertel. Diese Ergebnisse erklären sich vermutlich durch den Umstand, dass alle drei Distraktoren Mund- bzw. Schriftbildähnlichkeit zum gefragten Begriff aufweisen und sich dabei in jeweils nur einem Laut von diesem unterscheiden. So liegt die Annahme nahe, dass das Item für die gehörlosen Testpersonen zu schwer war. Die Verschiebung des Schwierigkeitsranges um sechs Plätze nach unten weist ebenfalls in diese Richtung. Die Ergebnisse der Testdurchläufe mit Hörenden sprechen außerdem für eine Beibehaltung des Items. Daher wurde der Distraktor ‚danken‘ durch ‚träumen‘ ersetzt (Gebärdensprache) und der Distraktor ‚lenken‘ durch ‚warnen‘ (Gebärde und Zeichnung).

Item 9: hinsetzen (anspitzen/sitzen/schwitzen)

Die Ergebnisse disqualifizieren das Item. Es ist in drei Durchläufen negativ trennscharf und differenziert im Schriftsprachtest mit Gehörlosen gar nicht. Ausschließlich im Schriftsprachtest mit Hörenden liegt es im Bereich der positiv Trennscharfen, ist von diesen jedoch das schlechteste, wenn auch mit Abstand schwerste. Hier haben 29 der 81 Testpersonen den Distraktor ‚sitzen‘ gewählt. Auch im Lautsprachtest mit Hörenden wählten sechs von 86 Testpersonen diesen Distraktor. Im Gebärdensprachtest entfallen ebenfalls deutliche Wahlhäufungen auf diesen Distraktor. Es liegt daher nahe anzunehmen, dass den Testpersonen die Unterscheidung zwischen ‚hinsetzen‘ und ‚sitzen‘ zu schwer fiel und verwirrend war. Da die Ergebnisse diese Interpretation nahelegen wurde beschlossen, das Item mit veränderter Darstellung und unter Austausch eines der Distraktoren erneut in den Itemkorpus aufzunehmen. So wurde der Pfeil am Rücken des sich Hinsetzenden als ungeeignete und verwirrende Symbolik weggelassen und beide Personen, der Sitzende sowie der sich Hinsetzende, von ihrer Körperhaltung deutlicher gestaltet. Außerdem wurde der Distraktor ‚anspitzen‘ durch ‚hinlegen‘ ersetzt, um so vier Personen auf der Testheftseite abzubilden und den Gesamteindruck homogener zu gestalten.

Item 10: warten (braten/halten/wandern)

Im Gebärdensprachtest sowie im Lautsprachtest mit Hörenden differenziert das Item nicht. In den übrigen beiden Testdurchläufen mit Gehörlosen jedoch sind die Werte herausragend gut. Es bestehen weder auffällige Wahlhäufungen noch Rangabweichungen. Dem Item fehlt ein Distraktor mit inhaltlicher Ähnlichkeit. Daher wurde ‚halten‘ durch ‚sitzen‘ ersetzt und der Seite eine homogenere Darstellung verliehen.

Item 11: verstecken (erschrecken/schlafen/stecken)

Die Ergebnisse fallen auffällig heterogen aus. So gehört das Item im Gebärden- sowie in den beiden Lautsprachtests zu den besten, in den beiden Schriftsprachtests jedoch ist es negativ trennscharf. Hier wählen vier der Gehörlosen und fünf der 81 Hörenden den Distraktor ‚schlafen‘. Da dieser ausschließlich Ähnlichkeit in der Gebärde hat, lässt sich kein plausibler Grund für diese Wahlhäufungen ausmachen. Daher wurde das Item auf dem Hintergrund der teilweise sehr guten Ergebnisse prinzipiell unverändert übernommen. Allein die Darstellung des Distraktors ‚stecken‘ wurde statt mit einer in einen Leuchter zu steckenden Kerze durch einen in eine Steckdose zu steckenden Stecker gezeichnet. Diese Darstellung war zweifelsfrei als ‚stecken‘ zu identifizieren, während vorher das (unerwünschte) Symbol eines Pfeils verwendet werden musste.

Item 12: weinen (regnen/schwitzen/warnen)

Die Ergebnisse in den beiden Lautsprachtests fallen sehr gut aus, sagen jedoch aufgrund der Wahlhäufungen weniger über die Qualität des Items aus, sondern zeigen Änderungsbedarf an. So wird in beiden Durchläufen gehäuft der Distraktor ‚schwitzen‘ gewählt, der jedoch keine Ähnlichkeit im Mundbild zum gesuchten Begriff hat. Dies weist eindeutig darauf hin, dass die Zeichnung missverständlich ist und dem Weinenden zu sehr ähnelt. Die Darstellungen wurden entsprechend umgestaltet. Um den fehlenden gebärdenähnlichen Distraktor zu ergänzen, wurde ‚wiegen‘ durch ‚warnen‘ ersetzt. Hiermit wurde gleichzeitig eine homogene Gestaltung der Seite erreicht, da nun alle Personen in Brüstungsgröße dargestellt werden.

Sollten die Missverständlichkeiten durch die Verdeutlichungen der Darstellungen ausgeräumt worden sein, so kann angenommen werden, dass das Item dadurch leichter wird. Gleichzeitig wird jedoch davon ausgegangen, dass diesem unerwünschten Effekt durch die darstellerische Angleichung von ‚warnen‘ entgegengetreten wird.

Item 13: schreiben (bohren/reiben/schlafen)

Die Ergebnisse stellen sich heterogen dar. Die Fehleranalyse legt keine Änderungen nahe. Auf der Rangliste verschiebt sich das Item um fünf Ränge nach oben. Es fehlt ein inhaltsähnlicher Ablenker. Da der Distraktor ‚bohren‘ in allen fünf Durchläufen kein Mal gewählt wurde, wurde dieser durch ‚malen‘ ersetzt. Auf diese Weise wurde dem gesuchten Begriff ‚schreiben‘ ein naheliegenderer Ablenker beigelegt, was zu einer Steigerung des Schwierigkeitsgrades führen kann.

Item 14: lesen (blasen/fegen/halten)

Mit Ausnahme des Lautsprachtests mit Hörenden, in dem das Item negativ trennscharf ist, schneidet es überall herausragend gut ab. Dieses eine negative Ergebnis wird deutlich relativiert durch den Umstand, dass hier nur eine Person fehlerhaft geantwortet hat. Es bestehen keine Wahlhäufungen. Auffällig ist die Verschiebung in den leichten Bereich um neun Ränge. Dem Item fehlt ein inhaltsähnlicher Ablenker. Daher wurde der Distraktor ‚fegen‘ durch ‚sitzen‘ ersetzt. Dies wurde durch eine im Sessel sitzende Person dargestellt, die ein geschlossenes Buch auf dem Schoß liegen hat. So wurde der inhaltliche Bezug zu dem gesuchten Begriff hergestellt und zusätzlich eine Erschwerung eingefügt.

Item 15: singen (geigen/sagen/schweigen)

Die Trennschärfe-/Schwierigkeitswerte fallen heterogen, jedoch relativ unauffällig aus. Es bestehen keine auffälligen Wahlhäufungen. Der gebärdenähnliche Ablenker fehlt und das Item verschiebt sich auf der Rangliste um fünf Plätze nach oben. Um einen gebärdenähnlichen Distraktor zu ergänzen, wurde ‚sagen‘ durch ‚warnen‘ ersetzt. Zudem wurde der Geiger gegen einen Flötisten ausgetauscht, da eine Flöte (auch gehörlosen) Kindern geläufiger sein dürfte als eine Geige.

Item 16: waschen (wanken/warten/wischen)

In allen fünf Durchläufen wurde, teilweise als einziger, teilweise vornehmlich, der Distraktor ‚wischen‘ gewählt. Dies ist für alle drei Wortarten plausibel, da in jeglicher Hinsicht Ähnlichkeiten bestehen. Die Trennschärfe-/Schwierigkeitswerte fallen jedoch heterogen aus, so dass sich keine einheitliche Interpretation ableiten lässt. Auffällig ist allein die Verschiebung des Schwierigkeitsgrades um neun Plätze in den leichten Bereich. Da der Distraktor ‚wanken‘ kaum gewählt wurde, wurde er durch ‚schwimmen‘ ersetzt, was dem gesuchten Begriff inhaltlich ähnelt und so eine Erschwernis bedeuten kann.

Item 17: brennen (brechen/rechnen/rennen)

Im Lautsprachtest mit Hörenden haben drei von 86 Testpersonen ‚rennen‘ gewählt. Hier liegt jedoch die Annahme nahe, dass der gefragte Begriff akustisch falsch verstanden wurde, da nicht davon auszugehen ist, dass hörenden Grundschulern das Verb ‚brennen‘ unbekannt ist. Dieses Ergebnis ist entsprechend wenig aussagekräftig. Im Schriftsprachtest mit Gehörlosen ist das Item das schwierigste, im Lautsprachtest das zweitschwierigste, wobei alle drei Distraktoren gewählt wurden. Es fehlte ein Distraktor mit inhaltlicher Ähnlichkeit. Daher wurde ‚brechen‘ durch ‚löschen‘ ersetzt. Zudem wurde die Person vor dem brennenden Haus eliminiert, da sie unnötig vom Objekt ablenken könnte.

Item 18: schieben (heben/schließen/ziehen)

Mit Ausnahme des Schriftsprachtests mit Gehörlosen sind die Ergebnisse durchgängig gut. Da die Zeichnung jedoch eindeutig ist, können keine Veränderungen vorgenommen werden. Das Item wird daher in gleicher Weise übernommen.

2.6.3 Diskussion der Testaufgaben/Adjektive

Item 1: ganz (gerade/halb/rund)

Im Gebärdensprachtest wählten neun der Testpersonen ‚rund‘. Dies ist plausibel, da ‚rund‘ der Gebärdensprachablenker ist. Das Item ist hier positiv trennscharf und das schwierigste aller Items. Dennoch wurde eine zeichnerische Modifikation vorgenommen, da ‚rund‘ als ‚ganze‘ Orange dargestellt wurde, die ja auch ‚ganz‘ und somit im Sinne der Testaufgabe missverständlich ist. Im Austausch wurde eine halbe Orange dargestellt, die in der Draufsicht auf die aufgeschnittene Fläche ja immer noch ‚rund‘ ist, und als durchgestrichener Gegensatz dazu eine gerade ganze Tafel Schokolade.

Im Schriftsprachtest mit Gehörlosen sowie in dem Durchlauf mit Hörenden wählten jeweils sechs Testpersonen den Distraktor ‚gerade‘. Dies ist missverständlicherweise plausibel, da der Begriff als ‚ganze‘ Tafel Schokolade dargestellt wurde, welche ja auch gerade ist. Daher ist die Wahlhäufung auf dieses Missverständnis zurückzuführen und das negativ trennscharfe Ergebnis in dem Durchlauf mit Gehörlosen entsprechend wenig aussagekräftig. Die Darstellungen von ‚ganz‘ und ‚halb‘ wurden daraufhin jeweils als Gegensatzpaar in Form eines Brotlaibes ausgeführt.

Im Lautsprachtest mit Gehörlosen wählten sechs Testpersonen ‚gerade‘, was als Mundbildablenker plausibel ist. Das Ergebnis ist jedoch negativ trennscharf. Im Lautsprachtest mit Hörenden entfielen die drei fehlerhaften Antworten auf den Distraktor ‚rund‘, was aufgrund der oben beschriebenen Missverständlichkeit wenig aussagekräftig ist.

Item 2: klein (groß/klug/schlecht)

Das negativ trennscharfe Ergebnis des Gebärdensprachtests ist aufgrund lediglich einer Fehlerantwort mäßig aussagekräftig. Die Ergebnisse der übrigen vier Durchläufe sind durchgängig gut bis sehr gut, jedoch zeigen die Fehleranalysen Modifikationsbedarf an. Im Schriftsprachtest entfallen sämtliche Fehlerantworten Hörender (vier von 81) sowie zwei von drei Fehlerantworten Gehörloser auf den Distraktor ‚groß‘. Dieser ähnelt dem gesuchten Begriff nicht im Schriftbild. Daher liegt die Annahme nahe, dass die Darstellung zu undeutlich ist. Auch im Lautsprachtest mit Hörenden schien das Vorkommen der Begriffe ‚groß‘ und ‚klein‘ innerhalb einer Testaufgabe Verwirrung zu stiften, wie die vierfache Wahl des Distraktors ‚groß‘ annehmen lässt. So wurde bei ‚klein‘ die Maus optisch hervorgehoben und der Elefant deutlicher durchgestrichen. ‚Groß‘ wurde ersetzt durch ‚dünn‘ (Gegensatzpaar mit ‚dick‘).

Das Ergebnis des Lautsprachtests mit Gehörlosen ist sehr gut. Hier wählten sechs von acht Testpersonen den mundbildähnlichen Ablenker ‚klug‘.

Der Distraktor ‚schlecht‘ fiel optisch aus dem Gesamtbild der Seite heraus. Zudem wurde angenommen, dass die Anwendung des Begriffs ‚schlecht‘ auf die fauligen Früchte eine Überforderung darstellen könnte. Der Begriff ließ sich jedoch nicht anders darstellen, daher wurde er gegen ‚fein‘ (Gegensatzpaar mit ‚schlampig‘) ausgetauscht.

Item 3: blau (braun/grau/schwarz)

Die Werte zeigen, dass das Item sehr leicht zu lösen ist. So differenziert es in den Durchläufen mit Hörenden sowie im Gebärdensprachtest gar nicht. In den übrigen beiden Durchläufen erzielt es jedoch gute Ergebnisse. Auf der Rangliste verschiebt es sich nach oben, sodass es nun den ersten Rang der leichtesten Testaufgabe einnimmt.

Da die Farbwörter aus dargestellten Gründen im Test vertreten sein sollten, das Konzept der Farb-Items keine Modifikationen zuließen und gleichzeitig die Ergebnisse z. T. positiv ausfielen, wurde das Item beibehalten.

Item 4: voll (falsch/stolz/viel)

Im Gebärdensprachtest wählten auffällig viele Testpersonen den Distraktor ‚viel‘ (9x). Das Schwierigkeits/Trennschärfe-Ergebnis ist zwar gut und ‚viel‘ ist der Gebärdensablenker, dennoch liegt aufgrund der Häufung die Vermutung nahe, dass die Zeichnungen nicht richtig verstanden wurden. Die Missverständlichkeit der Aufgabe kommt insbesondere im Schriftsprachtest zum tragen, denn wenn die Testperson lediglich die Zeichnung des ‚vollen‘ Glases vor sich sieht, ist die Wahl des Begriffs ‚viel‘ gleichermaßen plausibel. So wählten neun von 81 hörenden Testpersonen ‚viel‘. In der Folgeversion wurden daher beide Wörter als Gegensatzpaar dargestellt. Auf diese Weise kann differenziert werden, dass der Gegensatz eines leeren (durchgestrichenen) Glases präzise als ‚voll‘ denn als ‚viel‘ bezeichnet wird.

Im Lautsprachtest mit Gehörlosen differenziert das Item gar nicht, da alle Testpersonen fehlerhaft auswählten. Dies verdeutlicht, dass das Item missverständlich ist, was durch die Verschiebung auf der Rangliste um zwölf Plätze nach unten bestätigt wird. Die Wahlhäufung im Lautsprachtest mit Hörenden (zehn von 86 wählten ‚viel‘) stützen diese These. Um die Aufgabe verständlicher und den Schwierigkeitsgrad somit angemessener zu gestalten, wurde der Begriff ‚stolz‘ ebenfalls als Gegensatzpaar dargestellt. Zudem wurde der Distraktor ‚falsch‘ eliminiert, da die Darstellungsvariante mit einer fehlerhaften Rechenaufgabe symbolüberladen war. Statt dessen wurde ‚groß‘ als Gegensatzpaar zweier Gläser gezeichnet.

Item 5: blond (braun/grau/rot)

Die Trennschärfe/Schwierigkeitswerte sind sowohl bei den Gehörlosen als auch bei den Hörenden durchgehend auffällig gut. In allen Testdurchläufen kommen Fehler vor, im Gebärdensprachtest sowie im Lautsprachtest mit Gehörlosen sogar relativ zahlreich. Das Item ist demnach schwieriger als angenommen. Es differiert auf der Rangliste um fünf Plätze.

Unabhängig von den guten Ergebnissen wurde aus inhaltlichen Gründen eine Änderung vorgenommen: Der Distraktor ‚grau‘ wurde durch ‚schwarz‘ ersetzt, da ein junges Mädchen kaum grauhaarig sein kann.

Item 6: komisch (gemischt/kurz/neu)

Das Item ist sowohl im Lautsprachtest mit Gehörlosen als auch im Schriftsprachtest mit Hörenden negativ trennscharf. In den übrigen drei Durchläufen sind die Ergebnisse mittelmäßig schwierig, dabei jedoch nur mäßig trennscharf.

Verschiedene gestalterische Schwierigkeiten erschweren eine Modifikation der Aufgabe. So ist die Seite uneinheitlich und zu detailreich gestaltet, was einen verwirrenden Eindruck macht. Zudem lässt die Darstellung des Distraktors ‚kurz‘ auch eine Interpretation von ‚komisch‘ zu. Außerdem lässt sich der Distraktor ‚gemischt‘ kaum darstellen, ohne missverständlich zu sein.

Aufgrund dieser Schwierigkeiten sowie der teilweise negativ trennscharfen Ergebnisse wurde von modifikatorischen Versuchen abgesehen und das Item eliminiert.

Item 7: krumm (kaputt/krank/stumm)

Das Item ist in drei Testdurchläufen negativ trennscharf. Im Gebärdensprachtest differenziert es nicht. Im Lautsprachtest mit Hörenden ist es mäßig positiv trennscharf und dabei relativ leicht. Zudem ist die Seite heterogen gestaltet und es fehlt ein inhaltsähnlicher Distraktor. Diese Ergebnisse disqualifizieren die Aufgabe und sie wurde daher eliminiert.

Item 8: hellblau (dunkelblau/hellgrau/hellgrün)

Die Trennschärfe/Schwierigkeitswerte sind gut bis sehr gut, lediglich im Lautsprachtest mit Hörenden differenziert die Aufgabe nicht, da hier alle Testpersonen korrekt geantwortet haben. Alle übrigen Werte sind unauffällig bzw. erwartungsgemäß. Das Item zeichnet sich zudem besonders dadurch aus, dass jeder der Distraktoren gleichzeitig gebärden-, mund- bzw. schriftbild- sowie inhaltsähnlich ist. Es wurde somit ohne Modifikationen übernommen.

Item 9: orange (lila/rosa/rot)

In den Durchläufen mit Gehörlosen sind die Trennschärfe/Schwierigkeitsergebnisse sehr gut. Das negativ trennscharfe Ergebnis im Schriftsprachtest mit Hörenden, in dem zwei Testpersonen den Distraktor ‚rot‘ gewählt haben, ist vermutlich auf eine Ungenauigkeit der Darstellung zurückzuführen. So ist das ‚rot‘ so hell, dass es als orange angesehen werden kann. Das Ergebnis verliert entsprechend an Bedeutung.

Die Rangliste zeigt eine Verschiebung der Aufgabe um fünf Plätze in den leichten Bereich. Um diese Verschiebung möglichst zu verringern, wurde der Distraktor ‚lila‘, der sich am deutlichsten von den anderen abhebt, durch ‚gelb‘ ersetzt, welches den übrigen Farben ähnlicher ist.

Item 10: traurig (ernst/erstaunt/fröhlich)

Die Aufgabe ergibt teilweise sehr gute Ergebnisse. Im Gebärdensprachtest sowie im Lautsprachtest mit Hörenden antworteten alle Testpersonen korrekt. Das Item ist somit leichter als angenommen und verschiebt sich um sieben Plätze auf der Rangliste nach oben. Die guten Ergebnisse sprechen jedoch dennoch für eine prinzipiell unveränderte Übernahme.

Die Fehleranalyse im Schriftsprachtest mit Hörenden zeigt an, dass die drei Testpersonen, die fehlerhaft ausgewählt haben, sämtlich den Distraktor ‚ernst‘ tippten. Da im Schriftsprachtest nur die Zeichnung des gesuchten Begriffs zu sehen ist und somit nicht gegen die Zeichnung von ‚ernst‘ abgewogen werden kann, war diese Fehlerwahl plausibel. Um dies auszuschließen, wurden im Gesicht der ‚Traurigen‘ Tränen ergänzt.

Item 11: verletzt (besetzt/gesund/sauber)

In drei Testdurchläufen ist das Item negativ trennscharf. Im Lautsprachtest mit Hörenden differenziert es nicht und lediglich im Schriftsprachtest mit Gehörlosen ist das Ergebnis gut. Da in den beiden Lautsprachtestdurchläufen relativ häufig der Distraktor ‚gesund‘ gewählt wurde, lag die Annahme einer missverständlichen Darstellung nahe. So wurde statt eines Mädchens mit dem Arm in einer Schlinge ein mit Pflastern versehenes Knie dargestellt. Der mund- bzw. schriftbildähnliche Distraktor ‚besetzt‘ wurde fast nie gewählt. Er wurde durch den Begriff ‚verliebt‘ ersetzt, um so möglicherweise den Schwierigkeitsgrad zu erhöhen. Der gebärdensprachliche Ablenker ‚sauber‘ wurde im Schrift- sowie im Lautsprachtest mit Gehörlosen unverständlicherweise häufig gewählt. Er wurde daher durch ‚scharf‘ ersetzt.

Item 12: warm (heiß/kalt/weich)

Das Item ist in drei Testdurchläufen die schwierigste aller Aufgaben, einmal das zweitschwierigste und lediglich im Gebärdensprachtest differenziert es nicht. Die Fehleranalysen verdeutlichen jedoch, dass, unabhängig von diesen insgesamt sehr guten Ergebnissen, erheblicher Modifikationsbedarf besteht. So haben in allen Testdurchläufen annähernd sämtliche Testpersonen den Distraktor ‚heiß‘ gewählt, im Schriftsprachtest mit Hörenden entfielen außerdem zehn von sechsunddreißig Fehlantworten auf ‚kalt‘. Diese Ergebnisse machen überdeutlich, dass die Darstellung der verschiedenen Wärmegrade durch Thermometer mit den entsprechenden Wetterlagen sowie den Gradzahlen die angesprochene Altersgruppe vollkommen überfordern. Daher wurden die Zeichnungen entsprechend durch szenische Darstellungen ersetzt. Der Distraktor ‚weich‘ wurde überarbeitet, indem das Kissen und das Brett, die das Gegensatzpaar darstellen, deutlicher nebeneinander statt ineinander übergehend gestaltet wurden, um so weitere mögliche Missverständnisse auszuschließen.

Für den Schriftsprachtest wurde eine Extrazeichnung als Gegensatzpaar ‚warm/heiß‘ gefertigt, um die Differenzierung zwischen diesen beiden überhaupt zu ermöglichen.

Item 13: glatt (platt/satt/sauber)

Sämtliche gehörlosen Testpersonen haben in allen drei Durchläufen fehlerhaft ausgewählt, sodass das Item nicht differenziert. Im Lautsprachtest mit Hörenden wurde ebenfalls weit überwiegend fehlerhaft geantwortet und die Aufgabe schneidet hier negativ trennscharf ab. Im Gegensatz dazu steht allein der Lautsprachtest mit Hörenden, wo die Aufgabe bei ebenfalls zahlreichen Fehlantworten die trennschärfste und zweitschwierigste ist.

Diese sehr schlechten Ergebnisse legen eine umfassende Überarbeitung der Aufgabe nahe. Es ist insbesondere bei den hörenden Testpersonen nicht anzunehmen, dass aufgrund fehlender Kenntnis des Begriffes so zahlreich falsch geantwortet wurde, sondern es ist davon auszugehen, dass dies darstellungsimmanente Gründe hat. So wurde die Darstellung von ‚sauber‘ nun durch ein durchgestrichenes dreckiges Auto neben einem sauberen dar-

gestellt. Der gesuchte Begriff ‚glatt‘ wurde als ein auf einem zugefrorenen See ausrutschendes Mädchen statt durch glattes Papier gezeichnet.

Der Distraktor ‚satt‘ wurde vergleichsweise selten oder gar nicht gewählt. Insbesondere in den Schriftsprachtest-durchläufen hätte er im Verhältnis zu der Gesamtfehleranzahl als Schriftbildablenker häufiger gewählt werden müssen. Er wurde daher gegen den Begriff ‚lockig‘ ausgetauscht. Dieser ähnelt dem gefragten Begriff nicht in Schrift- oder Mundbild, was auch nicht notwendig ist, da ‚platt‘ diese Funktion hat und so weiterhin alle Ablenkerarten vorhanden sind.

Item 14: links (oben/rechts/unten)

Mit Ausnahme des Gebärdensprachtests, in dem das Item nicht differenziert, da alle Testpersonen fehlerhaft geantwortet haben, sind die Ergebnisse durchgehend sehr gut. Dabei wurde, auch in den Durchläufen mit Hörenden, relativ häufig fehlerhaft ausgewählt, so dass die Aufgabe zu den schwierigsten gehört. Sie wurde ohne Modifikationen übernommen.

Item 15: frei (erschrocken/frech/klein)

Die Ergebnisse sind durchgehend gut bis sehr gut. Dabei zählt das Item mit Ausnahme des Schriftsprachtests mit Gehörlosen zu den jeweils leichteren. Gleichzeitig ist festzustellen, dass der Distraktor ‚klein‘ von den gehörlosen Testpersonen kaum gewählt wurde (von insgesamt zwölf Fehlerantworten ein Mal). Außerdem fehlt der Aufgabe ein Distraktor mit inhaltlicher Ähnlichkeit zum Item. Daher wurde ‚klein‘ durch „gefangen“ ersetzt.

Item 16: schmutzig (sauber/zerknüllt/zerrissen)

Die Werte sind ausnahmslos gut. Auffällig stellt sich allein die Wahlhäufung des Distraktors ‚zerrissen‘ im Lautsprachtest mit Hörenden dar. Hier kann davon ausgegangen werden, dass die Zeichnung missverständlich ist, da die zerrissene Hose gleichzeitig schmutzig aussieht. Sie wurde entsprechend modifiziert dargestellt. Außerdem wurde der Distraktor ‚zerknüllt‘ gegen ‚zerknittert‘ ausgetauscht, da dieser präziser auf Textilien zutrifft.

Item 17 weich (wach/warm/weiß)

Trennschärfe und Schwierigkeit sind durchgehend gut bis sehr gut. In allen Testdurchläufen mit Ausnahme des Gebärdensprachtests wurden auffällig viele Fehler begangen. Dabei wurden alle drei Distraktoren gewählt, ohne dass sich hierfür plausible Erklärungen ablesen ließen.

Da dem Item zum einen ein Gebärdenablenker fehlte und gleichzeitig der Distraktor ‚weiß‘ als einziger als allein stehendes Objekt dargestellt wurde, wurde dieser durch ‚nass‘ ersetzt. Aufgrund der guten Ergebnisse wurde die Aufgabe trotz der hohen und unerklärlichen Fehlerfrequenz mit dieser einen Modifikation übernommen.

Item 18: leise (heiß/leicht/links)

Die Ergebnisse stellen sich heterogen dar. Im Gebärdensprachtest differenziert das Item nicht, da kein Fehler begangen wurde. Im Schriftsprachtest mit gehörlosen Testpersonen ist das Ergebnis nur mäßig trennscharf und dabei mittelmäßig schwierig, die Fehleranalyse jedoch unauffällig. Ebenso unauffällig ist diese im Lautsprachtest mit Gehörlosen, wobei die Ergebnisse hier, insbesondere im Hinblick auf die Trennschärfe, sehr gut ausfallen. Im Lautsprachtest mit Hörenden ist das Item mäßig positiv trennscharf und mittelmäßig schwierig. Hier ist die Häufung der Wahl des Distraktors ‚leicht‘ auffällig. Vierzehn von 86 Testpersonen haben diesen gewählt. Dieses Ergebnis ist eindeutig auf die Zeichnung zurückzuführen. So wurde ‚leicht‘ durch eine schwebende bzw. fallende Feder dargestellt, die ja ebenso leise zu Boden gleitet.

Im Schriftsprachtest mit Hörenden schließlich fällt das Item negativ trennscharf aus. Hier haben fünf von 81 Testpersonen den Distraktor ‚heiß‘ gewählt. Dies ist aufgrund der Zeichnung zwar in letzter Konsequenz nicht plausibel, jedoch erklärbar. So ist die Fingerhaltung des Mädchens, die dem neben ihr stehenden Jungen bedeutet, leise zu sein, zu schräg. Statt dessen hätte der Finger gerade gezeichnet werden müssen.

In der Überarbeitung wurde die Darstellung des gesuchten Begriffs insgesamt verändert. So wurden eine einem Lautsprecher lauschende neben einer sich vor dem Lautsprecher die Ohren zuhaltenden Person gezeichnet. Zudem wurden die Positionen der beiden Personen vertauscht, weil die den Begriff ‚leise‘ darstellende links stand und daher auch die Wahl des Distraktor ‚links‘ plausibel gewesen wäre.

Alle drei Distraktoren ähnelten dem gesuchten Begriff im Mund- bzw. Schriftbild. Daher mussten, um alle drei Varianten vertreten zu haben, zwei der Distraktoren ersetzt werden. Da ‚leicht‘ und ‚links‘, wie beschrieben, zu Missverständnissen führten, wurden diese durch ‚süß‘ und ‚laut‘ ersetzt.

2.7 Ergänzung zusätzlicher und schwererer Items

Die Daten legen insgesamt eine Steigerung des Schwierigkeitsgrades nahe, d. h. dass im Bereich der schwierigen Testaufgaben deutlichere Differenzierungsmöglichkeiten unter den Testpersonen erwirkt werden sollten. Zu diesem Zweck wurden der benannten Wortliste sechs weitere Items pro Wortart aus deren schwierigem Bereich entnommen und dem Test hinzugefügt. Gleichzeitig wurden die Items, die in Folge der Pilotstudie eliminiert werden mussten, durch neue ersetzt. So wurde der Itemkorpus auf die Anzahl von achtundsechzig aufgestockt. Mit dieser Maßnahme wurde gleichzeitig intendiert, diejenigen Items, die sich nach der Auswertung der Hauptuntersuchung als unbrauchbar erweisen würden, eliminieren zu können, ohne den Itemkorpus dabei unerwünscht zu verkleinern.

Zusätzlich wurde dem Test ein zweites Übungsbeispiel vorangestellt. Es hatte sich herausgestellt, dass ein solches bei Nichtverständnis anhand des ersten und einzigen Übungsbeispiels ‚Auge‘ gebraucht wird.

Bei der Einfügung der neuen Testaufgaben in den vorhandenen Itemkorpus wurden wie bereits zuvor die Rangplätze gemäß der Wortliste zu Grunde gelegt. Items, Distraktoren sowie Rangplatz sind der folgenden Liste zu entnehmen. Dabei kennzeichnen die abgekürzten Buchstaben in Klammern jeweils unter den Distraktoren deren Ähnlichkeiten zum Item.

Tab. 15a-B 2.7: Auswahl und Distraktorenähnlichkeit zusätzlicher und schwerer Items Nomen				
Nr.	Item	Distraktor I	Distraktor II	Distraktor III
0	Hund	Hahn (M/S, I/Z)	Katze (I/Z)	Pferd (G, I/Z)
7	Krone	Bohne (M/S)	Hut (G, I)	Kranz (M/S)
9	Garten	Baum (I/Z)	Geschenk (M/S)	Arbeit (G)
11	Schuh	Stiefel (G, I/Z)	Strumpf (G, M/S, I/Z)	Uhr (M/S)
15	Ei	Eis (M/S, I/Z)	Huhn (I/Z)	Kartoffel (G, I/Z)
18	Fisch	Boot (G, I/Z)	Flasche (M/S)	Frosch (M/S, I/Z)
22	Kanne	Kaffee (M/S, I/Z)	Kanister (G, M/S, I/Z)	Wanne (M/S)
23	Rose	Rasen (M/S, I/Z)	Sonnenblume (G, I/Z)	Tulpe (G, I/Z)
24	Welt	Ball (G, I/Z)	Planet (G, I/Z)	Vase (G, M/S, I/Z)

Tab. 15b-B 2.7: Auswahl und Distraktorenähnlichkeit zusätzlicher und schwerer Items Verben				
Nr.	Item	Distraktor I	Distraktor II	Distraktor III
5	fliegen	springen (I/Z)	winken (G)	fragen (M/S)
7	springen	sitzen (G/M/S)	klettern (G, I)	schimpfen (M/S)
11	anziehen	ausziehen (G, I/Z, M/S)	tragen (G)	ziehen (M/S)
15	werfen	fangen (G, M/S, I/Z)	strecken (I/Z)	wärmen (M/S)
22	stolpern	springen (G, I/Z)	stoßen (M/S)	streicheln (M/S)
23	stinken	riechen (G, I/Z)	staunen (M/S)	winken (M/S)
24	frieren	lachen (G)	zwicken (M/S)	schütteln (I/Z)

Tab. 15c-B 2.7: Auswahl und Distraktorenähnlichkeit zusätzlicher und schwerer Items Adjektive				
Nr.	Item	Distraktor I	Distraktor II	Distraktor III
6	spitz	struppig (G)	schief (I/Z,M/S)	spannend (M/S)
7	nass	trocken (I/Z)	blass (M/S)	nackt (M/S, I/Z)
10	dünn	eng (G, I/Z)	dick (G, I/Z)	dumm (M/S)
11	schwer	stark (G,M/S, I/Z)	leer (M/S)	schwarz (M/S)
15	schief	steil (G, I/Z, M/S)	gerade (I/Z)	stark (M/S)
22	gestreift	kariert (G, I/Z)	geblümt (M/S, I/Z)	gepunktet (I/Z)
23	viereckig	dreieckig (I/Z,G)	fünfeckig (I/Z,G,M/S)	sechseckig (I/Z,G)
24	nah	geschlossen (G)	fern (I/Z)	nass (M/S)

2.8 Testversion I: Itemkorpus

Im Folgenden wird der modifizierte und ergänzte Itemkorpus für Testversion I dargestellt. Dabei sind neue Items oder Distraktoren durch Schraffierung hervorgehoben:

Tab. 16a-B 2.8: Itemkorpus Nomen Version I					
neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	1	Mutter	Mädchen	Bauer	Motte
2	2	Auto	Bus	Laster	Zug
3	3	Haus	Hand	Kirche	Maus
4	4	Kopf	Zopf	Junge	Topf
5	5	Straße	Weg	Strand	Strauß
6	6	Bett	Decke	Brett	Sofa
7	-	Krone	Bohne	Hut	Kranz
8	7	Angst	Wut	Trauer	Freude
9	-	Garten	Baum	Geschenk	Arbeit
10	8	Stein	Felsen	Kaffee	Stange
11	-	Schuh	Stiefel	Strumpf	Uhr
12	10	Bart	Haare	Kinn	Post
13	11	Zahn	Mund	Hahn	Zange

neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
14	12	Leute	Paar	Mensch	Herde
15	-	Ei	Huhn	Kartoffel	Eis
16	13	Zahl	Regen	Buchstabe	Pfahl
17	14	Geld	Gold	Salz	Feld
18	-	Fisch	Boot	Flasche	Frosch
19	16	Holz	Baum	Brot	Gold
20	17	Arbeit	Eimer	Stein	Ampel
21	18	Schloß	Schaf	Kirche	Schlüssel
22	-	Kanne	Kaffee	Kanister	Wanne
23	-	Rose	Sonnenblume	Tulpe	Rasen
24	-	Welt	Planet	Ball	Vase

Tab. 16b-B 2.8: Itemkorpus Verben Version I

neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	1	sagen	fragen	singen	schweigen
2	2	gehen	laufen	sehen	stehen
3	3	fahren	feiern	frieren	reiten
4	4	essen	messen	trinken	zähneputzen
5	-	fliegen	springen	winken	fragen
6	5	gucken	schminken	denken	ducken
7	-	springen	sitzen	klettern	schimpfen
8	7	schlafen	schlagen	streicheln	liegen
9	8	denken	träumen	warnen	schenken
10	9	hinsetzen	sitzen	hinlegen	schwitzen
11	-	anziehen	ausziehen	tragen	ziehen
12	10	warten	sitzen	wandern	braten
13	11	verstecken	schlafen	erschrecken	stecken
14	12	weinen	reiben	schwitzen	wiegen
15	-	werfen	fangen	strecken	wärmen
16	13	schreiben	malen	reiben	schlafen
17	14	lesen	sitzen	halten	blasen
18	15	singen	sagen	warnen	flöten
19	16	waschen	wischen	schwimmen	warten
20	17	brennen	löschen	rechnen	rennen

neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
21	18	schieben	schließen	heben	ziehen
22	-	stolpern	springen	stoßen	streicheln
23	-	stinken	riechen	staunen	winken
24	-	frieren	lachen	zwicken	schütteln

Tab. 16c-B 2.8: Itemkorpus Adjektive Version I

neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	1	ganz	gerade	halb	rund
2	2	klein	klug	dünn	fein
3	3	blau	braun	grau	schwarz
4	4	voll	viel	groß	stolz
5	5	blond	braun	rot	schwarz
6	-	spitz	struppig	schief	spannend
7	-	nass	trocken	blass	nackt
8	8	hellblau	dunkelblau	hellgrau	hellgrün
9	9	orange	rosa	gelb	rot
10	-	dünn	eng	dick	dumm
11	-	schwer	stark	leer	schwarz
12	10	traurig	ernst	erstaunt	fröhlich
13	11	verletzt	scharf	gesund	verliebt
14	12	warm	heiß	kalt	weich
15	-	schief	steil	gerade	stark
16	13	glatt	sauber	lockig	platt
17	14	links	oben	rechts	unten
18	15	frei	erschrocken	gefangen	frech
19	16	schmutzig	sauber	zerknittert	zerrissen
20	17	weich	nass	wach	warm
21	18	leise	heiß	laut	süß
22	-	gestreift	kariert	geblümt	gepunktet
23	-	viereckig	dreieckig	fünfeckig	sechseckig
24	-	nah	geschlossen	fern	nass

3. Hauptuntersuchung

Mit der Version I des Wortschatztests wurde in der Zeit vom 28. August 1997 bis zum 19. Januar 1998 eine Stichprobe von 112 gehörlosen und hochgradig schwerhörigen Kindern untersucht. Da die Eichstichprobe repräsentativ für die Bevölkerungsgruppe sein soll, in der der Test später Anwendung finden wird, wurden Dritt-, Viert- und beginnende Fünftklässler aus Schulen in Deutschland sowie in Zürich ausgewählt. Damit wurde etwa ein Viertel aller Testpersonen der entsprechenden Grundgesamtheit erfasst. Dieser Umstand rechtfertigt die Wahl der anfallenden und nicht zufälligen Zusammenstellung der Stichprobe, da bei einer so umfangreichen Erhebung nahe an der Grundgesamtheit von Repräsentativität der Stichprobe ausgegangen werden muss. Die Annahme der Normalverteilung der Stichprobe wird gemäß der üblichen Vorgehensweise nach der Erhebung (s. B 3.3) überprüft.

3.1 Stichprobe

Nicht alle 112 Testpersonen haben an jedem der Subtests bzw. Subtestteile teilgenommen. Aus Krankheitsgründen waren zwei Testpersonen (Junge, 3. Klasse; Mädchen, 4. Klasse) zur Durchführung des Gebärdensprachteils, zwei Testpersonen (beides Mädchen, 4. und 5. Klasse) zur Durchführung des Schriftsprachteils und sechs Testpersonen (drei Mädchen, drei Jungen, alle 3. Klasse) zur Durchführung des Lautsprachteils jeweils in allen Wortarten nicht anwesend. Zudem haben zwei Testpersonen (beides Jungen, beide 3. Klasse) in der Modalität Schriftsprache aus zeitlichen Gründen lediglich die Wortart Nomen bearbeitet. (Im Folgenden werden demnach bei der Zuordnung der Testpersonen zu bestimmten Kategorien für die Modalität Schriftsprache stets zwei Angaben gemacht.) Es ergibt sich demzufolge für den Gebärdensprachteil eine Stichprobengröße von 110 Testpersonen, für den Schriftsprachteil von 110 (Nomen) bzw. 108 Testpersonen (Verben und Adjektive) und für den Lautsprachteil von 106 Testpersonen. Für die Wertung des Gesamttests wurden nur diejenigen Testpersonen berücksichtigt, die an allen Subtestteilen teilgenommen haben. Ihre Anzahl beläuft sich auf 101.

Bei der Auswahl der Schulen wurde versucht, im Sinne der Repräsentativität flächendeckend vorzugehen. Es fanden sich sieben Schulen in deutschen Städten sowie eine Schule in der deutschsprachigen Schweiz bereit, an der Untersuchung teilzunehmen. Nicht in jeder Schule standen Schüler aller drei gewünschten Klassenstufen zur Verfügung. Die daraus resultierende Aufteilung der Testpersonen in Klassenstufen ergibt sich wie folgt: Getestet wurden 40,2% der Testpersonen aus 3. Klassen, 27,7% aus 4. Klassen und 32,1% aus 5. Klassen. Für die Anwendung einiger statistischer Verfahren werden im weiteren Verlauf zur Erweiterung

des Datenpools der gehörlosen Grundschüler die beginnenden 5. zu den 4. Klassen gezählt, sodass 40,2% Drittklässler einem Anteil von 59,8% Viertklässlern gegenüberstehen. Diese Zuordnung legitimiert sich aus dem Umstand, dass die betreffenden Testpersonen zu Beginn der fünften Klassenstufe weder altersmäßig noch in Bezug auf Leistungs- bzw. Entwicklungsniveau gravierend von den Viertklässlern abweichen. Die Anzahl der Testpersonen verteilt sich folgendermaßen auf die 8 Schulen und die 3 Klassenstufen:

Tab. 17-B 3.1: Anzahl der Testpersonen				
Stadt	3. Klasse	4. Klasse	5. Klasse	gesamt
Berlin	11 (6/5)	6	-	17
Bremen	6	-	8	14
Hamburg	7	-	6	13
Köln	6	7	-	13
Münster	6	5	7	18
Schleswig	5	-	7	12
Würzburg	4	5	4	13
Zürich	-	8 (4/4)	4	12
gesamt	45	31	36	112

Die Stichprobe bestand zu 48,2% aus Mädchen und zu 51,8% aus Jungen:

Tab. 18-B 3.1: Geschlecht der Testpersonen			
Klassenstufe	Mädchen	Jungen	gesamt
3. Klasse	20	25	45
4. Klasse	15	16	31
5. Klasse	19	17	36
gesamt	54	58	112

5,4% der Testpersonen wurden nach dem bilingualen Konzept unterrichtet, 32,1% nach hörgerichteten bzw. oralen Konzepten und 62,5% mit Einsatz von Gebärden:

Tab. 19-B 3.1: Unterrichtskonzepte der Sprachvermittlung				
Klassenstufe	bilingual	mit Gebärden	oral	gesamt
3. Klasse	-	31	14	45
4. Klasse	-	20	11	31
5. Klasse	6	19	11	36
gesamt	6	70	36	112

Die Testpersonen waren zum Testzeitpunkt der Subtests Gebärden- und Schriftsprache zwischen 7;11 und 13;3, im Durchschnitt 10;5 Jahre, zum Testzeitpunkt des Subtests Lautsprache zwischen 8;0 und 13;4, im Durchschnitt 10;6 Jahre alt. Sie befanden sich beim Testtermin jeweils einer Sprachmodalität in folgenden Altersgruppen:

Tab. 20a-B 3.1: Alter der Testpersonen beim Subtest Gebärdensprache				
Alter (Jahre)	3. Klasse	4. Klasse	5. Klasse	gesamt
7;6 - 7;11	1	0	0	1
8;0 - 8;5	2	0	0	2
8;6 - 8;11	4	0	0	4
9;0 - 9;5	11	2	0	13
9;6 - 9;11	8	5	4	17
10;0 - 10;5	11	8	4	23
10;6 - 10;11	4	6	6	16
11;0 - 11;5	2	4	7	13
11;6 - 11;11	1	4	8	13
12;0 - 12;5	0	0	5	5
12;6 - 12;11	0	1	1	2
13;0 - 13;5	0	0	1	1
gesamt	44	30	36	110

Tab. 20b-B 3.1: Alter der Testpersonen beim Subtest Schriftsprache				
Alter (Jahre)	3. Klasse	4. Klasse	5. Klasse	gesamt
7;6 - 7;11	1	0	0	1
8;0 - 8;5	2	0	0	2
8;6 - 8;11	4	0	0	4
9;0 - 9;5	12	2	0	14
9;6 - 9;11	8	4	4	16
10;0 - 10;5	10 (9)	10	3	23 (22)
10;6 - 10;11	5 (4)	5	7	17 (17)
11;0 - 11;5	2	4	7	13
11;6 - 11;11	1	3	8	12
12;0 - 12;5	0	1	4	5
12;6 - 12;11	0	1	1	2
13;0 - 13;5	0	0	1	1
gesamt	45 (43)	30	35	110 (108)

Tab. 20c-B 3.1: Alter der Testpersonen beim Subtest Lautsprache				
Alter (Jahre)	3. Klasse	4. Klasse	5. Klasse	gesamt
7;6 - 7;11	0	0	0	0
8;0 - 8;5	2	0	0	2
8;6 - 8;11	4	0	0	4
9;0 - 9;5	9	2	0	11
9;6 - 9;11	9	4	4	17
10;0 - 10;5	10	8	3	21
10;6 - 10;11	5	7	6	18
11;0 - 11;5	0	4	6	10
11;6 - 11;11	0	4	7	11
12;0 - 12;5	0	1	7	8
12;6 - 12;11	0	1	2	3
13;0 - 13;5	0	0	1	1
gesamt	39	31	36	106

3.2 Untersuchungsergebnisse

Die 112 Testpersonen werden aus Datenschutzgründen im Folgenden nach einer nur den Autorinnen bekannten Reihenfolge durchnummeriert.

3.2.1 Darstellung der Testpunktwerte

Eine im Anhang (Anh. 1: B 3.2.1) aufgeführte Tabelle (Tab. 21-B 3.2.1) stellt die Ergebnisse aller mit Version I des Wortschatztests untersuchten Testpersonen (je Subtestteil maximal 24 Punkte) dar.

Die Punktzahl variiert im Subtestteil GSN von 11 bis 24, durchschnittlich 18 Punkte, im Subtestteil GSV von 9 bis 24, durchschnittlich 19 Punkte, im Subtestteil GSA von 10 bis 23, durchschnittlich 16 Punkte. Im Subtestteil SSN liegen die Ergebnisse zwischen 7 und 24, durchschnittlich bei 16 Punkten, im Subtestteil SSV zwischen 2 und 24, durchschnittlich bei 12 Punkten, im Subtestteil SSA zwischen 3 und 23, durchschnittlich bei 13 Punkten. Im Subtestteil LSN erreichten die Testpersonen Ergebnisse zwischen 5 und 24, durchschnittlich 16 Punkte, im Subtestteil LSV zwischen 4 und 23, durchschnittlich 14 Punkte, im Subtestteil LSA zwischen 4 und 24, durchschnittlich 13 Punkte.

3.2.2 Darstellung der Distraktorenwahl

Während in B 3.2.1 die Ergebnisse der einzelnen Testpersonen aufgeführt wurden, dokumentieren drei weitere Tabellen (Tab. 22a/b/c-B 3.2.2) im Anhang (Anh. 1: B 3.2.2) Besonderheiten in der Beantwortung einzelner Aufgaben und in der Wahl der Distraktoren. Dies gibt Aufschluss über die Qualität der Aufgaben im Hinblick auf die Eignung des Testinstruments. Diese Informationen werden in die Aufgaben- (s. B 4.1) bzw. Distraktorenanalyse (s. B 4.1.5) und in die Weiterarbeit an dem Instrument einfließen. Erläuterungen dazu erfolgen an den entsprechenden Stellen.

3.3 Überprüfung auf Normalverteilung

Die Analytestichprobe stellt eine repräsentative Auswahl aus der Grundgesamtheit deutscher (und deutschsprachiger schweizer) gehörloser Schüler aus 3., 4. und 5. Klassen dar. Das Zustandekommen einer Normalverteilung ist von daher grundsätzlich möglich, wenn zusätzlich die Darstellung des Häufigkeitspolygons Normalität ausweist.

Die Darstellung der Daten erfolgt in der üblichen Abstufung des zu messenden Merkmals auf der Abzisse und der Anzahl der vorkommenden Merkmalsausprägungen auf der Ordinate eines Koordinatensystems. Zur tabellarischen Aufbereitung des Materials wurden die individuellen Messwerte in Kategorien zusammengefasst. Mit Hilfe der Berechnungsregel nach Sturges (in Bortz 1993, 29) ergibt sich für die Subtestteile des Wortschatztests eine Anzahl von 7,8 Kategorien. Die Kategorienbreite wurde mit Hilfe einer Faustregel nach Bortz (vgl. 1993, 29) berechnet. Daraus ergeben sich folgende Kategorienbreiten für die Subtestteile²⁰: GSN: 1,7; GSV: 1,9; GSA: 1,7; SSN: 2,2; SSV: 2,8; SSA: 2,6; LSN: 2,4; LSV: 2,4; LSA: 2,6. Zum Erhalt einer übersichtlichen Darstellung wird eine einheitliche Einteilung für alle Subtestteile vorgenommen, sodass in Anlehnung an die Vorschläge von Sturges und Bortz die Testpunktwerte jedes Subtestteils in acht Kategorien zu je drei Testpunktwerten unterteilt wurden. Eine Tabelle (Tab. 23-B 3.3) im Anhang (Anh. 1: B 3.3a) verdeutlicht die Punkteverteilung nach diesem Vorgehen.

Die Darstellung der Testergebnisse in Form von kategorisierten Verteilungskurven (Abb. 11a-i-B 3.3) im Anhang (Anh. 1: B 3.3b) legt die Annahme einer Normalverteilung trotz geringfügiger Verschiebungen nach rechts nahe. Diese Verschiebung ist durch den Umstand gerechtfertigt, dass das Testinstrument konzeptionell darauf ausgerichtet wurde, den Grundwortschatz gehörloser Grundschüler zu überprüfen. Es wird demnach davon ausgegangen,

²⁰ Die Abkürzungen für die Subtestteile verstehen sich wie folgt: GSN: Gebärdensprache Nomen, GSV: Gebärdensprache Verben, GSA: Gebärdensprache Adjektive, SS: Schriftsprache (Wortarten entsprechend), LS: Lautsprache (Wortarten entsprechend)

dass alle Testpersonen zumindest einige Items kennen und richtig zuordnen können. Demgemäß kommen Gesamtergebnisse einzelner Subtestteile mit weniger als vier Punkten nur vereinzelt vor.

Für die Analyse der Verteilung der Testergebnisse wurden zusätzlich die Schiefe- und Exzesswerte sowie deren Signifikanzen betrachtet, die die Abweichung der Verteilung von der Normalverteilung beschreiben:

Tab. 24-B 3.3: Schiefe und Exzess									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
Schiefe	-.24	-1,23	-.14	-.07	-.04	-.21	-.51	-.31	-.06
p	.230	.230	.230	.230	.233	.233	.240	.240	.240
Exzess	-.36	2,28	-.99	-1,18	-1,15	-1,19	-.48	-.67	-1,03
p	.457	.457	.457	.457	.461	.461	.465	.465	.465

Die aufgeführten Signifikanzwerte verdeutlichen, dass diese oder extremere Schiefe- und Exzesswerte in der Realität wahrscheinlich häufig auftreten. Bei einer Analysenstichprobe dieses Umfangs sollte die Nullhypothese, nach der die Testpunktwerte des Wortschatztests normalverteilt sind, erst auf dem 1%-Niveau aufgegeben werden (vgl. Lienert/Raatz 1994, 148). Hier trifft sie mit hoher Wahrscheinlichkeit zu und soll demnach als angenommen gelten.

3.4 Aufgabenanalyse

3.4.1 Qualitätskriterien der Testaufgaben

Die – zu einem späteren Zeitpunkt nach Abschluss der Aufgabenanalyse zu überprüfenden – Anforderungen an Objektivität, Reliabilität und Validität kann der Wortschatztest nur dann erfüllen, wenn auch jede einzelne seiner Testaufgaben objektiv, reliabel und valide ist. Neben der Ausprägung dieser Testgütekriterien wird die Qualität jeder Aufgabe auch durch ihre Trennschärfe und ihre Schwierigkeit gekennzeichnet.

Eine Aufgabe ist objektiv, wenn unterschiedliche Testauswerter bei der Bewertung der Aufgabe zu dem gleichen Ergebnis kommen, sie also gleichermaßen als richtig bzw. falsch beantwortet interpretieren.

Eine Aufgabe ist reliabel, wenn sie beispielsweise bei einer Testwiederholung nach einem angemessenen Zeitraum zu gleicher Beantwortung führt (Retest-Reliabilität) oder wenn sie

in derselben Weise beantwortet wird wie eine streng vergleichbare Aufgabe einer anderen Testskala (Parallelttest-Reliabilität).

Die Beziehung zwischen Objektivität und Reliabilität gestaltet sich generell als einseitig monoton (vgl. Lienert/Raatz 1994, 31), bei hoher Objektivität kann demnach im allgemeinen auch von hoher Reliabilität einer Aufgabe ausgegangen werden. Unter Berücksichtigung der Tatsache, dass die Ergebnisinterpretation des Wortschatztests anhand der Auswertungsmaßstäbe im Testprotokoll eindeutig ist und somit hohe Objektivität der Aufgaben gewährleistet wird, kann aufgrund des beschriebenen Zusammenhangs zwischen Objektivität und Reliabilität gleichzeitig von hoher Aufgabenreliabilität ausgegangen werden. Dies wird zudem durch die Ergebnisse der Berechnung von Reliabilitätsindizes über Aufgabenstreuung und -trennschärfe (s. B 3.4.4) belegt. Auf die gesonderte Überprüfung der Aufgabenreliabilität in Form von Re- oder Parallelttestung konnte deshalb verzichtet werden (vgl. Lienert/Raatz 1994, 31).

Nach Objektivität und Reliabilität muss als drittes Testgütekriterium auch die Validität für einzelne Aufgaben überprüft werden. Eine Aufgabe ist valide, wenn sie von Testpersonen mit starker Merkmalsausprägung häufiger richtig beantwortet wird als von Testpersonen mit geringer Merkmalsausprägung. Die Validität des gesamten Wortschatztests (s. B 3.7.3) wird dabei bereits vorausgesetzt, um Auskunft über die Höhe der Korrelation der einzelnen Aufgabe mit dem Validitätskriterium des Tests zu erhalten.

Während der Validitätskoeffizient durch die punktbiseriale Korrelation zwischen der Aufgabenbeantwortung der Testpersonen (richtig, falsch) einerseits und dem Grad der Merkmalsausprägung (gemessen an einem Außenkriterium) andererseits beschrieben wird, ist der Trennschärfekoeffizient durch die punktbiseriale Korrelation zwischen der Aufgabenbeantwortung der Testpersonen (richtig, falsch) und den von ihnen erzielten Rohwerten im Gesamttest definiert. Der Validitätskoeffizient hat demnach für den Gesamttest dieselbe Bedeutung wie der Trennschärfekoeffizient für eine einzelne Aufgabe. Ein hoher Trennschärfekoeffizient besagt dabei, dass die dazugehörige Aufgabe dem Gesamtpunktwert entsprechend gute und schlechte Testpersonen deutlich voneinander trennt, indem sie von den guten meist richtig, von den schlechten Testpersonen meist falsch beantwortet wird.

Zur Bestimmung der Trennschärfekoeffizienten des Wortschatztests wurde auf die Berechnung mit Hilfe eines echten Außenkriteriums verzichtet, da kein hinreichend reliables Kriterium gewonnen werden konnte, welches sich unmittelbar an dem zu untersuchenden Persönlichkeitsmerkmal orientiert. Statt dessen wurde, wie generell üblich (vgl. Lienert/Raatz 1994, 59), im Sinne eines quantitativen Analyse Kriteriums der Rohwert der Testvorform als Binnenkriterium herangezogen, um mit ihm den Score einer einzelnen Aufgabe zur Ableitung

des Trennschärfekoeffizienten zu korrelieren. Da dieses Vorgehen im Gegensatz zu einer Aufgabenanalyse an einem echten Außenkriterium nur mittelbar zu einer Validitätsverbesserung führt, ist eine gesondert durchgeführte Validierung (s. B 3.7.3) unerlässlich. Im Sinne des Gebotes einer Kreuzvalidierung (vgl. Lienert/Raatz 1994, 60) bietet sich diese Aufteilung, bei der zur Aufgabenanalyse ein Binnen- und zur Validierung ein Außenkriterium gewählt wird, an.

Bei der Korrelation mit dem Testrohwert als Binnenkriterium wird jede Aufgabe als Bestandteil des gesamten Wortschatztests mit diesem korreliert. Dabei liegt der Trennschärfekoeffizient etwas höher als eigentlich berechtigt, da jedes Item zum Teil mit sich selbst korreliert wird und dort selbstverständlich eine hundertprozentige Korrelation erzielt. Aus diesem Grund wurde eine Part-Whole-Korrektur durchgeführt, die bei der Auszählung des Rohwertes jeweils diejenige Aufgabe auslässt, für die der Trennschärfekoeffizient berechnet werden soll.

Schließlich wird eine Testaufgabe durch ihre Schwierigkeit beschrieben. Im allgemeinen sollte ein Test in Bezug auf die entsprechende Grundgesamtheit mittlere Schwierigkeit aufweisen, wobei jede einzelne Aufgabe daraus sowohl schwer (für Testpersonen mit geringer Merkmalsausprägung) als auch leicht (für Testpersonen mit hoher Merkmalsausprägung) sein kann.

Zwischen Trennschärfe und Schwierigkeit besteht generell eine parabolische Abhängigkeit, da die Höhe des Trennschärfekoeffizienten einer Aufgabe nicht nur von dem Gemeinsamkeitsgrad dessen, was Gesamttest und einzelne Aufgabe messen, abhängt, sondern in starkem Maße von dem Grad der Aufgabenschwierigkeit geprägt wird. Dies erklärt sich in der Eigenschaft der Schwierigkeit, zwischen denjenigen Testpersonen, die eine Aufgabe gelöst, und denjenigen, die sie nicht gelöst haben, zu differenzieren. Eine Aufgabe mit mittlerer Schwierigkeit unterscheidet demnach vielfältiger (maximal $50 \cdot 50 = 2500$ Mal) als jede andere Aufgabe (minimal $1 \cdot 99 = 99$ Mal, maximal $49 \cdot 51 = 2499$ Mal) und birgt somit bessere Voraussetzungen für eine hohe Trennschärfe. Mittlere Schwierigkeit allein garantiert jedoch noch keine hohe Trennschärfe. Zusätzlich müssen die Unterscheidungen gemäß den Erwartungen verlaufen, d. h. dass tatsächlich die 50%, die die Aufgabe gelöst haben, zu der Hälfte der Testpersonen mit starker Merkmalsausprägung, die 50%, die sie nicht gelöst haben, zu der Hälfte mit schwacher Merkmalsausprägung gehören. Erst dann hat eine Aufgabe ihrer Schwierigkeit entsprechend höchstmögliche Trennschärfe erreicht.

Die Kriterien Trennschärfe und Schwierigkeit sind im Rahmen der Aufgabenanalyse von entscheidender Bedeutung. Dementsprechend genau werden sie im weiteren Verlauf untersucht.

3.4.2 Aufgabentrennschärfe und –schwierigkeit

Als primäres Auswahlkriterium wird im Falle des Wortschatztests als ein homogener Niveautest von inhaltlicher oder trivialer Validität (vgl. Lienert/Raatz 1994, 115) die Trennschärfe angesetzt. Damit soll direkt auf die Reliabilität, die bei hohen Trennschärfekoeffizienten höher liegt, und immerhin mittelbar auf die Validität des Instruments Einfluss genommen werden, die zusätzlich an das Konstanthalten aller anderen Bedingungen gebunden ist.

Je höher der Trennschärfekoeffizient einer Aufgabe liegt, je deutlicher die Aufgabe also bezüglich des Gesamtpunktwertes zwischen guten (im Sinne von starker Merkmalsausprägung) und schlechten (im Sinne von schwacher Merkmalsausprägung) Testpersonen trennt, desto brauchbarer wird sie für den Wortschatztest. Liegt der Trennschärfekoeffizient nahe Null, erweist sich die entsprechende Aufgabe als unbrauchbar; gute und schlechte Testpersonen werden überhaupt nicht voneinander getrennt, da sie die Aufgabe gleichermaßen richtig bzw. falsch oder gar nicht beantwortet haben. Ein negativer Trennschärfekoeffizient bringt zum Ausdruck, dass die Aufgabe gerade von den guten Testpersonen falsch und von den schlechten richtig gelöst wurde und somit die Leistungsverhältnisse geradezu umgekehrt werden. Testaufgaben mit fehlender oder negativer Trennschärfe werden im Rahmen der Weiterentwicklung zu Testversion II eliminiert, bei Aufgaben mit geringer Trennschärfe wird entsprechend anderer Kriterien wie der Schwierigkeit abgewogen, ob sie eliminiert oder beibehalten und nach Möglichkeit verbessert werden.

Entscheidendes Auswahlkriterium ist demzufolge auch die Schwierigkeit einer Testaufgabe. Zu Beginn des Tests sind einige besonders leichte Aufgaben wünschenswert, die als ‘Eisbrecher-Aufgaben’ fungieren sollen. Abgesehen davon sollen die Schwierigkeitsindizes eines Niveautests grundsätzlich von mindestens $P=.20$ bis $P=.80$ streuen (vgl. Lienert/Raatz 1994, 115). Dabei sollen zumindest so viele leichte Aufgaben vorhanden sein, dass keine Testperson punktlos abschneidet, gleichzeitig soll der Test so viele schwere Aufgaben beinhalten, dass nur wenige Testpersonen alle Aufgaben richtig beantworten können. Entscheidend ist, dass eine Häufung der Schwierigkeitsindizes an der Stelle des Tests auftritt, an der er besonders gut differenzieren soll (bei erwünschter gleichmäßiger Differenzierung bei $P=.50$). Da es sich im vorliegenden Fall um ein Messinstrument zur ausdrücklichen Überprüfung des Grundwortschatzes handelt, kann davon ausgegangen werden, dass alle Testpersonen zumindest einen Teil der Aufgaben richtig beantworten können. Die Schwierigkeitsindizes streuen also erwartungsgemäß nicht über den gesamten Bereich, sondern vornehmlich im hohen Wertebereich, also bei geringer bis mittlerer Schwierigkeit. Da der Test im Sinne dieses Grundwortschatzes im Bereich der leichten Aufgaben präzise differenzieren soll, ist eine Häufung der Schwierigkeitsindizes an dieser Stelle wünschenswert. Gleichzeitig bedeutet dies jedoch einen stückweisen Verzicht auf hohe Trennschärfeindizes, da der oben beschriebene Zusammenhang zwischen Trennschärfe und Schwierigkeit einen besonders hohen

Trennschärfewert vor allem bei mittleren und nicht bei hohen Schwierigkeitsindizes (also leichten Aufgaben) zulässt.

In den folgenden Tabellen sind Schwierigkeits- sowie Trennschärfekoeffizienten der ursprünglichen 24 Aufgaben eines jeden Subtestteils aus Testversion I dargestellt. Sie umfassen drei mal drei zueinander in Beziehung stehende Spaltenpaare, wobei jeweils die Nomen (bzw. Verben und Adjektive) durch alle Modalitäten (Gebärden-, Schrift- und Lautsprache) in einer Darstellung zu finden sind. Die Aufgaben mit kritischen Werten werden durch Schattierung hervorgehoben.

Tab. 25a-B 3.4.2: Schwierigkeit und Trennschärfe Nomen (Testversion I)						
Aufgabe	GSN		SSN		LSN	
	p	r_{it}	p	r_{it}	p	r_{it}
1. Mutter	.63	.35	.47	.24	.52	.37
2. Auto	.98	.14	1	.00	.96	.32
3. Haus	1	.00	.93	.19	.80	.46
4. Kopf	.75	.04	.72	.41	.48	.33
5. Straße	.50	.16	.64	.55	.61	.53
6. Bett	.89	.28	.86	.47	.88	.36
7. Krone	.91	-.13	.58	.50	.69	.30
8. Angst	.84	.22	.61	.50	.76	.63
9. Garten	.47	.40	.46	.58	.59	.47
10. Stein	.66	.37	.60	.58	.74	.49
11. Schuh	.61	.33	.96	.27	.72	.44
12. Bart	.96	.22	.50	.47	.72	.36
13. Zahn	.83	.10	.76	.34	.82	.27
14. Leute	.81	.29	.40	.26	.56	.51
15. Ei	.99	.18	.96	.17	.89	.16
16. Zahl	.95	.07	.71	.44	.42	.38
17. Geld	.56	.07	.76	.50	.68	.32
18. Fisch	1	.00	.93	.23	.88	.27
19. Holz	.56	.47	.66	.62	.60	.13
20. Arbeit	.69	.29	.66	.50	.62	.43
21. Schloss	.86	-.03	.36	.55	.43	.41
22. Kaffee	.54	-.01	.44	.42	.52	.26
23. Rose	.71	.20	.65	.44	.68	.31
24. Welt	.71	.29	.60	.56	.78	.38

Tab. 25b-B 3.4.2: Schwierigkeit und Trennschärfe Verben (Testversion I)						
Aufgabe	GSV		SSV		LSV	
	p	r _{it}	p	r _{it}	p	r _{it}
1. sagen	.48	.29	.36	.11	.43	.40
2. gehen	.90	.52	.69	.53	.47	.23
3. fahren	.99	- .01	.77	.51	.69	.43
4. essen	.97	.14	.86	.28	.88	.37
5. fliegen	.99	.30	.74	.55	.82	.32
6. gucken	.68	.41	.28	.15	.42	.31
7. springen	.96	.07	.65	.62	.78	.43
8. schlafen	.76	.22	.74	.51	.77	.32
9. denken	.46	.36	.23	.18	.29	.25
10. hinsetzen	.36	.20	.21	.37	.40	.28
11. anziehen	.79	- .13	.49	.57	.60	.48
12. warten	.75	.44	.57	.44	.58	.28
13. verstecken	.87	.27	.57	.56	.69	.44
14. weinen	.96	.08	.75	.53	.77	.32
15. werfen	.96	.38	.66	.43	.72	.37
16. schreiben	.96	.36	.71	.61	.85	.29
17. lesen	.96	.30	.77	.59	.77	.50
18. singen	.85	.33	.42	.64	.50	.39
19. waschen	.97	.10	.61	.64	.72	.50
20. brennen	.94	.18	.48	.48	.56	.42
21. schieben	.94	.29	.31	.27	.65	.41
22. stolpern	.70	.13	.32	.18	.43	.17
23. stinken	.96	.20	.44	.38	.82	.27
24. zittern	.48	.38	.14	.33	.28	.07

Tab. 25c-B 3.4.2: Schwierigkeit und Trennschärfe Adjektive (Testversion I)						
Aufgabe	GSA		SSA		LSA	
	p	r _{it}	p	r _{it}	p	r _{it}
1. ganz	.46	- .001	.36	.36	.26	.26
2. klein	.89	.17	.73	.42	.52	.29
3. blau	.92	.14	.84	.29	.97	- .10
4. voll	.44	.43	.54	.58	.59	.46
5. blond	.46	.46	.73	.46	.58	.50
6. spitz	.91	.16	.66	.49	.71	.41

Aufgabe	GSA		SSA		LSA	
	p	r _{it}	p	r _{it}	p	r _{it}
7. nass	.55	.43	.58	.53	.66	.31
8. hellblau	.79	.42	.66	.56	.81	.33
9. orange	.91	.14	.60	.30	.60	.51
10. dünn	.93	.03	.63	.50	.51	.45
11. schwer	.17	.25	.52	.64	.42	.41
12. traurig	.66	-.06	.62	.67	.71	.30
13. verletzt	.45	.45	.48	.51	.59	.39
14. warm	.79	.27	.57	.51	.66	.53
15. schief	.68	.07	.54	.42	.43	.51
16. glatt	.96	.09	.54	.36	.79	.35
17. links	.19	.24	.31	.36	.30	.45
18. frei	.41	.08	.65	.27	.40	.32
19. schmutzig	.77	.27	.68	.51	.76	.48
20. weich	.59	.31	.52	.57	.41	.56
21. leise	.92	.11	.59	.55	.84	.26
22. gestreift	.72	.19	.22	.30	.50	.56
23. viereckig	.92	.14	.65	.45	.54	.47
24. nah	.46	.20	.29	.34	.43	.30

Zusätzlich sind die Beziehungen der Trennschärfe- und Schwierigkeitsindizes der ursprünglichen 24 Aufgaben eines jeden Subtestteils anschaulich in den im Anhang (Anh. 1: B 3.4.2) einzusehenden Koordinatensystemen (Abb. 12a-i-B 3.4.2) dargestellt.

Da mit Hilfe des Wortschatztests explizit der Grundwortschatz überprüft werden soll, indem häufig verwendete und demzufolge leichte Items geprüft wurden, ist eine Häufung vornehmlich im hohen Schwierigkeits- und damit im mittleren Trennschärfewertebereich gerechtfertigt. Die generell erwartete parabolische Abhängigkeit zwischen Schwierigkeit und Trennschärfe muss demnach für den Wortschatztest einer Verteilungsform weichen, die im Koordinatensystem vor allem im rechten Bereich der Abzisse und im mittleren Bereich der Ordinate differenziert.

Selbstverständlich können sowohl Trennschärfe- als auch Schwierigkeitswerte nur in Relation zueinander betrachtet werden, sodass nach Eliminierung diesbezüglich ungeeigneter Aufgaben in der Endversion des Instrumentes in Ermangelung besserer Werte möglicherweise Aufgaben mit geringer Trennschärfe- und/oder Schwierigkeitsausprägung erhalten bleiben. Dies geschieht unter Berücksichtigung des Umstandes, dass ein Test mit zu wenigen

Aufgaben nur zufällige Augenblicksmomente widerspiegeln kann, was zu Lasten einer objektiven Einschätzung der Fähigkeiten einer Testperson fällt. Verstärkt wird dieser Sachverhalt durch die konzeptionelle Zusammensetzung des Itemkorpus aus Nomen, Verben und Adjektiven zu gleichen Anteilen, die bei Eliminierung eines Items zum gleichzeitigen Ausschluss von zwei weiteren Aufgaben der jeweils anderen Wortarten zwingt. Nicht auf jede Aufgabe mit nicht optimalen Werten kann demnach verzichtet werden, sodass ein Zustandekommen von Kompromissen unausweichlich ist. Der Umstand, dass bei der Beurteilung sämtlicher Aufgaben auch die psychologische Validität (Lienert/Raatz 1994, 116) von Bedeutung war, rechtfertigt die Übernahme dieser Aufgaben in die endgültige Testversion.

3.4.3 Aufgaben-Interkorrelationen

Zusätzlich zur Berechnung von Schwierigkeit und Trennschärfe wurden für die Aufgabenanalyse die korrelativen Zusammenhänge je zweier Aufgaben untersucht. Diese Aufgaben-Interkorrelationen sollen Aufschluss darüber geben, inwieweit je zwei Aufgaben gleiche bzw. verschiedene Aspekte des untersuchten Merkmals erfassen. Homogene Testaufgaben werden dementsprechend höher korrelieren als heterogene.

Die Aufgaben-Interkorrelation übt einen entscheidenden Einfluss auf die Aufgabenschwierigkeit aus. Diese ist nur dann zu so vielen Differenzierungen fähig, wenn gleichzeitig die Aufgaben-Interkorrelation niedrig ist. Ist dies nicht der Fall, wären also alle Aufgaben mit hoher Interkorrelation belegt, so würde eine Testperson, die eine Aufgabe löst, automatisch auch alle anderen richtig beantworten, der Test wäre also nur zu zwei Differenzierungen (alles richtig – alles falsch) fähig. Eine solche vollständige Interkorrelation und zugleich eine vollständige Homogenität der Aufgaben ist nur dann möglich, wenn alle Aufgaben gleich hohe Schwierigkeit aufweisen. Je größer die Schwierigkeitsunterschiede der Aufgaben sind, desto geringer werden die Interkorrelationskoeffizienten ausfallen. Die relativ niedrigen Interkorrelationen der inhaltlich homogen erscheinenden Aufgaben des Wortschatztests lassen also bereits darauf schließen, dass die Aufgaben dieses Tests von sehr unterschiedlicher Schwierigkeit sein müssen. Gute Differenzierung bzw. hohe Reliabilität ist demnach entweder bei unkorrelierten (heterogenen) Aufgaben gleicher Schwierigkeit oder, wie im vorliegenden Fall, bei mäßig positiv korrelierten (homogenen) Aufgaben unterschiedlicher Schwierigkeit gewährleistet.

Da sich die Anzahl der Kombinationsmöglichkeiten je zweier Aufgaben bei steigender Aufgabenanzahl potenziert, bedeutet die Aufgaben-Interkorrelation einen enormen Zeitaufwand. Für die 72 Aufgaben pro Subtest des Wortschatztests gibt es 2556, für den Gesamttest mit 216 Aufgaben sogar 23220 Kombinationsmöglichkeiten. Die Berechnung der Aufgaben-Interkorrelation wird dementsprechend selten (vgl. Lienert/Raatz 1994, 98) und an dieser Stelle

nur exemplarisch für den ersten Subtestteil GSN durchgeführt, da die Interkorrelationsbeziehungen in allen weiteren Subtestteilen ähnlich gelagert sind.

Die im Anhang (Anh. 1: B 3.4.3) aufgeführte Tabelle (Tab. 26-B 3.4.3) veranschaulicht diese interkorrelativen Zusammenhänge zwischen den ursprünglichen 24 Aufgaben des ersten Subtestteils GSN aus Testversion I. Aufgaben mit nicht vorhandener Trennschärfe können entsprechend nicht korrelieren, diese Fälle sind mit einem Leerzeichen (-) gekennzeichnet.

Grundsätzlich können bei der Betrachtung einzelner Aufgaben Zufälle stärkeren Einfluss ausüben als auf den Gesamtkorpus. Da die Schwierigkeitsindizes der Aufgaben des Wortschatztests zudem weit streuen und Items mit verschieden hoher Schwierigkeit nur mäßig miteinander korrelieren, können die Aufgaben-Interkorrelationen nicht hoch ausfallen. Darüber hinaus sind nur wenige dieser Interkorrelationen so signifikant, dass sie einen wahren Zusammenhang abbilden. Nur diejenigen Aufgaben, die wenigstens auf dem 5%- (gekennzeichnet durch *) oder auf dem 1%-Niveau (gekennzeichnet durch **) gesichert sind, liefern einen interpretierbaren Wert. Im vorliegenden Fall sind es erwartungsgemäß gerade diese, die die mäßig positiven Korrelationen liefern, alle anderen können als Nullkorrelationen angesehen werden und unbeachtet bleiben.

Zusätzliches Motiv für die Aufgabenanalyse war es, die Aufgaben, die mehrere negative, nicht signifikante Interkorrelationswerte zu anderen Aufgaben aufweisen, zu eliminieren, und gerade die Aufgaben in die Testversion II aufzunehmen, die signifikante und positive Interkorrelationswerte liefern.

3.4.4 Aufgabenstreuung und -reliabilität

Neben Aufgabentrennschärfe, -schwierigkeit und -interkorrelation spielt auch die Aufgabenstreuung eine Rolle bei der Beurteilung der Qualität einer Aufgabe. Die Streuung einer Aufgabe beschreibt die Standardabweichung ihrer Punktbewertung. Bei mittlerer Schwierigkeit wird die Aufgabenstreuung mit $s=.50$ maximal hoch. Je stärker sich die Schwierigkeit dem Wert 0 oder 1 nähert, desto geringer wird die Aufgabenstreuung. Die Aufgabe hört dann auf zu streuen, da sie entweder immer oder nie richtig beantwortet worden ist.

Multipliziert man die Aufgabenstreuung mit dem Trennschärfewert der Aufgabe, so erhält man eine Maßzahl für die Reliabilität einer Aufgabe, den sogenannten Reliabilitätsindex (vgl. Lienert/Raatz 1994, 111) einer Aufgabe, der, mal zwei genommen, im Sinne eines Korrelationskoeffizienten verstanden wird. Dieser Reliabilitätsindex ist für die Aufgabenanalyse des Wortschatztests von Bedeutung, denn die Reliabilität einer Aufgabe trägt mittelbar zur

Reliabilität des gesamten Tests bei. Ein entsprechender Validitätsindex einer Aufgabe kann im vorliegenden Fall nicht berechnet werden, da die Aufgabenanalyse nicht an einem Außenkriterium durchgeführt wurde.

Tabellen (Tab. 27a-i-B 3.4.4), die Schwierigkeits-, Trennschärfe-, Streuungs- und Reliabilitätsindizes jeder Aufgabe in jedem Subtestteil aufführen, befinden sich im Anhang (Anh. 1: B 3.4.4).

Bei den folgenden Tabellen handelt es sich jeweils um eine zusammenfassende Darstellung dieser Wertetabellen zu den drei Subtestteilen einer Wortart (Nomen bzw. Verben oder Adjektive in Gebärden-, in Schrift- und in Lautsprache). Sie enthalten die drei Reliabilitätsindizes einer Wortart sowie einen Gesamtreliabilitätsindex, jeweils berechnet aus dem Mittelwert der drei Indizes der Subtestteile. Zusätzlich ist in der letzten Spalte eine Rangfolge der Aufgaben bzgl. ihrer Reliabilitätsindizes abzulesen. Je kleiner der Index desto geringer die Reliabilität und desto höher der Rangplatz.

Tab. 28a-B 3.4.4: Gesamtreliabilität Nomen					
Aufgabe	GSN	SSN	LSN	ges.	Rang
1	.17	.12	.18	.16	9
2	.02	.00	.06	.03	24
3	.00	.05	.18	.08	21
4	.02	.19	.17	.12	15
5	.08	.27	.26	.20	3
6	.09	.17	.12	.12	14
7	-.04	.25	.14	.12	16
8	.08	.25	.27	.20	4
9	.20	.29	.23	.24	1
10	.18	.29	.22	.23	2
11	.16	.04	.09	.10	19
12	.04	.23	.07	.12	17
13	.04	.14	.10	.10	20
14	.11	.13	.25	.16	8
15	.02	.03	.05	.03	23
16	.02	.20	.19	.13	12
17	.02	.22	.15	.13	13
18	.00	.06	.09	.05	22
19	.23	.29	.06	.20	5
20	.13	.24	.21	.19	6

Aufgabe	GSN	SSN	LSN	ges.	Rang
21	-.01	.26	.20	.15	10
22	-.004	.21	.13	.11	18
23	.09	.21	.15	.15	11
24	.13	.28	.16	.19	7

Tab. 28b-B 3.4.4: Gesamtreliabilität Verben					
Aufgabe	GSV	SSV	LSV	ges.	Rang
1	.14	.05	.20	.13	16
2	.16	.24	.12	.17	6
3	-.004	.21	.20	.14	15
4	.02	.10	.12	.08	23
5	.09	.24	.12	.15	10
6	.19	.07	.16	.14	14
7	.01	.29	.18	.16	8
8	.09	.22	.13	.15	11
9	.18	.08	.11	.12	20
10	.10	.15	.14	.13	18
11	-.05	.29	.23	.16	9
12	.19	.22	.14	.18	4
13	.09	.28	.20	.19	2
14	.02	.23	.13	.13	19
15	.07	.20	.17	.15	13
16	.07	.28	.10	.15	12
17	.06	.25	.21	.17	5
18	.12	.32	.19	.21	1
19	.02	.31	.22	.19	3
20	.04	.24	.21	.16	7
21	.07	.12	.19	.13	17
22	.06	.08	.09	.08	24
23	.04	.19	.11	.11	22
24	.19	.12	.03	.11	21

Tab. 28c-B 3.4.4: Gesamtreliabilität Adjektive					
Aufgabe	GSA	SSA	LSA	ges.	Rang
1	-.005	.172	.12	.09	23
2	.05	.185	.15	.13	20
3	.04	.106	-.02	.04	24
4	.21	.291	.23	.24	1
5	.23	.205	.25	.23	3
6	.05	.234	.19	.16	15
7	.22	.261	.15	.21	5
8	.17	.265	.13	.19	8
9	.04	.144	.25	.15	16
10	.01	.242	.23	.16	13
11	.10	.319	.20	.21	6
12	-.03	.325	.14	.14	17
13	.23	.243	.19	.22	4
14	.11	.250	.25	.20	7
15	.03	.207	.25	.16	10
16	.02	.180	.14	.11	21
17	.09	.167	.21	.16	14
18	.04	.129	.16	.11	22
19	.11	.240	.20	.19	9
20	.15	.286	.28	.24	2
21	.03	.271	.10	.13	19
22	.09	.124	.28	.16	11
23	.04	.214	.24	.16	12
24	.10	.155	.15	.14	18

Im Zuge der Aufgabenanalyse wurden Aufgaben mit niedrigen Rangfolgeplätzen denjenigen mit hohen vorgezogen (sofern keine anderweitigen Gebote dies verhinderten), da sie eine höhere Aufgabenreliabilität gewährleisten und somit für einen Test mit möglichst hoher Reliabilität von Vorteil sind.

3.4.5 Distraktorenanalyse

Bei der Distraktorenanalyse im klassischen Sinn (vgl. Lienert/Raatz 1994, 124) werden Schwierigkeits- und Trennschärfewerte nicht nur für die richtige Antwort, sondern auch für alle Distraktoren berechnet, um eine systematische Überprüfung der Eignung von Ablenkern vornehmen zu können. Dabei wird eine Aufgabe als ideal bewertet, wenn alle Distraktoren

den gleichen Schwierigkeitsgrad aufweisen und die Trennschärfe aller Distraktoren – im Gegensatz zu der erwarteten positiven Trennschärfe des Items – gleichermaßen im negativen Bereich oder nahe Null liegt. Im Falle des Wortschatztests mussten aufgrund konzeptioneller Besonderheiten diese Bewertungsmaßstäbe zur Analyse der Distraktoren geringfügig verändert werden. So wurden die Ablenker nach verschiedenen Ähnlichkeitskriterien qualitativ abgestuft ausgewählt. (Die Einzelheiten dieser Konzeption ist dem Kapitel B 3.10 zu entnehmen und soll hier nicht erläutert werden, da sie im Rahmen der Aufgabenanalyse nicht von Bedeutung sind.) Eine rein quantitative Beurteilung der Eignung von Distraktoren aufgrund einer anzustrebenden gleichmäßigen Verteilung der Schwierigkeit auf alle Distraktoren wäre kontraindiziert zu dieser Konzeption. Indem gezielt für jedes Item in jeder Modalität einige Ablenker plausibler gestaltet wurden als andere, sollen über das individuelle Distraktorenprofil einer Testperson Informationen über ihre spezifischen Kenntnisse und Schwierigkeiten in unterschiedlichen Modalitäten und Wortarten gewonnen werden. Im Rahmen der Distraktorenanalyse des Wortschatztests richtet sich die Beurteilung der Distraktoren demnach nicht nach einer möglichst gleichhäufigen Wahl aller Distraktoren, sondern nach einer jeweils gerechtfertigt hohen Wahlhäufigkeit jedes Distraktors in jeder Modalität. So ist beispielsweise die Wahl des Gebärdenablenkers im Subtest Gebärdensprache naheliegend, da auch die Itemvorgabe in gebärdeter Form erfolgte und sich die Testperson somit in ihrer Antwortwahl möglicherweise an der entsprechenden Modalität orientiert hat. Im Subtest Lautsprache ist der gebärdenähnliche Distraktor dagegen als weniger plausible Alternative einzustufen, da die Vorgabe des Items in gesprochener Form ohne zusätzliche Gebärde erfolgte. In diesem Fall wäre eine zu häufige Wahl des gebärdenähnlichen Distraktors demzufolge weniger angemessen. Aufgaben, deren Ablenker der jeweiligen Modalität entsprechend verteilte Fehleranzahl liefern (vgl. B 3.2.2), wurden im Zuge der Aufgabenanalyse denjenigen Aufgaben mit nicht nachvollziehbarem Zustandekommen von Wahlhäufigkeiten vorgezogen.

Zusätzliches Beurteilungskriterium einer Aufgabe waren die Trennschärfewerte ihrer Distraktoren. Dabei wurde nicht zwingend für jeden Distraktor gesondert, sondern im Falle von mehreren Distraktoren eines Ähnlichkeitstyps innerhalb einer Aufgabe für deren Summenwert ein Trennschärfeindex berechnet. Aus diesen Werten lässt sich ablesen, wie attraktiv eine fehlerhafte Antwort für die in Bezug auf den Gesamttest leistungsstarke Testpersonen ist. Ein hoher Trennschärfewert eines Distraktors bringt zum Ausdruck, dass dieser einzelne Distraktor oder eine Gruppe von Distraktoren mit gleichartiger Ähnlichkeit zum Item zu attraktiv ist und gute Testpersonen dadurch auf die Wahl einer falschen Antwort gedrängt werden. Im Zuge der Eliminierung von Aufgaben wurden diejenigen Aufgaben als auffallend schlecht bewertet und u. U. eliminiert, die einen Distraktor mit positiver Trennschärfe über .40 oder mehr als einen Distraktor mit positiver Trennschärfe über .30 besitzen. Grundsätzlich wurden Items mit Distraktoren mit negativen Trennschärfewerten oder Werten nahe

Null denen mit höheren Werten möglichst vorgezogen. Einzelne Aufgaben mit hoher Distraktorentrennschärfe wurden geringfügig geändert, wenn dadurch eine Minderung der Attraktivität des Distraktors zu erwarten war.

Schließlich wurde im Rahmen der Distraktorenanalyse ergänzend untersucht, ob einzelne Ablenker deutlich von der kognitiven Entwicklung einer Testperson abhängig sind und somit möglicherweise ein anderes Konstrukt als die passive Wortschatzkenntnis messen. Dafür wurden gemittelte Werte der gewählten Distraktoren mit den Ergebnissen eines zusätzlich erhobenen, non-verbalen Intelligenztests dem Raven (Raven-Matrizen-Test, CPM 1980) korreliert. Ein signifikant hoher Koeffizient dieser Korrelation würde den unerwünschten Effekt zum Ausdruck bringen, dass besonders die nach den Testergebnissen des Raven intelligenten Kinder von diesem Distraktoren abgelenkt werden. Es wurde neben anderen Kriterien versucht, Aufgaben mit niedrigen und/oder nicht signifikanten Distraktoren-Raven-Korrelationen denen mit signifikant hohen Korrelationskoeffizienten vorzuziehen.

Tabellen mit entsprechenden Werten der Distraktorentrennschärfe (Tab. 29a-e-B 3.4.5) und der Korrelationen zwischen Distraktoren und Raven (Tab. 30a-e-B 3.4.5) sind im Anhang (Anh. 1: B 3.4.5a und Anh. 1: B 3.4.5b) aufgeführt.

3.4.6 Sonstige Kriterien

Zusätzliches Kriterium für die Übernahme einer Aufgabe in die endgültige Testversion des Wortschatztests war die Gewährleistung des Vorkommens aller drei Ablenkerarten. Aufgaben, für die jeweils mindestens ein gebärdenähnlicher sowie ein mundbildähnlicher und ein inhaltlich ähnlicher Distraktor bereitgestellt werden konnten, wurden denen mit einem (oder sogar zwei) fehlenden Distraktorentyp(en) vorgezogen. Nur in Ausnahmefällen wurden Items mit einseitigem Distraktorenbild zugunsten ihrer psychologischen Validität (vgl. Lienert/Raatz 1994, 36) in die Testversion II übernommen. Schließlich wurde für alle Aufgaben der (vorerst) endgültigen Version angemessene Schwierigkeit eines Items in allen Modalitäten vorausgesetzt. Aufgaben, die nur in einzelnen Modalitäten als sehr leicht oder sehr schwer empfunden wurden, wurden denen, die durchgängig extreme Schwierigkeitswerte auswiesen, vorgezogen.

3.5 Konsequenzen der Aufgabenanalyse

Ziel der Aufgabenanalyse war die Eliminierung ungeeigneter und die Revision verbesserungsbedürftiger Aufgaben des Wortschatztests für die Erstellung einer im Rahmen der vor-

liegenden Arbeit endgültigen Testversion II. Neben dem Hauptkriterium der Trennschärfe wurden die anderen aufgeführten Aspekte einer Aufgabe zu ihrer Beurteilung herangezogen. Für jede Aufgabe wurden die Vor- und Nachteile gegeneinander abgewogen und die Eliminierung bzw. Änderung dementsprechend vorgenommen.

3.5.1 Aufgabenselektion

In der Wortart der Nomen erweisen sich sechs Items (Nr. 2, 3, 7, 18, 21, 22) durch Trennschärfewerte in mindestens einer Modalität im negativen Bereich oder um Null als absolut ungeeignet. Weitere zwei Aufgaben (Nr. 4, 15) erscheinen aufgrund niedriger Trennschärfewerte nicht optimal geeignet, ohne dabei gravierend schlechte Werte aufzuweisen. Da der Itemkorpus laut Konzeption des Tests in den drei Wortarten gleiche Itemanzahl aufweisen soll, würde das Eliminieren einer weiteren Aufgabe gleichzeitig den Verzicht auf je eine Aufgabe aus den beiden übrigen Wortarten bedeuten, die aufgrund ihrer Werte nicht zwingend ausgeschieden werden müssten. Es müssen demnach stets drei Aufgaben gleichzeitig ausgesondert werden, was zu einer gravierenden Reduzierung des Korpus führt. Unter Berücksichtigung dieser Tatsache und den verschiedenen Aspekte der Aufgabenanalyse wurde beschlossen, jeweils sechs Items pro Wortart zu eliminieren. Der Itemkorpus der Version I war bereits auf drei mal 24 Aufgaben aufgestockt worden, um nach Eliminierung der ungeeigneten Items genügend Aufgaben zurückzubehalten. Die so entstehende Anzahl von drei mal 18, also 54 Aufgaben, erscheint aufgrund der Testerfahrungen mit der Pilotversion (ebenfalls 54 Aufgaben) und Version I (72 Aufgaben) geeignet, ein differenziertes Bild über den Leistungsstand der Testperson zu liefern, ohne eine Überforderung darzustellen.

Folgende Aufgaben wurden von Version I zu Version II eliminiert:

Tab. 31-B 3.5.2: Eliminierte Aufgaben			
Nummer	Nomen	Verben	Adjektive
1	2	1	1
2	3	3	3
3	7	4	12
4	18	11	17
5	21	22	18
6	22	24	21

Hauptausschlusskriterium war in jedem Fall die Trennschärfe, eine Gewichtung der übrigen Mängel wurde jeweils gesondert vorgenommen. Dabei war, mit Ausnahme der Fälle negativer oder nicht vorhandener Trennschärfe, stets ein Zusammenwirken mehrerer Gründe oder das Vorkommen eines einzelnen Grundes in mehreren Modalitäten notwendig, um eine Auf-

gabe als ungeeignet zu bewerten und auszuschließen. Alle entscheidenden Gründe werden in den folgenden Tabellen nach Wortarten getrennt für jede eliminierte Aufgabe, bezogen auf ihr Vorkommen, in einzelnen Modalitäten aufgeführt.

Tab. 32a-B 3.5.2: Gründe für Eliminierung von Nomen						
Nr.	negative Trennschärfe	keine Trennschärfe	geringe Trennschärfe	fehlender Ablenker	zu leicht	positive Distr.-trennschärfe
2		Schriftsprache		Mund-/Schriftbild	in allen Modalitäten	
3		Gebärdensprache	Schriftsprache		Gebärdens-/Schriftsprache	Mundbild-distraktor
7	Gebärdensprache					
18		Gebärdensprache			in allen Modalitäten	
21	Gebärdensprache					Mundbild-distraktor
22	Gebärdensprache					

Tab. 32b-B 3.5.2: Gründe für Eliminierung von Verben					
Nr.	negative Trennschärfe	geringe Trennschärfe	zu leicht	zu schwer	positive Distr.-trennschärfe
1		Schriftsprache		in allen Modalitäten	
3	Gebärdensprache		Gebärdens-/Schriftsprache		Mundbild-distraktor
4		in allen Modalitäten	in allen Modalitäten		Inhaltsdistraktor
11	Gebärdensprache				Gebärdensdistraktor
22		in allen Modalitäten			
24		Lautsprache		Schriftsprache	

Tab. 32c-B 3.5.2: Gründe für Eliminierung von Adjektiven							
Nr.	negative Trennschärfe	geringe Trennschärfe	Ablenker fehlt	zu leicht	zu schwer	pos. Distr.-trennschärfe	sign. pos. Korr. Distr.-Raven
1	Gebärdensprache				in allen Mod.		
3	Lautsprache			in allen Mod.			Lautsprache I,M,MOD
12	Gebärdensprache					Inhaltsablenker	Gebärdensprache HRP
17			Mund-/Schriftbild		in allen Mod.	Inhaltsablenker	
18		Gebärdensprache			Gebärdensprache		Gebärdensprache I
21		Gebärdensprache	Gebärdensprache				

Bei der so entstandenen Auswahl zu eliminierender Items wurden Aufgaben aus dem Itemkorpus entfernt, die auch bzgl. Aufgaben-Interkorrelation und Aufgabenstreuung bzw. -reliabilität schlechte Werte aufwiesen. Entsprechend der in Kapitel B 3.4.4 berechneten Rangplätze aufgrund der Itemreliabilität wurden vielfach die Items mit den höchsten Rangplätzen, also mit niedrigen Reliabilitätsindizes eliminiert. Wären im optimalen Fall die diesbezüglich jeweils sechs schlechtesten Items (also die mit Rangplatz 19 bis 24) eliminiert worden, so würde dies einen durchschnittlich eliminierten Rangplatz von 21,5 bedeuten. Tatsächlich wurden Nomen mit den Rangplätzen 24, 21, 16, 22, 10, 18, also durchschnittlich Rangplatz 18,5; Verben mit den Rangplätzen 16, 15, 23, 9, 24, 21, damit durchschnittlich Rangplatz 18, und Adjektive mit den Rangplätzen 23, 24, 17, 14, 22, 19, also durchschnittlich Rangplatz 19,8, eliminiert. Es bleiben demnach Items mit verhältnismäßig hohen Reliabilitätsindizes im Itemkorpus der Testversion II erhalten.

3.5.2 Aufgabenrevision

Grundsätzlich sprachen im vorliegenden Fall zwei Gründe für eine Aufgabenrevision: Zum einen war abzusehen, dass die Werte einiger Aufgaben durch Revision noch zu verbessern

sein würden, zum anderen sollten bestimmte, bisher nicht geeignete Aufgaben zugunsten der psychologischen oder inhaltlichen Validität beibehalten werden. Eine Revision durch Neugestaltung ganzer Aufgaben war im fortgeschrittenen Entwicklungsstadium jedoch nicht mehr angezeigt, da derart revidierte Aufgaben erneut empirischer Überprüfung bedürfen. Statt dessen wurde eine Revision durch Distraktoren-Veränderung einzelner Aufgaben vorgenommen. Dabei wurden keine neuen Distraktoren formuliert, sondern lediglich in Einzelfällen Item- oder Distraktorendarstellungen geringfügig verändert, wenn eine sehr hohe oder sehr geringe Attraktivität des Distraktors erkennbar war.

Folgende Änderungen wurden von Version I zu Version II vorgenommen. Sofern nicht ausdrücklich Bezug auf das Testheft II (Schriftsprache) genommen wird, beziehen sich alle Änderungen auf das Testheft I, in dem vier Zeichnungen auf einer Seite abgebildet sind:

Nomen 1. ‘Mutter’

Der inhaltliche Distraktor ‘Mädchen’ wurde im Schriftsprachteil nach wie vor unverhältnismäßig häufig (77,6% aller Fehler) gewählt. Dies wurde auf die zeichnerische Darstellung der zu jungen ‘Mutter’ mit Baby zurückgeführt, die ebenso als großes ‘Mädchen’ mit Puppe gelten könnte. Eine Minderung der zu großen Attraktivität des Distraktors ‘Mädchen’ wurde durch eine Anpassung der Zeichnungen ermöglicht, indem sich die ‘Mutter’ altersmäßig deutlicher von dem ‘Mädchen’ abhebt und das Baby menschlichere Züge erhält.

Nomen 4. ‘Kopf’

Eine zu starke Ähnlichkeit des Distraktors ‘Zopf’ in inhaltlicher sowie in mundbildartiger Weise wurde als Grund dafür angesehen, dass das Item ‘Kopf’ trotz sehr frühem Vorkommen in der Wortliste von den Testpersonen als schwierig empfunden wird. Dem wurde entgegengewirkt, indem sich die neue Darstellung der Haare des ‘Kopfes’ deutlicher von einem ‘Zopf’ unterscheidet. Zusätzlich wurden die Positionen der Distraktoren ‘Junge’ und ‘Topf’ ausgetauscht, um die bevorzugte – da in Leserichtung als erste ins Auge fallende – Platzierung des für den Lautsprachteil stark attraktiven Distraktors ‘Topf’ (47,3% aller Fehler) oben links auf der Seite für diesen Distraktor auszuschließen.

Nomen 9. ‘Garten’

Gemäß der Systematik der schlichten Darstellung des Wesentlichen wurde die Darstellung des Distraktors ‘Geschenk’ ohne Blumenvase vorgenommen.

Nomen 10. ‘Stein’

Da der Gebärdendistraktor ‘Kaffee’ gerade in der gebärdensprachlichen Modalität ausgesprochen selten gewählt wurde (5,3% der Fehler), wurde eine zu große Attraktivität der übrigen Darstellungen angenommen. Diese wurden nach Möglichkeit im Sinne einer wenig aus-

geschmückten Darstellung verändert. Für das Item ‘Stein’ selbst geschah dies durch eine nur angedeutete Darstellung der Gräser im Hintergrund, die jedoch für den Größenvergleich erhalten bleiben müssen. Für den Distraktor ‘Stange’ mußte die den Sinn vermittelnde Schaukel ebenfalls unauffälliger und heller gestaltet werden.

Nomen 19. ‘Holz’

In der Modalität Gebärdensprache wurde der gebärdenähnliche Distraktor ‘Brot’ auffallend häufig gewählt (85,4% aller Fehler). Um seine Attraktivität etwas zu mindern, wurde die Darstellung des Messers als Blickfang am ‘Brot’ unauffälliger und realitätsnah vorgenommen.

Nomen 20. ‘Arbeit’

Auch in diesem Fall sollten die Distraktoren eindeutig dargestellt werden und nicht zuviel schmückendes Beiwerk aufweisen. Die Halme des Distraktors ‘Stein’ sowie das Auto des Distraktors ‘Ampel’ wurden demzufolge heller und unauffälliger gezeichnet.

Verb 8. ‘schlafen’

In den Subtests Gebärden- und Lautsprache, in denen alle vier Darstellungen im Testheft zu sehen sind, wurde der Distraktor ‘liegen’ auffallend häufig (96,2% und 66,7% aller Fehler) gewählt. Die Ähnlichkeit zwischen den Darstellungen von ‘schlafen’ und ‘liegen’ war so gravierend, dass nur bei genauem Hinsehen ein Unterschied erkennbar wurde. Um einem Testen von Konzentration entgegenzuwirken, musste sich die Darstellung von ‘liegen’ deutlicher vom Item ‘schlafen’ unterscheiden. Das Item wird in Version II als eine auf einer Gartenliege liegenden Person mit deutlich geöffneten Augen dargestellt.

Verb 14. ‘weinen’

Der Distraktor ‘schwitzen’ wurde in den Modalitäten Gebärden- und Lautsprache, in denen der Testperson alle vier Zeichnungen vorliegen, auffallend häufig (75,0% und 52,0% aller Fehler) gewählt. Die Darstellung musste sich deutlicher von dem Item ‘weinen’ absetzen, indem die dargestellte Person weniger verzweifelt guckt und keine Schweißtropfen im Gesicht hat.

Verb 15. ‘werfen’

Im Subtest Schriftsprache wurde auffallend häufig (70,3% aller Fehler) der Distraktor ‘fangen’ gewählt. Die Testpersonen erkannten demnach in dem Werfer vermutlich einen Fänger, da – wie im Testheft II der Fall – die zeichnerische Darstellung der Distraktoren, also auch der fangenden Person, fehlt. Die dargestellte Person bei ‘werfen’ sollte demzufolge durch eine hängende Hand den Prozess des Werfens deutlich abgeschlossen haben, um mit der Position der Hand nicht das Fangen eines Balls zu kennzeichnen.

Verb 19. 'waschen'

Die Darstellung der Person des Distraktors 'schwimmen' zog während der Untersuchung die Aufmerksamkeit vieler Testpersonen auf sich, da sie für die Benutzung eines Schwimmreifens zu alt erschien. Durch Umwandlung der Person in ein Kind wurde eine entsprechende Attraktivitätsminderung des Distraktors im Lautsprachteil (60% aller Fehler) vorgenommen.

Adjektiv 2. 'klein'

Obwohl die Darstellung von 'klug' Ähnlichkeiten in der Gebärde und im Mundbild aufweist, wurde dieser Distraktor in den Subtests Gebärden- und Lautsprache verhältnismäßig selten (16,7% und 15,7% aller Fehler) gewählt. Dies wurde darauf zurückgeführt, dass einerseits die Darstellung der 'klugen' Person mit Hilfe von Ausrufezeichen erfolgte, welche als Symbole missverstanden bzw. überhaupt nicht eingeordnet werden können, und dass andererseits die Darstellung der 'dummen' Person zu sehr einem Betrunkenen ähnelte. Die Darstellung dieses Gegensatzpaares musste entsprechend überarbeitet werden. Gleichzeitig fiel auf, dass im Gebärdensprachteil statt des Gebärdensablenkers der inhaltliche Distraktor 'dünn' häufig (58,3% aller Fehler) gewählt wurde. Die Darstellung des Gegensatzpaares musste durch sich in Größe und Position ähnelnde Hunde erfolgen, die sich lediglich durch ihre Statur (dick/dünn) unterscheiden. Somit wurde eine präzisere Unterscheidung von groß/klein gewährleistet. Die Attraktivität des Distraktors 'fein' wurde im Lautsprachteil neben der deutlichen Mundbildähnlichkeit durch Haltung und modische Kleidung der Person provoziert (70,6% aller Fehler). Zur Minderung der Attraktivität wird die feine Person in Testversion II mit klassischem Anzug dargestellt.

Adjektiv 4. 'voll'

Der Distraktor 'stolz' wurde sowohl im Gebärden- als auch im Lautsprachteil auffallend häufig gewählt (77,4% und 61,4% aller Fehler). Dies wurde auf eine Ähnlichkeit der Gebärde für 'voll' mit der Andeutung einer Stufe oder eines Podests zurückgeführt, wie sie in der Darstellung von 'stolz' zu sehen ist. Diese mögliche Verwechslung wird durch die neue Darstellung der stolzen Person ohne Siegerpodest aber mit erhobener Siegesfaust vermieden.

Adjektiv 6. 'spitz'

Die neue Darstellung des Items 'spitz' erfolgt durch zwei parallel und gerade liegende Stifte. Somit kann im Schriftsprachteil in Testheft II, in dem nur diese eine Darstellung zu sehen ist, eine Verwechslung mit dem Distraktor 'schief' (48,6% aller Fehler) vermieden werden, der mit der bisher vorliegenden Darstellung ebenfalls als richtige Antwort gelten müsste.

Adjektiv 9. 'orange'

Die Farbnuance des Items ‘orange’ musste genauer abgestimmt werden, da der Farbton im Schriftsprachteil in Testheft II bei alleiniger Abbildung dieser Farbe auch als ‘rot’ (69,8% aller Fehler) identifiziert werden konnte.

3.5.3 Modifizierter Itemkorpus

Nach Aufgabenselektion und -revision ergibt sich ein neuer Aufgabenkorpus von je 18 Nomen, Verben und Adjektiven:

Tab. 33a-B 3.5.3: Itemkorpus Nomen Version II					
neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	1	Mutter	Mädchen	Bauer	Motte
2	4	Kopf	Zopf	Junge	Topf
3	5	Straße	Weg	Strand	Strauß
4	6	Bett	Decke	Brett	Sofa
5	8	Angst	Ärger	Trauer	Freude
6	9	Garten	Baum	Geschenk	Arbeit
7	10	Stein	Felsen	Kaffee	Stange
8	11	Schuh	Stiefel	Strumpf	Uhr
9	12	Bart	Haare	Kinn	Post
10	13	Zahn	Mund	Hahn	Zange
11	14	Leute	Paar	Mensch	Herde
12	15	Ei	Huhn	Kartoffel	Eis
13	16	Zahl	Regen	Buchstabe	Pfahl
14	17	Geld	Gold	Salz	Feld
15	19	Holz	Baum	Brot	Gold
16	20	Arbeit	Eimer	Stein	Ampel
17	23	Rose	Sonnenblume	Tulpe	Rasen
18	24	Welt	Planet	Ball	Vase

Tab. 33b-B 3.5.3: Itemkorpus Verben Version II					
neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	2	gehen	stehen	rennen	sehen
2	5	fliegen	springen	winken	fragen
3	6	gucken	schminken	denken	ducken
4	7	springen	sitzen	klettern	schimpfen
5	8	schlafen	schlagen	streicheln	liegen
6	9	denken	träumen	warnen	schenken
7	10	hinsetzen	sitzen	hinlegen	schwitzen
8	12	warten	sitzen	wandern	braten
9	13	verstecken	schlafen	erschrecken	stecken
10	14	weinen	reiben	schwitzen	wiegen
11	15	werfen	fangen	strecken	wärmen
12	16	schreiben	malen	reiben	schlafen
13	17	lesen	sitzen	halten	blasen
14	18	singen	sagen	warnen	flöten
15	19	waschen	wischen	schwimmen	warten
16	20	brennen	löschen	rechnen	rennen
17	21	schieben	schließen	heben	ziehen
18	23	stinken	riechen	staunen	winken

Tab. 33c-B 3.5.3: Itemkorpus Adjektive Version II					
neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	2	klein	klug	dünn	fein
2	4	voll	viel	groß	stolz
3	5	blond	braun	rot	schwarz
4	6	spitz	struppig	schief	spannend
5	7	nass	trocken	blass	nackt
6	8	hellblau	dunkelblau	hellgrau	hellgrün
7	9	orange	rosa	gelb	rot
8	10	dünn	eng	dick	dumm
9	11	schwer	stark	leer	schwarz
10	13	verletzt	scharf	gesund	verliebt
11	14	warm	heiß	kalt	weich
12	15	schief	steil	gerade	stark

neue Nummer	alte Nummer	Item	Distraktor I	Distraktor II	Distraktor III
13	16	glatt	sauber	lockig	platt
14	19	schmutzig	sauber	zerknittert	zerrissen
15	20	weich	nass	wach	warm
16	22	gestreift	kariert	geblümt	gepunktet
17	23	viereckig	dreieckig	fünfeckig	sechseckig
18	24	nah	geschlossen	fern	nass

Diese Itemauswahl stellt den neuen Itemkorpus für die Testversion II dar. Hinzu kommen zwei Übungsbeispiele aus dem Bereich der Nomen:

Hund mit den Distraktoren Katze (I), Pferd (I/G) und Hahn (I/M/S)

Auge mit den Distraktoren Nase (I/G), Auto (M/S) und Stirn (HRP/I/G)

In der Regel wird ein neuer Itemkorpus an einer neu erhobenen Stichprobe überprüft. Da die Modifikation von Testversion I zu II des Wortschatztests jedoch anhand der Ergebnisse einer im Bezug auf die relativ kleine Grundgesamtheit besonders umfangreichen Stichprobe vorgenommen wurde, wird im vorliegenden Fall auf die Erhebung einer Kontrollstichprobe verzichtet. Alternativ wurde für die ursprüngliche Analysestichprobe eine neue Punktbewertung der Version I ohne die eliminierten Aufgaben vorgenommen und eine Aufgaben-Re-Analyse (vgl. Lienert/Raatz 1994, 126) durchgeführt. (Aufgabenschwierigkeit und -trennschärfe nach einer zufälligen Halbierung dieser Re-Analysen-Stichprobe weisen die Testendform als hinreichend stabil aus, werden jedoch neben den Werten für den Gesamttest (s. B 3.5.6) nicht gesondert aufgeführt. Die Testergebnisse der Testpersonen aus dieser Testversion I-6 (Tab. 34-B 3.5.3) sind im Anhang (Anh. 1: B 3.5.3) dargestellt.) Auch die Überprüfung der Reliabilität (B 3.7.2) und die Normierung (B 3.8) erfolgen demgemäß anhand der Daten der in B 3.1 beschriebenen Analysestichprobe, für deren Bewertung die eliminierten Aufgaben ausgenommen werden.

3.5.4 Modifizierte Reihenfolge

Die ursprüngliche Reihenfolge der Items in der Pilotversion und in Version I war ihrem Vorkommen in der Wortliste, also ihrer Häufigkeit im Gebrauch angepasst. Dabei wurde davon ausgegangen, dass die Items mit abnehmender Verwendungshäufigkeit zunehmend schwierig werden. Während der Untersuchung haben sich Abweichungen der Schwierigkeitsgrade einzelner Aufgaben von ihrer ursprünglichen Schwierigkeitsplatzierung gemäß der Wortliste herausgestellt. Um das zweckmäßige Testprinzip der ansteigenden Schwierigkeitsreihung innerhalb der Wortarten aufrechtzuerhalten, wurde die vorläufig angenommene Reihung revidiert, indem die Aufgaben entsprechend ihrer neuen Schwierigkeitsrangplätze sortiert wurden.

den. Durch eine derartige Aufgabenreihung soll eine möglichst normale Testpunktverteilung erzielt werden, die als Abbildung der „Fähigkeitsverteilung“ in der Analytestichprobe angesehen werden kann (vgl. Lienert/Raatz 1994, 140).

Es ergeben sich dabei potentiell unterschiedliche Platzierungen eines Items in allen drei Subtests, wenn die Aufgabe von den Testpersonen bei verschiedener Anbietungsform als unterschiedlich schwierig empfunden wurde. So wurde beispielsweise das Item ‘Schuh’ von den Testpersonen in der Modalität Gebärdensprache als durchschnittlich 14.-leichtestes, in der Modalität Schriftsprache als leichtestes und in der Modalität Lautsprache als siebt-leichtestes Nomen eingestuft. Drei Tabellen (Tab. 35a/b/c-B 3.5.4) im Anhang (Anh. 1: B 3.5.4a) geben die so entstehende neue Itemreihenfolge jedes Subtestteils wieder.

Aufgrund dieser neuen Verteilungen würde sich die Notwendigkeit ergeben, drei statt bisher zwei Testhefte bereitzustellen, da die Modalitäten Gebärden- und Lautsprache nicht wie bisher bei gleicher Reihenfolge der Items in einem Testheft dargestellt werden können. Ein aus drei Testheften bestehender Test wäre neben dem Aspekt einer größeren finanziellen Aufwendung unübersichtlicher und damit weniger anwenderfreundlich. Aus diesen Gründen wurde es vorgezogen, aus den Schwierigkeitswerten der beiden Modalitäten eine gemeinsame Itemreihenfolge zu erzielen, sodass der Gebärden- und der Lautsprachteil weiterhin in einem Testheft dargestellt werden können. Dafür wurde für jede Aufgabe ein Mittelwert der zwei Schwierigkeitswerte und daraus eine neue Platzierung berechnet, die den endgültigen Rangplatz der Aufgabe auf der Schwierigkeitsskala beschreibt. Eine Darstellung dieser Werte ist in der Tabelle (Tab. 36-B 3.5.4) im Anhang (Anh. 1: B 3.5.4b) einzusehen.

So ergeben sich zwei Testhefte, in denen der gleiche Itemkorpus unterschiedlich sortiert aufgeführt wird. Wenn im weiteren Verlauf dieser Arbeit auf die neue Itemreihung Bezug genommen wird, ist damit immer die Gliederung des Itemkorpus für Testheft I gemeint.

Testheft I für die Modalitäten Gebärden- und Lautsprache gliedert sich folgendermaßen:

Tab. 37a-B 3.5.4: Gliederung des Itemkorpus in Testheft I (Gebärdensprache und Lautsprache)			
Nr.	Nomen	Verben	Adjektive
1	Ei	schreiben	glatt
2	Bett	fliegen	spitz
3	Bart	stinken	hellblau
4	Zahn	springen	schmutzig
5	Angst	weinen	orange
6	Welt	lesen	viereckig
7	Stein	waschen	warm

Nr.	Nomen	Verben	Adjektive
8	Rose	werfen	dünn
9	Zahl	schieben	klein
10	Leute	verstecken	gestreift
11	Schuh	schlafen	nass
12	Arbeit	brennen	schief
13	Kopf	gehen	verletzt
14	Geld	singen	blond
15	Holz	warten	voll
16	Mutter	gucken	weich
17	Straße	hinsetzen	nah
18	Garten	denken	schwer

Fotos und Ausführungsanweisungen der Gebärden sind im Anhang (Anh. 1: B 3.5.4c) aufgeführt (Abb. 13-B 3.5.4).

Testheft II für die Modalität Schriftsprache gliedert sich folgendermaßen:

Tab. 37b-B 3.5.4: Gliederung des Itemkorpus in Testheft II (Schriftsprache)			
Nr.	Nomen	Verben	Adjektive
1	Schuh	lesen	klein
2	Ei	weinen	blond
3	Bett	fliegen	schmutzig
4	Zahn	schlafen	spitz
5	Geld	schreiben	hellblau
6	Kopf	gehen	viereckig
7	Zahl	werfen	dünn
8	Holz	springen	orange
9	Arbeit	waschen	nass
10	Rose	warten	warm
11	Straße	verstecken	voll
12	Angst	brennen	schief
13	Stein	stinken	glatt
14	Welt	singen	schwer
15	Bart	schieben	weich
16	Mutter	gucken	verletzt
17	Garten	denken	nah
18	Leute	hinsetzen	gestreift

Schließlich wurde neben der Reihenfolge der Items selber auch die Reihenfolge der Wortarten und Modalitäten auf Schwierigkeitssteigerung hin untersucht. Dafür wurde die Gesamtanzahl aller vorkommenden Fehler in der Analysestichprobe herangezogen:

Tab. 38-B 3.5.4: Fehlerverteilung				
	Gebärdenspr.	Schriftspr.	Lautspr.	ges.
Nomen	538	667	629	1834
%	27,2	33,7	33,0	33,0
Verben	305	881	681	1867
%	15,4	45,3	35,7	34,2
Adjektive	619	850	795	2264
%	31,3	43,7	41,7	41,5
ges.	1462	2398	2105	5965
%	24,6	41,1	36,8	36,5

Die prozentuale Fehleranzahl bestätigt die Annahme, dass die Wortart Nomen am leichtesten (33,0% aller Fehler) und die Wortart Adjektive am schwierigsten (41,5% aller Fehler) ausfällt. Die ursprünglich gewählte Reihenfolge Nomen – Verben – Adjektive wird demzufolge für Testversion II beibehalten.

Auch die Reihung der Modalitäten Gebärden-, Schrift- und schließlich Lautsprache wird durch die obigen Werte für die Testversion II bestätigt und daher übernommen. Wie erwartet erweist sich die Modalität Gebärdensprache für die Testpersonen als geeigneter Einstieg, da sie sich mit der geringsten Fehleranzahl (24,6% aller Fehler) als die Sprachmodalität herausstellt, zu der die Testpersonen den leichtesten Zugang haben. Wegen des potentiellen Erinnerungseffekts zwischen Gebärden- und Lautsprache mit demselben Testheft muss der Schriftsprachtest mit einem eigenständigen Heft an mittlerer Stelle positioniert werden, sodass zwischen der Anwendung des Gebärden- und des Lautsprachteils ca. vier Wochen liegen. Damit ist auch die Position der Modalität Lautsprache an dritter Stelle festgelegt, obwohl in diesem Teil weniger Fehler (36,8% aller Fehler) gemacht wurden als in der Schriftsprache (41,1% aller Fehler).

3.5.5 Modifizierte Distraktorenverteilung

Im Folgenden wird für den neuen Itemkorpus das Vorkommen der verschiedenen Distraktorenarten beschrieben. Mit (begründeter) Ausnahme von drei Adjektiven (schmutzig, gestreift und blond) konnten für jedes Item die drei Distraktoren mit Ähnlichkeit in der Gebärde (G), im Mundbild (M) bzw. Schriftbild (S) und im Inhalt (I) bereitgestellt werden. Bei 15 Nomen, zwölf Verben und sieben Adjektiven weist zudem einer der inhaltsähnlichen Distraktoren so

starke Wahlplausibilität auf, dass er als annähernd korrekte Antwort angesehen werden muss und damit untersucht werden kann, ob seine Wahl Hinweise auf elaborierte Sprachkenntnis gibt. Alle diese Distraktoren mit besonderer inhaltlicher Ähnlichkeit und höchster relativer Plausibilität (HRP) befinden sich demnach automatisch ebenfalls in der Liste der einfach inhaltlich ähnlichen Distraktoren (I). Einige sind – abhängig von der Form der Vorgabe des Items – nur in einzelnen Modalitäten mit besonderer Ähnlichkeit behaftet und demgemäß in der Liste gekennzeichnet. Schließlich weisen zehn Nomen, 17 Verben und 13 Adjektive in der Modalität Schriftsprache Distraktoren mit Schriftbildähnlichkeit zum Item auf. Diese besondere Form der Ähnlichkeit überschneidet sich mit Ausnahme von zwei Fällen mit den mundbildähnlichen Distraktoren, wobei nicht alle mundbildähnlichen Distraktoren ebenfalls schriftbildähnlicher Art sind.

Tab. 39a-B 3.5.5: Distraktorenverteilung Nomen Version II							
Nr.	Item	Distraktor	HRP	G	M	I	S
1	Ei	Huhn	x (GS, LS)			x	
		Kartoffel		x		x	
		Eis			x		x
2	Bett	Decke	x	x		x	
		Brett			x		x
		Sofa				x	
3	Bart	Haare	x			x	
		Kinn	x	x		x	
		Post		x	x		
4	Zahn	Mund	x (GS, LS)	x		x	
		Hahn			x		x
		Zange			x	x	x
5	Angst	Ärger		x	x	x	
		Trauer		x		x	
		Freude				x	
6	Welt	Planet	x	x		x	
		Ball		x			
		Vase		x	x		
7	Stein	Felsen	x (GS, LS)			x	
		Kaffee		x			
		Stange			x		x
8	Rose	Sonnenblume		x		x	
		Tulpe		x		x	
		Rasen			x	x	x
9	Zahl	Regen		x			
		Buchstabe				x	
		Pfahl			x		x
10	Leute	Paar	x (GS, LS)			x	
		Mensch		x		x	
		Herde			x	x	
11	Schuh	Stiefel	x (GS, LS)	x		x	
		Strumpf		x		x	
		Uhr			x		

Nr.	Item	Distraktor	HRP	G	M	I	S
12	Arbeit	Eimer	x (GS, LS)		x	x	
		Stein		x			
		Ampel			x		
13	Kopf	Zopf	x (GS, LS)		x	x	x
		Junge		x			
		Topf			x		x
14	Geld	Gold	x		x	x	x
		Salz		x			
		Feld			x		x
15	Holz	Baum	x			x	
		Brot		x	x		
		Gold			x		
16	Mutter	Mädchen	x			x	
		Bauer		x			
		Motte			x		x
17	Straße	Weg	x	x		x	
		Strand			x		x
		Strauß			x		x
18	Garten	Baum	x			x	
		Geschenk		x			
		Arbeit			x		

Tab. 39b-B 3.5.5: Distraktorenverteilung Verben Version II							
Nr.	Item	Distraktor	HRP	G	M	I	S
1	schreiben	malen		x		x	
		reiben		x	x		x
		schlafen			x		x
2	fliegen	springen	x			x	
		winken		x			
		fragen			x		x
3	stinken	riechen	x	x		x	
		staunen			x		x
		winken			x		x
4	springen	sitzen		x			
		klettern				x	
		schimpfen			x		
5	weinen	reiben	x (GS, LS)	x		x	x
		schwitzen				x	
		wiegen			x		x
6	lesen	sitzen	x (GS, LS)			x	
		halten	x (SS)	x		x (SS)	
		blasen			x		x
7	waschen	wischen	x	x	x	x	x
		schwimmen		x			
		warten			x		x
8	werfen	fangen		x		x	
		strecken				x	
		wärmen			x		x

Nr.	Item	Distraktor	HRP	G	M	I	S
9	schieben	schließen	x (GS, LS)	x	x	x	x
		heben				x	
		ziehen				x	
10	verstecken	schlafen		x			
		erschrecken			x	x	x
		stecken			x		x
11	schlafen	schlagen		x	x		x
		streicheln		x			
		liegen	x (SS)			x	
12	brennen	löschen	x			x	
		rechnen		x	x		x
		rennen			x		x
13	gehen	stehen		x	x	x	x
		rennen				x	
		sehen			x		x
14	singen	sagen		x	x	x	x
		warnen		x			
		flöten				x	
15	warten	sitzen	x (GS, LS)			x	
		wandern		x	x		x
		braten			x		x
16	gucken	schminken	x (GS, LS)	x		x	
		denken		x			
		ducken			x		x
17	hinsetzen	sitzen	x	x	x	x	x
		hinlegen			x	x	x
		schwitzen			x		
18	denken	träumen	x	x		x	
		warnen		x			
		schenken			x		x

Tab. 39c-B 3.5.5: Distraktorenverteilung Adjektive Version II

Nr.	Item	Distraktor	HRP	G	M	I	S
1	glatt	sauber		x			
		lockig				x	
		platt			x		x
2	spitz	struppig		x			
		schief	x (SS)			x	x
		spannend			x		
3	hellblau	dunkelblau		x	x	x	x
		hellgrau		x	x	x	x
		hellgrün		x	x	x	x
4	schmutzig	sauber				x	
		zerknittert				x	
		zerrissen				x	
5	orange	rosa		x	x	x	
		gelb				x	
		rot				x	

Nr.	Item	Distraktor	HRP	G	M	I	S
6	viereckig	dreieckig		x	x	x	x
		fünfeckig		x	x	x	x
		sechseckig		x	x	x	x
7	warm	heiß	x	x		x	
		kalt				x	
		weich			x		x
8	dünn	eng	x (GS, LS)	x	x	x	
		dick				x	
		dumm			x		x
9	klein	klug		x	x		
		dünn				x	
		fein			x		x
10	gestreift	kariert		x		x	
		geblümt				x	
		gepunktet				x	
11	nass	trocken		x		x	
		blass			x		x
		nackt			x		x
12	schief	steil	x (GS, LS)	x		x	
		gerade				x	
		stark			x		
13	verletzt	scharf	x (GS, LS)	x		x	
		gesund				x	
		verliebt			x		x
14	blond	braun			x	x	
		rot				x	
		schwarz				x	
15	voll	viel	x	x	x	x	x
		groß		x	x	x	
		stolz		x	x		
16	weich	nass		x			
		wach			x	x	x
		warm			x		
17	nah	geschlossen		x			
		fern				x	
		nass			x		x
18	schwer	stark	x (GS, LS)	x		x	
		leer			x		
		schwarz			x		x

3.5.6 Neue Trennschärfewerte

Im Rahmen der Aufgaben-Re-Analyse wurde für den neuen Itemkorpus die Trennschärfe der verbleibenden Aufgaben neu berechnet. Sowohl die tabellarische Darstellung (Tab. 40a/b/c-B 3.5.6) als auch die anschauliche Darstellung der neuen Werte im Koordinatensystem (Abb. 13a-i-B 3.5.6) sind im Anhang (Anh. 1: B 3.5.6a und Anh. 1: B 3.5.6b) aufgeführt.

Ein qualitativer Vergleich dieser Verteilungsformen mit denen aus Version I (vgl. B 3.4.2) macht deutlich, dass neben der notwendigen Veränderung, die Werte unterhalb der x-Achse zu entfernen, für Testversion II die Annäherung an eine parabolische Darstellung besser realisiert werden konnte. Auch der quantitative Vergleich der neuen Werte mit denen aus Version I ergibt durch Eliminierung der ungeeigneten Items eine durchgehende Verbesserung der Trennschärfekoeffizienten in allen Subtestteilen:

Tab. 41-B 3.5.6: Mittlere Trennschärfe im Vergleich Version I und II									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
mittlere Trennsch. I	.18	.41	.37	.24	.44	.34	.21	.46	.39
mittlere Trennsch. II	.25	.43	.38	.27	.47	.35	.25	.46	.43

Darüber hinaus wird deutlich, dass vor allem in den Subtestteilen, in denen Items aufgrund negativer Trennschärfe eliminiert wurden (GSN, GSV, GSA und LSA), mehr Items eine Trennschärfesteigerung als eine -minderung erfahren haben. Aufgrund dieser überwiegend positiven Veränderungen der Aufgabenindizes und der Annäherung der dargestellten Zusammenhänge an eine parabolische Form war eine Sekundärselektion (vgl. Lienert/Raatz 1994, 126) der Aufgaben nicht erforderlich.

3.5.7 Modifiziertes Testprotokoll

Nach Abschluss der Aufgabenanalyse und Festlegung eines endgültigen Itemkorpus für Testversion II wurde schließlich der Protokollbogen erneuert. Sein Aufbau (Tab. 42a-j-B 3.5.7) mit einem Deckblatt sowie neun Subtestteil-Protokollen wird im Anhang (Anh. 1: B 3.5.7) dargestellt.

Auf dem Deckblatt werden Name, Testdatum und –uhrzeit des Subtests, der Geburtstag bzw. das Alter (auf den Monat genau; gerundet) sowie die Klassenstufe eingetragen. Auf jedem folgenden Protokollbogen befinden sich die jeweils 18 Items eines Subtestteils mit den dazugehörigen Distraktoren, in der Wortart Nomen jeweils inklusive der beiden Übungsbeispiele. Neben der Spalte mit den Items befinden sich in drei Spalten die Distraktoren zu jedem Item. Jeweils hinter den Items bzw. Distraktoren ist ein freies Feld, das bei entsprechender Antwort angekreuzt wird. Zum Zeitpunkt der Testdurchführung sind nur diese schwarz- und fettgedruckten Angaben von Bedeutung.

Die Summe der richtigen Antworten wird nach dem Testdurchgang in einem grau umrandeten Feld unterhalb der Itemspalte eingetragen. Anhand dieser Angabe lassen sich aus den Normierungstabellen (s. B 3.8.4) Prozentränge für die Leistung der Testperson ablesen.

Unter jedem Distraktor ist in grau dessen Art der Ähnlichkeit zum Item angegeben. Durch die Konzeption der Distraktoren kann eine falsche Antwort mehrere Distraktorenähnlichkeiten gleichzeitig bedeuten. Im Falle der Wahl eines Distraktors kann der Anwender im Anschluss an den Testdurchgang in vier (Subtest Gebärden- und Lautsprache) bzw. fünf (Subtest Schriftsprache) zusätzlichen grauen Spalten auf der rechten Seite des Testprotokolls die Art der Distraktoren ankreuzen. Das gibt ihm die Möglichkeit, die Anzahl der gewählten Distraktoren entsprechend ihrer Ähnlichkeiten zum Item getrennt auszuzählen und die Summe am Ende der Spalten einzutragen. Diese Information wird für das Ablesen von Werten aus den Tabellen zur Distraktoren-Normierung (s. B 3.10.2) benötigt.

Die in dem alten Testprotokoll aufgeführten Kategorien 'weiß nicht', 'Kind wiederholt' (differenziert in Gebärde, Lautsprache, Mundbild) sowie 'Testleiter wiederholt' sind bei der Modifikation entfallen. Sie dienten der Informationsgewinnung im Zuge der Testentwicklung und sind für zukünftige Anwender ohne Bedeutung. Demnach muss sich eine Testperson in jedem Fall für eine der vier Möglichkeiten entscheiden, auch wenn sie sich nicht sicher ist oder keine Antwort weiß. Die Wahrscheinlichkeit dieser Zufallstreffer wurde in den entsprechenden statistischen Berechnungen berücksichtigt.

3.6 Exkurs: Ikonizität von Gebärden

Prillwitz / Wudtke weisen auf die häufige "Ikonizität von Gebärden" (1988, 64) hin. Gemeint ist der Umstand, dass bestimmte Gebärden das durch sie Bezeichnete motorisch nachbilden. Auf diese Weise ist dessen Gestalt oder Sinn auch ohne Kenntnis der Bedeutung der betreffenden Gebärde offensichtlich oder kann leicht erschlossen werden, bzw. besteht umgekehrt die Möglichkeit, eine Gebärde zufällig zutreffend auszuführen. Dieser Hinweis zielt auf die dadurch bestehende gewisse Wahrscheinlichkeit für die gehörlose Testperson ab, eine dem abgebildeten Objekt ähnelnde Gebärde zu dessen Bezeichnung zu erfinden. Dadurch würde das angestrebte Ziel, die lexikalisch-semantische Kompetenz zu erfassen, von dem Kind unterlaufen werden. So merken Prillwitz / Wudtke an, dass die Gebärden auf Grund ihrer Ikonizität gerade bei dem Wortschatz von Kindern zu viel Hilfen gewähren.

Dieser Kritikpunkt fällt so nur ins Gewicht, sofern es sich um einen aktiven Wortschatztest handelt. Hier erhält die Testperson die Aufgabe, Gegenstände oder Abbildungen, die ihm

vorgelegt werden, von sich aus (aktiv) mit dem richtigen Wort bzw. der richtigen Gebärde zu benennen. Verarbeitet das Kind nun die bildlichen Momente des Gegenstandes oder der Abbildung zu einer spontan eigens kreierte Gebärde, so könnte mit gewisser Wahrscheinlichkeit die zutreffende Gebärde von der Testperson ausgeführt werden, ohne dass sie dabei über die Kenntnis des betreffenden Wortes verfügt. Im vorliegenden Fall der Erstellung eines passiven Wortschatztests stellt sich die Situation verändert dar. Dennoch führt benannte Kritik zu einem angrenzenden Problem, dem im Folgenden nachgegangen wird.

Prillwitz / Wudtke vermuten, dass das Kind von den bildlichen Momenten der von dem Testleiter angebotenen Gebärden auf die richtige der vier vorgelegten Abbildungen schließen könnte. Auch dies hätte zur Folge, nicht mehr die lexikalisch-semantische Kompetenz des Kindes zu testen, sondern dessen Fähigkeit, Gebärde und Bild vergleichend aufeinander zu beziehen. Diese Schwierigkeit trifft ausschließlich im Gebärdensprachtest zu, in dem der Testleiter der Testperson das gebärdete Item anbietet und das Kind daraufhin aus den vier zugehörigen Zeichnungen auswählt. (Im Schriftsprachtest muss ein geschriebenes Wort ausgewählt werden, im Lautsprachtest wird lediglich das gesprochene Wort angeboten. In beiden Fällen kann sich die Testperson die bildlichen Momente einer Gebärde nicht zu Nutze machen.)

Maßgeblich entlastend ist im vorliegenden Fall jedoch vor allem der Umstand, dass, wie oben geschildert, jedem Item das Konzept zugrundeliegt, die drei Distraktoren derart zuzuordnen, dass jeweils mindestens einmal eine Ähnlichkeit zum Item in der Gebärde, im Mundbild und in inhaltlicher Hinsicht besteht. (Diese Systematik war in der Erstversion noch nicht lückenlos eingehalten, später jedoch vervollständigt worden.) Gesetzt den Fall, es trifft bei einem Item Ikonizität der angebotenen Gebärde zu, so kann die Testperson durch dieses Konzept von den bildlichen Momenten der Gebärde nicht zielsicher auf das Item schließen, sondern könnte ebenso gut auf den gebärden- oder den inhaltsähnlichen Distraktor verfallen. Dabei liegt die Ablenkungsfunktion des Distraktors mit Parallelen in der Gebärde auf der Hand. Jedoch haben auch inhaltlich übereinstimmende Elemente ablenkende Funktion im Sinne des dargestellten Problems. Zur Verdeutlichung soll das erste betroffene Nomen 'Kopf' mit den Distraktoren 'Zopf' (Ähnlichkeit in Mundbild und Inhalt), 'Junge' (Ähnlichkeit in der Gebärde) und 'Topf' (Ähnlichkeit im Mundbild) dienen: Versucht eine Testperson, von der Gebärde für 'Kopf', die mit dem Tippen der flachen Hand an den Kopf oberhalb der Schläfe ausgeführt wird, auf eine der Zeichnungen zu schließen, so bietet sich neben dem Abgelenktsein durch die Gebärde für 'Junge', bei der sich lediglich die Stellung der Hand im Raum, nicht aber ihre Form, Bewegung und Position unterscheiden, auch ein Abgelenktsein durch die inhaltlichen Übereinstimmungen zum Distraktor 'Zopf' an. Denn die Zeigefunktion der Hand auf das Haupthaar stimmt hier überein.

Das Ziel, dem von Prillwitz / Wudtke aufgezeigten Problem in dieser Weise entgegenzuwirken, stellt einen der Gründe dar, aus denen in der Phase der Überarbeitung des Tests zur endgültigen Testversion die Zuordnung der Distraktoren nochmals dahingehend modifiziert wurde, in jedem Fall das Vorkommen der drei Ähnlichkeiten zu sichern. Somit wird dem von Prillwitz / Wudtke formulierten Anspruch nachgekommen, standardisierte Tests für Gehörlose speziell unter dem Gesichtspunkt der möglichen Hilfen, die die Ikonizität bestimmter Gebärden gewähren, zu konzipieren.

Um die Auswirkung der Problematik auf die Erhebung des Wortschatzes im Gebärdensprachtest hier jedoch konsequent aufzugreifen und die durchgängige Einhaltung der Maßnahmen, die zu seiner Aufhebung führen, zu verdeutlichen, wurde von den Autorinnen eine entsprechende Überprüfung durchgeführt. Diese wird in den folgenden drei Tabellen für die jeweiligen Wortarten dargestellt. Dabei wird der Itemkorpus der endgültigen Testversion II untersucht. Die Tabellen listen den gesamten Itemkorpus auf, wobei dahinter die Angabe folgt, ob Ikonizität vorliegt oder nicht. Bei den betreffenden Items wird nachfolgend angegeben, auf welche Distraktoren bildhafte bzw. inhaltliche Parallelen zum Item zutreffen. Damit wäre der Nachweis erbracht, dass jeweils zwei zusätzliche im Sinne der Ikonizität des Items plausible Wahlmöglichkeiten bestehen.

Tab. 43a-B 3.6: Ikonizität der Nomen				
Nr.	Item	Ikonizität	Bildhafte Parallelen	Inhaltliche Parallelen
01.	Ei	x	Kartoffel	Huhn/Kartoffel
02.	Bett			
03.	Bart	x	Kinn/Post	Haare/Kinn
04.	Zahn	x	Mund	Mund/Zange
05.	Angst			
06.	Welt	x	Planet/Ball/Vase	Planet/Ball
07.	Stein			
08.	Rose	x	Sonnenblume/Tulpe	Sonnenbl./Tulpe/Rasen
09.	Zahl			
10.	Leute			
11.	Schuh	x	Stiefel/Strumpf	Stiefel/Strumpf
12.	Arbeit			
13.	Kopf	x	Junge	Zopf
14.	Geld	x	Salz	Gold
15.	Holz			
16.	Mutter			
17.	Straße	x	Weg	Weg
18.	Garten			

Tab. 43b-B 3.6: Ikonizität der Verben

Nr.	Item	Ikonizität	Bildhafte Parallelen	Inhaltliche Parallelen
01.	schreiben	x	malen	malen
02.	fliegen	x	winken	springen
03.	stinken	x	riechen	riechen
04.	springen	x	sitzen	klettern
05.	weinen	x	reiben	reiben/schwitzen
06.	lesen	x	halten	sitzen
07.	waschen	x	wischen	wischen/schwimmen
08.	werfen	x	fangen	fangen/strecken
09.	schieben	x	schließen	schließen/heben/ziehen
10.	verstecken	x	schlafen	erschrecken
11.	schlafen	x	schlagen/streicheln	liegen
12.	brennen	x	rechnen	löschen
13.	gehen	x	stehen	stehen/rennen
14.	singen	x	sagen/warnen	sagen/flöten
15.	warten			
16.	gucken	x	schminken/denken	schminken
17.	hinsetzen	x	sitzen	sitzen/hinlegen
18.	denken	x	träumen/warnen	träumen

Tab. 43c-B 3.6: Ikonizität der Adjektive

Nr.	Item	Ikonizität	Bildhafte Parallelen	Inhaltliche Parallelen
01.	glatt	x	sauber	lockig
02.	spitz	x	struppig	schief
03.	hellblau			
04.	schmutzig			
05.	orange	x	rosa	rosa/gelb/rot
06.	viereckig	x	drei-/fünf-/sechseckig	drei-/fünf-/sechseckig
07.	warm			
08.	dünn	x	eng	eng/dick
09.	klein	x	klug	dünn
10.	gestreift	x	kariert	kariert/geblümt/gepun.
11.	nass			
12.	schief	x	steil	steil/gerade
13.	verletzt	x	scharf	scharf/gesund
14.	blond	x		braun/rot/schwarz
15.	voll	x	viel/groß/stolz	viel/groß
16.	weich	x	nass	wach
17.	nah	x	geschlossen	fern
18.	schwer	x	stark	stark

Zunächst ist abzulesen, dass in 40 Fällen von insgesamt 54 Items aller Wortarten Ikonizität der Gebärde vorliegt, was einem Anteil von 74% entspricht. D. h., dass diese 40 Gebärden, die der Testperson angeboten werden, derart bildhaften Charakter haben, dass ein Rückschluss auf die entsprechende Zeichnung möglich erscheint. Betroffen sind neun Nomen, 17 Verben und 14 Adjektive. In jedem dieser Fälle (mit einer speziellen Ausnahme) ist dem Item mindestens ein Distraktor mit Ähnlichkeit zum Item in der Gebärde zugeordnet. Zusätzlich ist jeweils mindestens ein Distraktor mit inhaltlicher Ähnlichkeit zum Item vorhanden. In einigen Fällen übernehmen Distraktoren beide Ähnlichkeitsfunktionen. In jedem Fall ist ausgeschlossen, dass das Kind eindeutige Rückschlüsse ziehen kann, da unter den angebotenen Distraktoren mindestens einer, häufig sogar zwei und in einigen günstigen Fällen alle drei bildhafte bzw. inhaltliche Parallelen mit der Gebärde aufweisen. Bei den bildhaften Parallelen zu einem der 40 Items ist in 29 Fällen ein Distraktor vorhanden, in sieben Fällen sind zwei, in drei Fällen drei und in einem Fall ist kein Distraktor vorhanden. Bei den inhaltlichen Parallelen ist in 18 Fällen ein Distraktor, in 16 Fällen sind zwei und in sechs Fällen drei Distraktoren vorhanden.

Zwei spezielle Fälle sollen kurz Erwähnung finden. Für das Verb Nr. 10 'verstecken' ist als inhaltsähnlicher Distraktor 'erschrecken' angegeben. Die inhaltliche Parallelität besteht hier nicht über den Sinn der Wörter, sondern ergibt sich aus den zeichnerischen Darstellungen im Testheft. Die Zeichnung des Wortes 'verstecken' stellt eine hinter einen Pfeiler geduckte Person dar. Die Darstellung des Wortes 'erschrecken' unterscheidet sich davon nur unwesentlich, nämlich im Gesichtsausdruck und in der Haltung des Oberkörpers der Person. Der Bildaufbau bleibt grundsätzlich erhalten, sodass sich beide Bilder stark ähneln und insofern inhaltliche Parallelität besteht. Die Gebärde für das Item Nr. 14 der Adjektive 'blond' wird mit einer bestimmten Bewegung der Hand am Haar ausgeführt. Demnach trifft hier, wenn strenge Maßstäbe zu Grunde gelegt werden, Ikonizität der Gebärde zu, ohne dass ein gebärdenähnlicher Distraktor vorhanden ist. Denn die Distraktoren 'braun', 'rot' und 'schwarz' sind alle inhaltsähnlich und 'braun' weist zusätzlich Parallelen im Mundbild auf. Da jedoch die inhaltlichen Parallelen auf alle drei Distraktoren zutreffen und es so für die Testperson unmöglich ist, das Item auf Grund der Ikonizität der Gebärde zu wählen, muss der Umstand, dass keine Parallelen in der Gebärde bestehen, als unerheblich bewertet werden.

Abschließend ist festzustellen, dass auf Grund des hohen Anteils von 74% der Items, bei denen Ikonizität der Gebärden zutrifft, das von Prillwitz / Wudtke aufgezeigte Problem deutlich auf den vorliegenden Test zutrifft. Seine Bearbeitung war daher unbedingt angezeigt. Die entsprechenden Untersuchungen erbringen den Nachweis, dass die kritisch angemerkten Aspekte durch die konzeptionelle Grundlage des Tests in den drei Sprachmodalitäten und die konsequente Modifizierung der Distraktorenauswahl behoben wurden. Das Testinstrument entspricht demnach auch in dieser Hinsicht dem Anspruch, die in bestehenden Testinstru-

menten vorfindlichen Mängel nicht fortzusetzen, sondern grundlegend zu vermeiden, indem die Itemauswahl entsprechend getroffen wurde, und zu verbessern, indem die Distraktorauswahl entsprechend überarbeitet wurde.

Gesondert betrachtet werden sollen unter dem Aspekt der Ikonizität der Gebärden zusätzlich zu den ausgeführten Untersuchungen diejenigen Fälle, in denen Items ausschließlich in der Modalität Gebärdensprache richtig beantwortet wurden. Dies könnte bei konsequenter Verfolgung der möglichen Folgen der Ikonizität, wie sie Prillwitz / Wudtke anmerken, als Hinweis darauf bewertet werden, dass tatsächlich die Ikonizität die Testperson zur korrekten Antwort des Items verholfen hat, obgleich diese eigentlich keinen Begriff von dem Wort haben. Begründet wäre diese Annahme damit, dass die betreffenden Testpersonen das Item andernfalls mit einiger Wahrscheinlichkeit auch in einer der anderen Modalitäten beherrscht hätten. Die folgenden Tabellen stellen diesen Aspekt dar. Zu den oben bereits aufgeführten Informationen ist hier die Anzahl derjenigen Testpersonen angeben, die das betreffende Item ausschließlich in der Modalität Gebärdensprache richtig, in den Modalitäten Schrift- und Lautsprache jedoch falsch beantwortet haben. Hinter der Anzahl befindet sich die Angabe jeweils in Prozent. Bezugsgröße ist die Summe der Testpersonen, die die jeweilige Wortart in allen Modalitäten beantwortet haben. Dies sind bei den Nomen 103, bei Verben und Adjektiven je 101 Testpersonen.

Tab. 44a-B 3.6: Ausschließlich in Gebärdensprache richtig beantwortete				
Nomen				
Nr.	Item	Ikonizität	nur in GS richtig	%
01.	Ei	x	1	1,0
02.	Bett		2	2,1
03.	Bart	x	9	9,3
04.	Zahn	x	7	7,2
05.	Angst		7	7,2
06.	Welt	x	7	7,3
07.	Stein		8	8,2
08.	Rose	x	14	14,4
09.	Zahl		24	24,7
10.	Leute		27	27,8
11.	Schuh	x	1	1,0
12.	Arbeit		7	7,2
13.	Kopf	x	13	13,4
14.	Geld	x	4	4,1
15.	Holz		5	5,2
16.	Mutter		14	14,4
17.	Straße	x	5	5,2
18.	Garten		6	6,2

Tab. 44b-B 3.6: Ausschließlich in Gebärdensprache richtig beantwortete Verben				
Nr.	Item	Ikonizität	nur in GS richtig	%
01.	schreiben	x	7	7,1
02.	fliegen	x	13	13,1
03.	stinken	x	11	11,1
04.	springen	x	10	10,1
05.	weinen	x	12	12,1
06.	lesen	x	11	11,1
07.	waschen	x	16	16,2
08.	werfen	x	18	18,2
09.	schieben	x	30	30,3
10.	verstecken	x	14	14,1
11.	schlafen	x	6	6,1
12.	brennen	x	30	30,3
13.	gehen	x	17	17,2
14.	singen	x	26	26,3
15.	warten		18	18,2
16.	gucken	x	32	32,3
17.	hinsetzen	x	15	15,2
18.	denken	x	26	26,3

Tab. 44c-B 3.6: Ausschließlich in Gebärdensprache richtig beantwortete Adjektive				
Nr.	Item	Ikonizität	nur in GS richtig	%
01.	glatt	x	11	11,1
02.	spitz	x	17	17,2
03.	hellblau		2	2,0
04.	schmutzig		8	8,1
05.	orange	x	13	13,1
06.	viereckig	x	24	24,2
07.	warm		12	12,1
08.	dünn	x	19	19,2
09.	klein	x	11	11,1
10.	gestreift	x	29	29,3
11.	nass		8	8,1
12.	schief	x	23	23,2
13.	verletzt	x	7	7,1
14.	blond	x	3	3,0
15.	voll	x	3	3,0
16.	weich	x	15	15,2
17.	nah	x	14	14,1
18.	schwer	x	3	3,0

Die Zahlen lassen erkennen, dass jedes Item, bei dem Ikonizität der Gebärde vorliegt, mehr oder weniger häufig ausschließlich in der Modalität Gebärdensprache richtig beantwortet wurde. Eine Festlegung derjenigen Werte, die auf Grund ihrer relativen Höhe etwa besondere Beachtung erfordern, soll hier nicht vorgenommen werden. Eine derartige Festlegung könnte nur willkürlich sein und würde eines objektiven Erklärungshintergrundes entbehren. Die Ergebnisse werden stattdessen im Folgenden in dreierlei Hinsicht diskutiert:

Auffällig ist das gehäufte Vorkommen der bezeichneten Fälle bei den Verben. Die Erklärung dafür liegt nach Ansicht der Autorinnen in wortartspezifischen zeichnerischen Bedingungen. Da Verben Tätigkeiten bezeichnen und diese i. d. R. in Bewegung ausgeführt werden, muss die entsprechende Zeichnung diese Bewegung abbilden. Die zugehörige Gebärde bildet die Bewegung jedoch ebenfalls ab. Folge ist eine besonders ausgeprägte Parallelität zwischen dem bildhaften Charakter der Gebärde und der zugehörigen Zeichnung. Da aber die Gründe dafür in der Wortart liegen, wäre die einzige Möglichkeit zur Behebung dieser Schwierigkeit deren Eliminierung. (Vorausgesetzt, die Annahme von Prillwitz / Wudtke ist zutreffend.) Die Abwägung dieses Arguments gegen die Gründe, die zur Aufnahme der Wortart in den Test zu gleichen Anteilen wie die übrigen beiden Wortarten geführt hat, ergibt jedoch eine eindeutige Entscheidung für die Beibehaltung von Verben im Itemkorpus.

Die Erwartung der Autorinnen ging, wie oben erwähnt, von vornherein dahin anzunehmen, dass die gehörlosen Testpersonen generell am besten in ihrer Muttersprache, der Gebärdensprache, abschneiden würden (vgl. C 1.1 und C 1.3). Dies wurde sogar unabhängig von den jeweiligen pädagogischen Konzepten, mit denen die Kinder erzogen wurden, angenommen. Zeigen die Werte nun also durchgängig das Vorkommen von Fällen auf, in denen ein Item ausschließlich in der Modalität Gebärdensprache korrekt beantwortet wurde, so ist eine Möglichkeit, dass dies die Annahme bestätigt, dass bei den Testpersonen hohe Gebärdensprachkenntnis vorliegt. Es wird dabei davon ausgegangen, dass nicht etwa die Ikonizität der jeweiligen Gebärden zur korrekten Beantwortung der Items geführt hat, sondern eine relativ große Anzahl von Items tatsächlich ausschließlich in der Modalität Gebärdensprache beherrscht werden. Das hieße, dass die Kinder durchaus einen Begriff von dem jeweiligen Item haben, in den Modalitäten Schrift- bzw. Lautsprache jedoch modalitätsspezifische Eigenarten dazu führen, dass keine korrekte Antwort erfolgt. Diese Annahme wird hier der Überlegung Prillwitz' und Wudtkes gegenübergestellt.

Träfe die Vermutung von Prillwitz / Wudtke zu, dass das Zustandekommen dieser Fälle möglicherweise ursächlich mit der Ikonizität der Gebärden zusammenhängt, so müssten folgerichtig bei den Items, deren Gebärden keinen bildhaften Charakter aufweisen, keine oder zumindest deutlich weniger Fälle auftreten, in denen Testpersonen bestimmte Items ausschließlich in der Modalität Gebärdensprache korrekt beantworten. Als Vergleichsmaßstab

werden daher die Werte der Items herangezogen, die trotz nicht bestehender Ikonizität der Gebärde ausschließlich in der Gebärdensprache richtig beantwortet wurden. Die folgende Tabelle veranschaulicht diesen Vergleich.

Tab. 45-B 3.6: Ausschließlich in Gebärdensprache richtig beantwortete Items mit und ohne Ikonizität				
	Items mit Ikonizität der Gebärde		Items ohne Ikonizität der Gebärde	
	Durchschnittswerte		Durchschnittswerte	
	Anzahl	%	Anzahl	%
Nomen	6,8	7	11,1	11,4
Verben	17,3	17,5	18	18,2
Adjektive	13,7	13,8	7,5	7,6

Bei den Items, deren Gebärden Ikonizität aufweisen, besteht bei den Nomen der durchschnittlich niedrigste Wert von 6,8 Fällen bzw. 7% aller Fälle, bei den Verben der höchste Wert von durchschnittlich 17,3 Fällen bzw. 17,5% aller Fälle und bei den Adjektiven der durchschnittliche Wert von 13,7 Fällen bzw. 13,8% aller Fälle. Wird nun unter vorangehend formulierter Fragestellung ein Vergleich der genannten Werte mit denen derjenigen Items, deren Gebärden keinen bildhaften Charakter haben, gezogen, so wird deutlich, dass zunächst bei den Nomen die Annahme Prillwitz' / Wudtkes nicht nachweisbar ist. Denn hier haben mit 11,1 Fällen bzw. 11,4% aller Fälle weit weniger Testpersonen Items mit Ikonizität der Gebärden ausschließlich in der Modalität Gebärdensprache richtig beantwortet als Items ohne Ikonizität der Gebärden mit 6,8 Fällen bzw. 7% aller Fälle. Dabei eignet sich die Wortart der Nomen besonders gut für diesen Vergleich, da sich die Vergleichswerte auf jeweils 50% der vorhandenen Nomen beziehen, d. h. bei der Hälfte der Nomen trifft Ikonizität der Gebärden zu, bei der anderen Hälfte nicht. Die Verben eignen sich für eine Untersuchung unter diesem Aspekt kaum, da nur bei einem von 18 Items Ikonizität nicht zutrifft. Die durchschnittlichen Werte der Vergleichsgruppe der übrigen 17 Verben sind mit 17,3 Fällen bzw. 17,5% aller Fälle zwar niedriger als der Wert des Verbs 'warten' mit 18 Fällen bzw. 18,2% aller Fälle, was dem Ergebnis der Untersuchung der Nomen entsprechen würde; die Untersuchung ist jedoch auf Grund des Missverhältnisses der Vergleichsgruppen nicht angemessen. Die Adjektive weisen Werte auf, welche die gegenteilige Annahme unterstützen könnten. Hier sind die Werte der Items mit Ikonizität der Gebärden mit 13,7 Fällen bzw. 13,8% aller Fälle höher als die Werte der Items ohne Ikonizität der Gebärden mit 7,5 Fällen bzw. 7,6% aller Fälle.

Es sollte deutlich geworden sein, dass die oben stehende Untersuchung als Folge- und Exkursuntersuchung der Haupterhebung der Daten zur Erstellung des Tests bewertet werden muss. Ziel war es, die Annahme von Prillwitz / Wudtke konsequent zu verfolgen. Der dazu

betriebe Aufwand sollte sich jedoch in angemessenen Grenzen halten. Daher wurde das vorhandene Datenmaterial in Hinblick auf die differierenden Annahmen ausgewertet und bearbeitet. Im Ergebnis muss festgestellt werden, dass mit den angeführten Werten keine der beiden Annahmen, weder die von Prillwitz / Wudtke noch die der Autorinnen des vorliegenden Tests, verifiziert oder falsifiziert werden kann. Folge dieses Ergebnisses, die Annahme Prillwitz' und Wudtkes mit dem vorhandenen Datenmaterial nicht beweisen zu können, ist zum einen der Beschluss, sie nicht zur Grundlage möglicher Modifikationen des Tests heranzuziehen. Zum anderen stellt der offene Ausgang der Untersuchung den Ausgangspunkt möglicher Folgeuntersuchungen unter diesem Aspekt dar. Diese werden jedoch im Rahmen der vorliegenden Arbeit nicht vorgenommen.

3.7 Testgütekriterien

3.7.1 Objektivität

Unter dem ersten Gütekriterium der klassischen Testtheorie, der Objektivität, wird Unabhängigkeit der Ergebnisse von der Durchführung, der Auswertung und der Interpretation verlangt. Die Durchführungsobjektivität ist für den Wortschatztest durch die in schriftlicher Form vorliegende Testanweisung eindeutig gewährleistet, sofern sich der Testleiter an diese Anweisung hält oder sie zumindest nicht inhaltlich verändert. Im Subtest Gebärdensprache werden dem Testleiter die zu verwendenden Gebärden als Gebärdenphotos mit Ausführungsanweisung vorgegeben. Die Durchführungsobjektivität wird demnach zusätzlich durch die Art und Weise bestimmt, in der der Testleiter im Stande ist, diese Vorgaben in für die Testperson verständliche Gebärden umzusetzen. Die Auswertungsobjektivität ist dadurch gewährleistet, dass eindeutig festgelegt ist, welche der Antwortmöglichkeiten richtig ist. Die Interpretationsobjektivität spielt auf dieser Stufe der Testkonstruktion, auf der für erzielte Rohwerte Prozentrangangaben in Hinblick auf unterschiedliche Bezugsgruppen gemacht werden, noch keine entscheidende Rolle. Die Umwandlung von Rohwerten in Prozentränge ist nicht durch Interpretationsspielräume seitens des Testleiters beeinflussbar und somit objektiv. Lienert und Raatz sprechen in diesem Zusammenhang von der Trivialität der Objektivität (vgl. 1994, 8).

3.7.2 Reliabilität

Die Reliabilität eines Tests ist gewährleistet, wenn mit ihm ein reproduzierbares Testergebnis gewonnen werden kann, das die Leistung einer Testperson genau auf der Testskala einordnet.

Bevor mit der Überprüfung der Reliabilität des Wortschatztests begonnen werden konnte, wurden die in diesem Zusammenhang üblicherweise verwendeten statistischen Berechnungsmethoden auf ihre Eignung für dieses Testinstrument untersucht. Dabei handelt es sich zunächst um die Beurteilung der Zweckmäßigkeit des Einsatzes verschiedener Verfahren. Die vier zur Verfügung stehenden Verfahren lassen sich grundsätzlich in zwei Kategorien zusammenfassen. Die Methode der Inneren Konsistenz stellt eine Verallgemeinerung der Testhalbierung dar, denn sie geht davon aus, dass sich ein Testinstrument nicht nur in zwei, sondern auch in mehrere bzw. im äußersten Fall in so viele äquivalente Teile unterteilen lässt, wie Aufgaben vorhanden sind. Zudem ist der Split-half-Methode und der Inneren Konsistenz gemein, dass sie aufgrund einmaliger Erhebung bestimmte Aspekte der Fehlervarianz vermeiden, die auf Retest- und Paralleltest-Verfahren Einfluss nehmen. Während die Methoden der einmaligen Untersuchung lediglich durch die Auswertungsobjektivität, die Inhomogenität der Items sowie punktuelle Schwankungen der Untersuchungsbedingungen (z. B. Störungen) und/oder des Merkmals beeinflusst werden, sind Paralleltest- und Retest-Verfahren zusätzlich von der Veränderlichkeit der Testdurchführungsbedingungen bestimmt. Demzufolge geben die Innere Konsistenz und die Split-half-Methode Aufschluss über die Leistungsfähigkeit des Tests und kommen dem eigentlichen Wesen der Reliabilität, nämlich ihrer instrumentellen Seite, am nächsten (vgl. Lienert/Raatz 1994, 201). Daher werden diese beiden Methoden zur Überprüfung der Reliabilität des Wortschatztests eingesetzt. Zudem ist die Reliabilitätsschätzung über Retest-Verfahren erst üblich und sinnvoll, wenn die Methoden Split-half und Innere Konsistenz durchgeführt und angemessene Ergebnisse erzielt wurden. Dies ist in der Regel zu einem späteren Zeitpunkt nach Abschluss der Testentwicklungsphase der Fall und wird daher im Rahmen der vorliegenden Arbeit nicht berücksichtigt. Zur Anwendung der Paralleltest-Methode ist ein mit dem zu untersuchenden Test streng vergleichbares Instrument nötig. Da ein entsprechendes Instrument für den Wortschatztest nicht existiert und nicht ausschließlich zur Reliabilitätskontrolle entwickelt werden soll, wurde auf die konsequente Durchführung der Paralleltest-Methode verzichtet, wenngleich die beiden aus der Split-half-Methode hervorgehenden Testhälften als quasi-parallel bewertet werden können.

Der Anspruch, für den Reliabilitätskoeffizienten Werte von .90 zu erreichen, wird in Anlehnung an Lienert/Raatz für den Wortschatztest als unrealistisch hoch angesehen, wenn man den Einfluss verschiedener Faktoren (vgl. 1994, 202ff) berücksichtigt: Für Tests, die einen umschriebenen Wissensstand wie den Wortschatz prüfen, gilt die Untersuchung der Reliabilität als erschwert. In einem solchen Fall kommen Paralleltests aufgrund des speziellen Aufgabeninhaltes nicht in Betracht und auch Testwiederholungen sind aufgrund des Erinnerungseffekts nicht möglich. Grundsätzlich gilt zudem, dass mit der Länge eines Tests und bei sonst vergleichbaren Bedingungen seine Reliabilität steigt. Eine Verlängerung des Wortschatztests würde demzufolge zwar eine mögliche Reliabilitätsverbesserung mit sich brin-

gen, ist aber unter Berücksichtigung der zeitlichen und kognitiven Leistungsmöglichkeiten der Testpersonen zu verwerfen. Wie bereits in anderem Zusammenhang (vgl. B 3.4.4) erwähnt, ist auch die Streuung der Aufgabenschwierigkeiten von großer Bedeutung für die Reliabilität, die bei mittelschweren Aufgaben und bei sonst vergleichbaren Bedingungen potentiell am höchsten liegt. Die spezielle Konzeption des Wortschatztests sieht jedoch vor allem Aufgaben von geringer Schwierigkeit (also hohen Schwierigkeitskoeffizienten) vor, sodass von vornherein keine optimalen Reliabilitätsbedingungen für diesen Test gegeben sind. Zusätzlich vermindert der Zufallseinfluss bei der Darbietung von Mehrfachwahlaufgaben die Reliabilität. In Anbetracht dieser verschiedenen Bedingungen werden im vorliegenden Fall niedrigere Werte ab .70 als zufriedenstellend bewertet.

Schließlich wurde die Auswahl einer für das vorliegende Instrument geeigneten Methode zur Testhalbierung für das Split-half-Verfahren nach folgenden Gesichtspunkten getroffen. Die Methode der zufälligen Aufteilung vermeidet zwar systematische Fehler, ist jedoch nur bei schwierigkeitsgleichen Aufgaben sinnvoll und muss demnach im vorliegenden Fall ausgeschlossen werden. Auch die Halbierung nach Testzeit kommt nicht in Frage, da es sich bei dem Wortschatztest um einen reinen Niveautest handelt, bei dem Schnelligkeit unberücksichtigt bleibt. Statt dessen wurde die Testhalbierung nach der auch üblicherweise angewendete odd-even-Methode vorgenommen. Dabei wurden die Aufgaben des neu sortierten Itemkorpus der Testversion II nach gerad- und ungeradzahlig Items aufgeteilt. Da die Aufgaben dieser Version nach den sich aus der Aufgabenanalyse ergebenden Schwierigkeitswerten und nicht mehr nach ihrem Vorkommen in der Wortliste sortiert sind, ist anzunehmen, dass eine gleichmäßige Schwierigkeitsverteilung auf beide Testhälften gewährleistet ist. Darüber hinaus wurde über die übliche Methode hinaus eine Halbierung anhand der Analysedaten vorgenommen. Damit sollte sichergestellt werden, dass die zusätzliche Einbeziehung der Trennschärfewerte die Gleichwertigkeit der Hälften nicht vorteilhafter gewährleistet und damit höhere Reliabilitätskoeffizienten provoziert. Dafür wurden anhand der vorliegenden Daten aus der Aufgabenanalyse Aufgabenpaare zusammengestellt, die jeweils ähnliche Schwierigkeits- und ähnliche Trennschärfewerte besitzen. Daraufhin wurde je eine Aufgabe eines Paares zufällig einer Testhälfte zugeordnet. Dies ermöglicht, bezüglich ihrer statistischen Kennwerte gleichwertige Testhälften zu konstruieren. Die Ergebnisse unterscheiden sich nur marginal von denen der Halbierung nach der odd-even-Methode und werden daher an dieser Stelle nicht gesondert aufgeführt (Ergebnisse siehe Anh. 1: B 3.7.2a, Anh. 1: B 3.7.2b und Anh. 1: B 3.7.2c mit den Tabellen Tab. 46-B 3.7.2, Tab. 47-B 3.7.2 und Tab. 48-B 3.7.2).

Basierend auf diesen Vorüberlegungen, die die spezifischen Bedingungen des Wortschatztests berücksichtigen, wurden die Berechnungen zur Reliabilität wie folgt vorgenommen. Um zunächst die Gleichwertigkeit der Testhälften A und B zu prüfen, werden in der folgen-

den Tabelle die Gesamtwerte (Summenmittelwerte) der Schwierigkeit beider Testhälften nach Halbierung über die odd-even-Methode in jedem Subtestteil einander gegenübergestellt. Zudem werden die jeweiligen Rohwertmittelwerte und Varianzen der Testhälften aufgelistet, die für die Berechnung und Auswahl der verschiedenen Halbierungskonsistenz-Formeln (vgl. Lienert/Raatz 1994, 184) zur Ermittlung der Reliabilität notwendig sein werden.

Tab. 49-B 3.7.2: Gesamtwerte der Testhälften A und B nach Halbierung über die odd-even-Methode			
Subtest-teil	Mittlere Schwierigkeit	Rohwert- Mittelwert	Varianz
GSN A	.74	6,5	3,44
GSN B	.72	6,3	3,30
GSV A	.86	7,6	2,85
GSV B	.83	7,3	2,39
GSA A	.71	6,3	3,59
GSA B	.66	5,8	3,01
SSN A	.68	6,0	6,21
SSN B	.64	5,7	6,05
SSV A	.55	4,8	6,63
SSV B	.53	4,6	6,61
SSA A	.57	5,0	7,34
SSA B	.55	4,8	6,65
LSN A	.68	5,8	6,09
LSN B	.66	5,6	6,33
LSV A	.66	5,6	5,80
LSV B	.62	5,3	5,21
LSA A	.59	5,0	6,73
LSA B	.57	4,9	7,03

Es wird deutlich, dass die Aufteilung der Aufgaben zu vergleichbaren Testhälften geführt hat, da die Mittelwerte der Aufgabenschwierigkeiten in den Hälften A und B aller Subtestteile weitestgehend übereinstimmen. Sie weichen durchschnittlich 2,9%, davon nie stärker als 5% (dies nur in Subtestteil GSA), meistens 2% (in fünf Subtestteilen) voneinander ab. Trotz der Begünstigung der Werte gegenüber einer vollkommen zufälligen Testhalbierung sind die Ergebnisse als zufriedenstellend zu bewerten.

Nach Konstruktion zweier gleichwertiger Testhälften wurden aus den Daten der Analysestichprobe für jede Testperson zwei Rohwerte ermittelt, deren Korrelation den Halbttest-Reliabilitätskoeffizienten r_{12} (vgl. Lienert/Raatz 1994, 184) ergibt. Verschiedene Verfahren ermöglichen es, aus diesem die Reliabilität des Gesamttests in Form des Halbierungs-Konsistenzkoeffizienten r_{tt} abzuschätzen. Die für die Split-half-Methode gebräuchlichste Halbierungskonsistenz-Formel nach Spearman-Brown (in Lienert/Raatz 1994, 185) liefert nur dann zuverlässige Werte, wenn die Varianzen beider Testhälften annähernd gleich sind, da sie in der Formel nicht speziell aufgeführt werden. Sollen die Varianzen gesondert einbezogen werden, so bietet sich die Formel von Flanagan (in Lienert/Raatz 1994, 185) an. Da diese Formel bei kleinem Stichprobenumfang, wie bei der Analysestichprobe des Wortschatztests der Fall, keine erwartungsgemäßen Angaben über die Grundgesamtheit liefert, wurde hier vor allem die abgeänderte Formel von Kristof (in Lienert/Raatz 1994, 187) verwendet, die sich auf die Formel von Flanagan bezieht. Aufgrund annähernd gleicher Varianzen beider Testhälften (s. o.) wurde zusätzlich die Formel nach Spearman-Brown angewendet, während das Ergebnis der Berechnungsmethode nach Flanagan vernachlässigt und daher nicht gesondert aufgeführt wurden. Es fließt jedoch in die Formel nach Kristof ein.

Im Rahmen der Methode der Inneren Konsistenz wurde im vorliegenden Fall auf eine Unterteilung in drei, vier oder mehr Teile verzichtet und statt dessen jede einzelne Aufgabe als gesonderter Testteil mit allen anderen korreliert. Für eine derartige Aufteilung in gleich lange, äquivalente Testteile wird üblicherweise der bekannte Alpha-Koeffizient von Cronbach (in Lienert/Raatz 1994, 192) berechnet. Zu gleichen Ergebnissen führt auch die Berechnung mit der Formel 20 von Kuder/Richardson (in Lienert/Raatz 1994, 193). Da die Formel 8 von Kuder/Richardson (in Lienert/Raatz 1994, 193) zusätzlich die einzelnen Trennschärfewerte der Aufgaben einbezieht, wurde zur Berechnung eines Schätzwertes der Reliabilität für den Wortschatztest vor allem diese Formel berücksichtigt. Zu einer Schätzung der Konsistenz unter Berücksichtigung des zufälligen Einflusses bei Mehrfachwahlaufgaben, wie sie dem Wortschatztest zugrunde liegen, gelangt man mit der Formel von Horst (in Lienert/Raatz 1994, 195).

In der folgenden Tabelle sind die Werte der Umwandlung des Halbttest-Reliabilitätskoeffizienten mittels verwendeter Formeln in einen geschätzten Reliabilitätskoeffizienten für jeden Subtestteil aufgeführt. Zum Erhalt realistischer Koeffizienten für den Gesamttest wurde darauf verzichtet, die Formeln auf eine Gesamtskala mit 81 Aufgaben pro Testhälfte anzuwenden, da dies zu einer Überschätzung der Reliabilität führen würde. Statt dessen wurde die Berechnung des Reliabilitätskoeffizienten für den Gesamttest über Umwandlung in z' -Werte nach Fisher (in Lienert/Raatz 1994, 417) vorgenommen.

Tab. 50-B 3.7.2: Reliabilitätskoeffizienten

	Split-half		Innere Konsistenz	
	r_{tt} Spearman- Brown	r_{tt} Kristof	r_{tt} Horst	r_{tt} Kuder / Richard- son 8
GSN	.79	.79	.85	.67
GSV	.83	.83	.86	.68
GSA	.79	.79	.83	.69
SSN	.87	.87	.91	.85
SSV	.86	.86	.90	.87
SSA	.87	.87	.91	.86
LSN	.85	.85	.90	.80
LSV	.87	.87	.88	.77
LSA	.87	.88	.91	.84
gesamt über z'	.85	.85	.89	.80

Die Reliabilitätskoeffizienten, die sich nach den Berechnungen mit unterschiedlichen Methoden ergeben, weisen den Gesamttest und alle Subtestteile als gut bis sehr gut reliabel aus. (Dieser Bewertung geht die oben erwähnte Einschränkung der Anforderung an Reliabilitätswerte um .90 voraus.) Die jeweiligen Besonderheiten, die in die unterschiedlichen Berechnungsmethoden einfließen, nehmen demnach keinen maßgeblichen Einfluss, der etwa gesondert berücksichtigt werden müsste.

3.7.3 Validität

Die Validität eines Tests beschreibt den Grad, mit der er das Merkmal, welches er zu messen beansprucht, tatsächlich misst.

Unter Berücksichtigung der Besonderheiten des Wortschatztests ergeben sich spezifische Rahmenbedingungen zur Untersuchung verschiedener Validitätsaspekte und der Interpretation der somit erhaltenen Werte. Zu Beginn des Validierungsprozesses steht in der Regel kein zufriedenstellendes Kriterium zur Verfügung (vgl. Lienert/Raatz 1994, 221). Erst im prozessualen Verlauf der Suche nach einem geeigneten Kriterium können vorläufige Kompromisse gegen bessere Kriterien ausgetauscht werden. So wurde auch für die Validierung des Wortschatztests zunächst ein vorläufiges, subjektives Außenkriterium in Form eines Ratingverfahrens (vgl. Lienert/Raatz 1994, 221ff) erhoben. Mit diesem Rating wurde ein Krite-

rium gewählt, das Übereinstimmungscharakter aufweist und zu diagnostischen statt zu prognostischen Zwecken verwendet wird. Die praktische Bewährung des Wortschatztests spricht bei sich herausstellender hoher Korrelation zwischen Test und Rating in diesem Zusammenhang für seine Durchführung, welche der Erhebung von Kriterienwerten gegenüber als präziser, objektiver und differenzierter angesehen wird.

Ein über Rating erhobener Koeffizient weist bestimmte nachteilige Eigenarten auf. Zum einen treten durch die Einzelschätzwerte bei einem individuellen Schätzverfahren relativ hohe Zufallsfehler auf. Durch die Auswahl eines solchen Kriteriums wurde daher die Objektivität des Wortschatztests zugunsten des subjektiven Ratingverfahrens zumindest teilweise geopfert. Zum anderen ist die Reliabilität eines individuellen, d. h. nur durch einen einzelnen Beobachter erhobenen Ratingverfahrens nur schwer zu bestimmen. Eine zweimalige Erhebung des Kriteriums war im vorliegenden Fall nicht möglich. Ein niedriger Zusammenhang zwischen Test und Kriterium kann deshalb zwar durch eine geringe Validität des Tests, aber auch durch ein unreliaables Kriterium mit zufälligen Skalenwerten zustande kommen. Eine eindeutige Aussage zur Tauglichkeit des Tests in Hinblick auf Validität ist demnach nur bei hoher Korrelation möglich, wodurch der Test als geeignetes Instrument zur Erfassung des Konstrukts ‚passive Wortschatzkenntnis‘ ausgewiesen wird.

Die Erwartungen an die Höhe des Validitätskoeffizienten hängen wesentlich von dem jeweiligen Verwendungszweck des Tests ab. Da der vorliegende Test zwar ein eindeutiges Urteil über den Ausprägungsgrad des zu untersuchenden Merkmals einer Testperson erlaubt, jedoch keine Beurteilung der Testperson überwiegend oder ausschließlich von seinem Testergebnis abgeleitet werden soll, muss der Test einen Validitätsanspruch von $r_{tc} \geq 0,5$, und nicht etwa von $r_{tc} > 0,7$ (vgl. Lienert/Raatz 1994, 270) erfüllen.

Schließlich hängt der Grad der externen bzw. der internen Validität von der Auswahl und der Kontrolle von Störvariablen ab. Bei der Erhebung des Wortschatzes können die Uhrzeit, der Termin, die Lichtverhältnisse u. ä. potentiell bedeutsam sein, obwohl sie weder als abhängige noch als unabhängige Variable auf die Untersuchung Einfluss nehmen. Während einige dieser Variablen wie Lichtverhältnisse oder Gebärdenvorgabe von den Testleitern kontrolliert wurden, wurden andere Variable wie Termin und Uhrzeit zufällig gewählt. Durch das so festgelegte Maß der Kontrolle dieser untersuchungsbedingten Störvariablen ist die Untersuchung im mittleren Bereich des Kontinuums von Untersuchungsvarianten angesiedelt. Dieser Kompromiss aus Labor- und Felduntersuchung gewährleistet interne und externe Validität. Damit sind die Ergebnisse der Untersuchung sowohl eindeutig interpretierbar, ohne dass Alternativerklärungen für ihr Zustandekommen herangezogen werden müssen, als auch über die besonderen Bedingungen der Untersuchungssituation und über die untersuchten Personen

hinaus allgemein gültig, indem sie die Natürlichkeit der Untersuchungsbedingungen und die Repräsentativität der untersuchten Stichprobe berücksichtigt (vgl. Bortz 1993, 8).

Innerhalb dieser Rahmenbedingungen erfolgte der Validierungsprozess des Wortschatztests nach den üblichen Untersuchungsaspekten. Die inhaltliche Validität eines Instrumentes ist gewährleistet, wenn der Test oder seine Teile das zu messende Merkmal so unzweifelhaft repräsentieren, dass sie selber das Kriterium für dieses Merkmal darstellen. Dies ist bei Kenntnistests wie dem Wortschatztest grundsätzlich der Fall (vgl. Lienert/Raatz 1994, 225). Die Items des Wortschatztests entstammen einer den Grundwortschatz hörender Grundschüler abbildenden Wortliste und wurden aus dieser systematisch in regelmäßigen Abständen über den Gesamtkorpus verteilt ausgewählt. Sie bilden somit eine repräsentative Stichprobe aus dem Grundwortschatz. Demgemäß wurde auf eine Einschätzung durch Experten verzichtet, da die inhaltliche Validität des Wortschatztests trivial ist (vgl. Lienert/Raatz 1994, 225).

Für die Überprüfung der Konstruktvalidität des Wortschatztests wurden aus der theoretischen Analyse des Konstruktes Hypothesen deduziert, die daraufhin empirisch überprüft wurden. Da es sich bei der 'Kenntnis des passiven Wortschatzes' (vgl. Begrifflichkeiten A 2.3) um eine vielschichtige, nicht direkt operational fassbare Einheit (vgl. Lienert/Raatz 1994, 226) handelt, ist eine differenzierte Auseinandersetzung mit diesem Konstrukt erforderlich. Theoretisch analysiert stellt sich das Konstrukt als die Wiedererkennung und richtige Zuordnung von konventionellen Bezeichnungen dar. Die Verfügbarkeit dieser Bezeichnungen wird als Anzeiger für die entwickelte Versprachlichung kognitiver Prozesse angesehen. Der Ausprägungsgrad dieses Konstruktes wird von verschiedenen Faktoren bestimmt. Generell wächst die Kenntnis des Wortschatzes mit zunehmendem Alter und mit steigendem Intelligenzniveau und ist vom Geschlecht relativ unabhängig. Einen weit deutlicheren Einfluss als das Alter übt bei Gehörlosen jedoch die Einordnung der Testpersonen in Klassenstufen aus. Diese erfolgt nach dem jeweiligen Förderbedarf, sodass Schüler unterschiedlichen Alters, aber gleichen Leistungsniveaus in einer Klasse zusammengefasst werden. Darüber hinaus nehmen sowohl schulische als auch familiäre Sozialisationsbedingungen Einfluss auf die Wortschatzentwicklung gehörloser Kinder: Sie wird auf schulischer Seite maßgeblich von dem angewandten sprachlichen Erziehungskonzept geprägt. Wie Untersuchungsergebnisse des Hamburger Bilingualen Schulversuchs u. a. belegen, muss bei bilingual erzogenen gehörlosen Kindern von einem differenzierter ausgebildeten Wortschatz ausgegangen werden, als bei rein oral- oder gebärdenorientierter Erziehung. Als zentrale Begründung dafür gilt der Umstand, dass den bilingual erzogenen Kindern mit der Gebärdensprache ein Sprachsystem zur Verfügung steht, auf dessen Basis nicht muttersprachliche, lautsprachliche Fähigkeiten vollständiger erworben werden können (vgl. Günther u. a. 1999 a; Prillwitz / Wudtke 1988). Ein Vergleich der Fähigkeiten bzgl. der Sprachmodalitäten weist demnach die bilingual er-

zogenen Kinder als kenntnisreich sowohl in der Gebärden- als auch in der Lautsprache aus, wohingegen die oral erzogenen Schüler vor allem Stärken in der Laut-, die gebärdenorientiert erzogenen Schüler in der Gebärdensprache aufweisen. Dabei wird davon ausgegangen, dass aufgrund der generellen Neigung Gehörloser, mit der Gebärdensprache als der ihnen am leichtesten zugänglichen Modalität zu kommunizieren, alle Kinder in der Modalität Gebärdensprache am leistungsstärksten abschneiden, und dies unabhängig von dem ihrer jeweiligen Erziehung zugrundeliegenden Konzept (vgl. C 1.1 und C 1.3). Zudem üben familiäre Rahmenbedingungen wie der Umstand, ob die Eltern des gehörlosen Kindes selber gehörlos oder hörend sind, ob also von der Gebärdensprache als Muttersprache ausgegangen werden kann oder nicht, einen Einfluss auf seine Wortschatzentwicklung aus. Gehörlose Kinder gehörloser Eltern erlernen aufgrund der Basis der Gebärdensprache, die sie sich von Geburt an entsprechend der lautsprachlichen Lernprozesse hörender Kinder aneignen, die Lautsprache als Zweitsprache leichter und vollständiger als gehörlose Kinder hörender Eltern. Daher kann davon ausgegangen werden, dass gehörlose Kinder gehörloser Eltern ein höheres Wortschatzniveau in allen Sprachmodalitäten aufweisen.

Zunächst wurde zur Überprüfung der Konstruktvalidität eine Analyse der interindividuellen Unterschiede in den Testresultaten vorgenommen. Untersuchungsziel dabei war die Überprüfung der Annahme, dass der Test für die verschiedenen, die Wortschatzentwicklung beeinflussenden Faktoren sensibel ist. Zeigen sich deutliche Zusammenhänge in oben beschriebener Weise, so wird dies als Beleg dafür angesehen, dass das Testinstrument tatsächlich das beabsichtigte Konstrukt 'Kenntnis des passiven Wortschatzes' misst. Herangezogen werden folgende Faktoren, korreliert nach Pearson und minderungskorrigiert:

Zeile 1: Klasse, in den gegebenen Abstufungen 3 und 4 (+5)

Zeile 2: Intelligenzquotient nach Raven, in den Abstufungen der Testpunktwerte
67 bis 132

Zeile 3: Konzept, in den Abstufungen 1 (oral), 2 (gebärdenorientiert) und 3 (bilingual)

Zeile 4: Hörstatus der Eltern, in den Abstufungen 1 (hörende Eltern) und 2 (gehörlose Eltern)

Zeile 5: Geschlecht, in den Abstufungen 1 (weiblich) und 2 (männlich)

Zeile 6: Alter, in den gegebenen Abstufungen 7;11 bis 13;3 Jahre

Tab. 51-B 3.7.3: Korrelationen Wortschatztest – Faktoren									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
1. Korr. Klasse	.35	.36	.33	.44	.41	.40	.31	.29	.34
Mind.-korrektur	.39	.40	.37	.47	.43	.44	.33	.32	.36
p	.000	.000	.000	.000	.000	.000	.000	.000	.000
2. Korr. IQ	.33	.32	.40	.32	.36	.42	.30	.22	.38
Mind.-korrektur	.39	.37	.46	.36	.39	.47	.35	.25	.43
p	.000	.001	.000	.001	.000	.000	.002	.026	.000
3. Korr. Konzept	.31	.10	.29	.29	.27	.29	.20	.22	.18
Mind.-korrektur	.35	.12	.32	.31	.28	.31	.22	.24	.19
p	.001	.287	.002	.002	.005	.003	.042	.024	.073
4. Korr. Eltern	.38	.20	.37	.31	.31	.37	.26	.24	.30
Mind.-korrektur	.42	.22	.41	.33	.32	.41	.29	.26	.32
p	.000	.039	.000	.001	.001	.000	.006	.013	.002
5. Korr. Sex	.12	-.10	-.07	-.07	-.03	-.01	.07	-.01	.02
Mind.-korrektur	.13	-.11	-.07	-.07	-.03	-.01	.08	-.02	.02
p	.213	.305	.501	.490	.794	.894	.472	.889	.879
6. Korr. Alter	-.00	.08	.01	.09	.06	.04	.05	.07	.03
Mind.-korrektur	-.01	.09	.01	.09	.06	.04	.06	.08	.04
p	.969	.395	.934	.359	.561	.686	.587	.475	.738

Gemäß der Hypothese bestehen sehr signifikante Zusammenhänge zwischen der Wortschatzentwicklung und den Einflussgrößen ‘Klassenstufe’, ‘Intelligenzniveau’ sowie ‘Hörstatus der Eltern’. In sieben von neun Subtestteilen hängt die Kenntnis des passiven Wortschatzes signifikant von dem ‘Erziehungskonzept’ ab. Im Falle von ‘Geschlecht’ und ‘Alter’ ist kein signifikanter Einfluss auf das Wortschatzniveau zu erkennen. Die Variation der Testergebnisse steht demnach in erwarteter Relation zu der Variation der schulischen und familiären Sozialisationsbedingungen. Diese Ergebnisse lassen darauf schließen, dass es sich bei dem vorliegenden Test tatsächlich um ein Messinstrument zur Erfassung des Konstruktes ‘Kenntnis des passiven Wortschatzes’ handelt, da sich die erwarteten Beziehungen zwischen dem Testergebnis und den untersuchten, den Wortschatz beeinflussenden Variablen bestätigt haben. Dadurch wird dem Wortschatztest eine Konstruktvalidität im Sinne der Abbildung der Wirkung unterschiedlicher Sozialisationsbedingungen zugeschrieben.

Zudem wurden die Ergebnisse des Wortschatztests mit denen eines Tests korreliert, der ein verwandtes Persönlichkeitsmerkmal erfasst. Da Wortschatz aus Sicht der Intelligenzforschung als ein Teilaspekt der Intelligenz angesehen wird (vgl. HAWIK-R Hamburg-Wechsler Intelligenztest für Kinder Revision 1983, Handbuch und Testanweisung 1985, 15/16), wurde der Analysenstichprobe neben dem Wortschatztest auch der Raven-Matrizen-Test vorgelegt. Der Raven wird in diesem Zusammenhang als ökonomisches Kontrollinstrument angesehen, mit dem in non-verbaler Form gravierende Einflüsse auf die Wortschatzentwicklung im Allgemeinen und die Überprüfung der passiven Wortschatzkenntnis mit dem Wortschatztest im Speziellen festgestellt und ausgeschlossen werden können.

Die Ergebnisse des Wortschatztests und des Raven wurden nach Pearson (erste Zeile) sowie einfach (zweite Zeile) und zweifach minderungskorrigiert (dritte Zeile) korreliert. Die Signifikanzwerte werden in der vierte Zeile aufgeführt. Bei der Minderungskorrektur wurde für den Raven ein Reliabilitätskoeffizient von .90 (Raven-Matrizen-Test, CPM, Manual 1980, 14), für den Wortschatztest der Koeffizient nach Kristof (s. B 3.7.2) verwendet.

Tab. 52-B 3.7.3: Korrelation Wortschatztest – Raven										
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA	ges.
Korrela-tion										
Raven	.33	.315	.40	.32	.36	.42	.30	.22	.38	.36
einfach										
mind.-korr.	.35	.332	.42	.34	.38	.44	.32	.23	.40	.37
zweifach										
mind.-korr.	.39	.371	.46	.36	.39	.47	.35	.25	.43	.39
p	.000	.001	.000	.001	.000	.000	.002	.026	.000	.000

Diese Ergebnisse sind unter zwei Gesichtspunkten zu interpretieren: Die theoretische Bedeutsamkeit einer Korrelation zeigt sich generell in der Unterschreitung eines Signifikanzniveaus. Sie liegt in diesem Fall hoch, da die Signifikanzergebnisse in acht von neun Fällen, wie auch für den Gesamttest, auf der 1%-Stufe, im neunten Fall immer noch auf der 5%-Stufe gesichert sind. Somit besteht bei geringer Irrtumswahrscheinlichkeit ein wahrer, interpretierbarer Zusammenhang zwischen Test und Kriterium. Die Wahl des Raven als Außenkriterium impliziert und rechtfertigt gleichzeitig die verhältnismäßig niedrigen Korrelationskoeffizienten. Der Wortschatztest berücksichtigt in stärkerem Maße als ein Intelligenztest (und besonders als der non-verbale Raven) die verbale Entwicklung. Zudem kann Wortschatz nur als ein Teilaspekt der gesamten Intelligenz angesehen werden, sodass nur ein bedingter Grad der Gemeinsamkeit zwischen Test und Kriterium besteht. Die entsprechenden Koeffizienten sind als Anzeiger der praktischen Relevanz demnach als gut zu bewerten, da ein stärkerer Zusammenhang zwischen Test und Kriterium nicht zu erwarten ist.

Für die kriterienbezogene Validitätsprüfung des Wortschatztests wurde parallel zu den Testpunktwerten im Sinne eines gesondert erhobenen Außenkriteriums ein Schätzurteil über die mutmaßliche Leistung einer Testperson durch die Klassenlehrer erhoben. Für diese Übereinstimmungsvalidität des Wortschatztests nahmen die Klassenlehrer der Testpersonen eine Einschätzung der passiven Wortschatzkenntnis ihrer Schüler vor, indem sie sie auf einer qualitativ abgestuften Ratingskala einordneten. Es wurde eine siebenstufige Ratingskala (vgl. Lienert/Raatz 1994, 237) mit folgenden Einteilungen vorgegeben:

Abb. 15-B 3.7.3: Ratingskala						
1	2	3	4	5	6	7
sehr schlecht	schlecht	eher schlecht	durch- schnittlich	eher gut	gut	sehr gut

Für jede Testperson wurden drei Ratings erhoben, um ihre mutmaßliche Leistung in den drei Sprachmodalitäten getrennt zu erfassen. Nicht alle Klassen konnten an diesem Rating teilnehmen, sodass aufgrund zu geringer Testpersonenanzahl keine zufällige Stichprobe aus der Analysestichprobe gewählt werden konnte. Da eine Validitätsstichprobe von einer Größe ab 30 Testpersonen als ausreichend umfangreich angesehen wird (vgl. Lienert/Raatz 1994, 241), stellen die in etwa normalverteilten Rating-Ergebnisse von insgesamt 62 Testpersonen eine genügend große Stichprobe dar. (Die Ergebnisse des Ratings sind im Anhang Anh. 1: B 3.7.3 in der Tab. 53-B 3.7.3 dargestellt.)

Die Skalenzuordnung wurde nach der Repräsentativgruppen-Methode (vgl. Lienert/Raatz 1994, 245) mit den Testergebnissen der dritten Version korreliert. Es wurden dabei jeweils die Testergebnisse der drei Subtestteile einer Modalität mit dem entsprechenden Rating-Ergebnis korreliert. Die Gesamtergebnisse einer Modalität oder einer Wortart sowie der Gesamtscore des Tests wurden über Transformation in z' -Werte nach Fisher berechnet:

Tab. 54-B 3.7.3: Korrelation Wortschatztest – Rating				
Modalität	Nomen	Verben	Adjektive	gesamt
Gebärdensprache	.58	.44	.58	.54
Schriftsprache	.67	.79	.78	.76
Lautsprache	.62	.62	.70	.65
gesamt	.63	.64	.70	.65

Alle Ergebnisse sind sehr signifikant, sodass ein wahrer Zusammenhang zwischen Test und Kriterium angezeigt wird. Da bei der Erhebung des Kriteriums Skalenwerte für die Fähigkeiten in einzelnen Modalitäten verteilt wurden, die die Fähigkeiten in den einzelnen Wortarten der entsprechenden Modalität beinhalten und zusammenfassen, können die erhaltenen

Koeffizienten möglicherweise als zu wenig differenzierend und damit zu gering angesehen werden. Die Koeffizienten der Subtestteile erfüllen dennoch mit Werten zwischen .44 und .80 die Anforderungen an einen Test, der zur Informationsgewinnung über den Kenntnisstand des passiven Wortschatzes einer Testperson beitragen soll. Auch die Gesamtwerte der Modalitäten von .54 bis .76 und der Wortarten von .63 bis .70 weisen den Test nach dieser ersten Validierung als gut bis sehr gut valide aus. Dies Urteil wird durch den Umstand unterstützt, dass sowohl Test als auch Kriterium Zufallsmomente aufweisen, deren gemeinsame Variation den in der Korrelation beschriebenen Zusammenhang möglicherweise niedriger ausfallen lässt.

Unter Berücksichtigung des oben beschriebenen Sachverhalts wurden für die Gesamtwerte der Modalitäten zusätzlich direkte Korrelationen mit den erhobenen Kriterienwerten ohne Transformation in z' -Werte berechnet. Dafür wurde für jede Testperson ein Summenwert für jede Modalität aus den drei entsprechenden Subtestteilergebnissen des Wortschatztests errechnet und mit dem Ratingergebnis der Modalität korreliert. Genaugenommen entspricht nur dieser Wert der mutmaßlichen Ratingeinstufung einer Testperson und nicht ihr Ergebnis in einzelnen Subtestteilen, für die kein gesondertes, weiter differenzierendes Rating erhoben wurde. Durch derartige Korrelation ergeben sich sogar Werte von .64 (GS), .79 (SS) und .70 (LS). Über Transformation in z' -Werte lässt sich daraus eine Gesamtkorrelation für den Wortschatztest von .72 berechnen, die die Validität des Tests noch höher einstuft. Während die Korrelationswerte durch die z' -Transformation leicht unterschätzt werden, fallen die Zusammenhänge bei dieser Methode etwas zu hoch aus. Der wahre Korrelationswert wird vermutlich in der Mitte der Ergebnisse beider Verfahren, also für Gebärdensprache um .59, für Schriftsprache um .78 und für Lautsprache um .68 anzusiedeln sein. Der wahre Wert für den Gesamttest liegt damit um .69.

3.8 Normierung

Um eine eindeutige Interpretation der Rohdaten einer Testperson aus dem Wortschatztest zu gewährleisten, ist eine Orientierung an vergleichbaren Messträgern nötig. Derartige Vergleichswerte sollen durch die Normierung bereitgestellt werden. Ziel der Normierung des Wortschatztests war dabei, die erhaltenen Rohdaten anschaulicher, direkter vergleichbar und messtechnisch möglichst höherwertig zu machen. Diese Qualitätssteigerung in Hinblick auf die Messtechnik der Rohdaten wurde vorgenommen, um für die empirische Untersuchung von vornherein das höchstmögliche Niveau anzusetzen. Transformationen der gewonnenen Daten auf ein niedrigeres Niveau sind danach immer möglich, ein nachträgliches Anheben auf ein höheres Niveau jedoch nicht.

3.8.1 Skalenniveaus

Zu jedem Vorgang des Messens gehört eine Menge von Objekten, deren Beziehungen untereinander durch eine oder mehrere Relationen beschrieben werden und die zusammen mit diesen Relationen ein Relativ bildet. Gegenstand des Messens ist dabei, ein empirisches Relativ mit Hilfe einer homomorphen Abbildung derart in ein numerisches Relativ mit gleichen Relationen abzubilden, dass jede Beziehung, die zwischen zwei Objekten des empirischen Relativs besteht, auch zwischen den zugeordneten Zahlen gewährleistet ist und umgekehrt (vgl. Orth 1983, 138). Mit Hilfe der zugeordneten Zahlen sollen Aussagen über den Ausprägungsgrad von Variablen getroffen werden. Welcher Art diese zu treffenden Aussagen sind, hängt von den jeweiligen Variablen und der benutzten Messskala ab. Für den Wortschatztest setzt sich das empirische Relativ aus der Menge von 112 Testpersonen und ihrer Beziehung ‘Objekt a hat eine genauso hohe oder höhere passive Wortschatzkenntnis wie/als Objekt b’ zusammen. Diese Beziehung wird in 72 (Version I) bzw. 54 (Pilotstudie und Version II) Variablen, also Testaufgaben erhoben. Durch eine homomorphe Abbildung sollen diesem Relativ Zahlen zugeordnet werden, die genau diesen Zusammenhang beibehalten und widerspiegeln. Derartige Zuordnungen können sich auf verschiedenen hohen Niveaus abspielen. Auf welchem Niveau sich das Messen mit dem Wortschatztest bewegt, soll im Folgenden geklärt werden. Dabei werden die Skalenniveaus in hierarchischer Reihenfolge betrachtet, da jede Skala die Axiome der jeweils niedrigeren Skalen beinhaltet.

Eine eindeutige Identifikation der Testpersonen und ihre Einteilung in Klassen (ohne Größer-Kleiner-Relation) wird durch die Zahlenzuordnung des Wortschatztests in jedem Fall gewährleistet. Den Objekten mit gleicher Merkmalsausprägung wird die gleiche Zahl, Objekten mit verschiedener Merkmalsausprägung werden verschiedene Zahlen zugeordnet. Die Größe der Zahlen kann dabei bisher als unerheblich angesehen werden, da auf diesem Niveau noch keine Aussage über den Wert der Zahl getroffen werden muss. Gleichzeitig gelten für das empirische Relativ des Wortschatztests die Bedingungen der Reflexivität (Objekt a steht zu sich selbst in Relation), der Symmetrie (Wenn Objekt a zu b in Beziehung steht, dann auch umgekehrt) und der Transitivität (Wenn Objekt a zu b und gleichzeitig b zu c in Relation steht, dann steht auch a zu c in Relation). Somit erfüllt der Test die Axiome der Nominalskala. Bei der Zuordnung anderer Zahlen durch Transformation der homomorphen Abbildung wäre in diesem Fall lediglich darauf zu achten, dass Objekte, die vorher in derselben Klasse waren, es auch unter Verwendung der neuen Zahlencodierung noch sind.

Auch die Anforderungen an das zweite Messniveau, die Ordinalskala, können von dem Wortschatztest erfüllt werden, da die Zuordnung von Zahlen sogar in Form einer Rangfolge mit geltender Größer-Kleiner-Relation erfolgt. Auf diesem Niveau müssen gleiche Zahlenabstände noch nicht bedeuten, dass auch zwischen dem Grad der Merkmalsausprägung gleiche Abstände herrschen. Im empirischen Relativ des Wortschatztests gilt – neben den für die

schwächere Nominalskala geltenden Eigenschaften Reflexivität, Symmetrie und Transitivität – auch die Bedingung der Konnexität, durch die immer feststellbar ist, ob das zu untersuchende Merkmal bei Objekt a oder bei Objekt b stärker ausgeprägt ist, oder aber, ob beide Objekte bzgl. des Merkmals äquivalent sind. Das empirische Relativ wird dabei durch alle Abbildungen, die die Rangfolge der Objekte beibehalten, in dem entsprechenden numerischen Relativ repräsentiert. Bei jeder zulässigen Transformation muss dieses Größer-Kleiner-Verhältnis beibehalten werden, sodass die quantitativen Aussagen der Ordinalskala des Tests gegenüber jeder beliebigen monotonen Transformation invariant sind.

Für eine Intervallskala entsprechen gleich große Abstände zwischen den Messwerten auch gleich großen Abständen zwischen den Ausprägungsgraden der Variablen. Wie bei einer Ordinalskala wird auch hier eine Rangfolge abgebildet, jedoch eine Rangordnung der Differenzen je zweier Objekte und nicht mehr nur eine Rangordnung der Objekte selbst. Diese Bedingungen kann der Wortschatztest mit seinen erhobenen Rohdaten in diesem Stadium noch nicht erfüllen. Abbildungen, die die Verhältnisse der Abstände zueinander beibehalten und lediglich in Messeinheit und -ursprung variieren, existieren für die Rohwertskalen des Tests nicht. Da die Skalenniveaus nach ihrer Komplexität geordnet sind, kann der Wortschatztest bei nicht geltendem Intervallskalenniveau die Bedingungen der Verhältnisskala sicher nicht erfüllen. Für Messungen auf diesem Niveau müsste die Existenz eines absoluten Nullpunktes gesichert sein, sodass nicht nur die Gleichabständigkeit von Differenzen, sondern auch die Gleichheit von Verhältnissen gewährleistet ist. Dies ist bei dem vorliegenden Instrument nicht der Fall, da über eine Testperson mit einem Rohwert von Null nicht schlussfolgernd die Aussage getroffen werden darf, dass sie über keinerlei Kenntnis des passiven Wortschatzes verfügt.

Für den Wortschatztest wird demnach zunächst vom Ordinalskalenniveau ausgegangen. Für die empirische Menge von Testpersonen und die Relation ‘Objekt a hat eine genauso hohe oder höhere passive Wortschatzkenntnis wie/als Objekt b’ gelten die Bedingungen der Reflexivität, der Symmetrie, der Transitivität und der Konnexität uneingeschränkt. Von Gleichabständigkeit der Wortschatzausprägung entsprechend den erzielten Rohwerten kann dagegen nicht unbedingt ausgegangen werden. Um die Daten messtechnisch dennoch auf Intervallskalenniveau zu bringen und sie gleichzeitig anschaulich und direkt vergleichbar zu machen, müssen sie auf neue Skalen transformiert werden.

3.8.2 Prozentrangskalen

Um eine fundierte Interpretation der gewonnenen Daten zu gewährleisten, wurden die Rohscores des Tests zunächst auf die Prozentrangskala transformiert, um sie anschaulicher und direkt vergleichbar zu machen. Eine Schwierigkeit in der Prozentrangberechnung vor-

liegender Daten lag in der ganzzahligen Abstufung der Variablen, obgleich diese Variable als psychologische Fähigkeit an sich stetig ist. Genaugenommen befindet sich die Leistung einer Testperson, die den Rohscore x erhält, also irgendwo zwischen $x-0,5$ und $x+0,5$ Punkten. Aus diesem Grund wurde die Hilfsannahme getroffen, dass sich alle Testpersonen mit dem Rohscore x den Platz zwischen $x-0,5$ und $x+0,5$ gleichabständig teilen. So kommt es, dass möglicherweise mehr als die objektiv schlechteren Testpersonen (Rohscore $x-1$ und schlechter), nämlich auch diejenigen mit dem Rohscore x , die zwischen $x-0,5$ und $x+0,5$ angesiedelt sind, den Prozentrang der Testperson bestimmen.

Die Skala der Prozentränge erfüllt die Ansprüche an Anschaulichkeit und Vergleichbarkeit, weist allerdings auch Nachteile auf. Zum einen täuscht sie ein nicht geltendes Skalenniveau vor, da 60% zwar doppelt so viel sind wie 30%, eine Testperson mit dem Prozentrang 60 jedoch nicht zwangsläufig doppelt so gut ist wie eine mit Prozentrang 30. Diese Skala hat demnach kein Intervallskalenniveau. Zum anderen gibt die Prozentrangskala eine unrealistische Messpräzision vor, wenn aufgrund der Angabe von Prozenträngen angenommen wird, das Messinstrument sei zu hundert Unterscheidungen fähig. Das Instrument kann nur der Anzahl der Scores entsprechend viele Prozentränge, also maximal 18 pro Subtestteil, angeben.

3.8.3 Lineare Standardwerte

Um zur Anschaulichkeit und Vergleichbarkeit auch die Anhebung auf Intervallskalenniveau zu gewährleisten, wird zusätzlich zu den Prozenträngen die Transformation der Rohwerte in lineare z -Werte vorgenommen. Diese Werte haben den Mittelwert 0 und die Standardabweichung 1 und erfüllen somit den Anspruch an direkte Vergleichbarkeit.

Wenn die Rohwertverteilungen der Stichprobe annähernd normal verteilt sind, erhält man bei der linearen Transformation nicht nur z -Werte, sondern Standard- z -Werte einer Standardnormalverteilung, die automatisch Intervallskalenniveau besitzen. Für den Fall, dass die Rohwerte nicht normal verteilt sind, muss zunächst eine Flächentransformation durchgeführt werden, um die Zuordnung von linearen Standard- z -Werten zu ermöglichen. Dabei handelt es sich um eine nachträgliche Zerrung und Stauchung der Messwertskala, bis die Häufigkeitsverteilung annähernd der Normalverteilung gleicht. Dieser Vorgang ist nötig und erlaubt, da es sich bei nicht normalverteilten Messwerten bestenfalls um ordinalskalierte Daten handelt, denen Gleichabständigkeit fehlt. Auf Ordinalskalen sind alle monotonen Transformationen erlaubt und eine solche stellt das oben erwähnte Zerren und Stauchen dar, denn es behält die Reihenfolge der Daten bei.

Um negative z-Werte zu vermeiden, werden die z-Werte in anschaulichere Z-Werte transformiert, die einen Mittelwert von 100 und eine Standardabweichung von 15 aufweisen.

3.8.4 Normierungstabellen

Für die Normierung des Wortschatztests wurden die Ergebnisse der Dritt-, Viert- (und Fünft-)klässler zusammengefasst. Dies geschah zum einen um eine Reduzierung des Datenpools und damit Zufallseinflüsse zu vermeiden, zum anderen unter der Voraussetzung, dass mit den Ergebnissen des Tests keine expliziten Aussagen über Leistungsunterschiede in der lexikalisch-semantischen Kompetenz zwischen Dritt- und Viertklässlern getroffen werden sollen. Vielmehr soll der Test dem unterrichtenden Pädagogen eine Aussage darüber ermöglichen, ob ein hörgeschädigter Grundschüler (egal welcher Klassenstufe) einen vergleichbaren Grundwortschatz besitzt wie ein hörender.

Die Normierung wurde vollständigshalber zweigleisig vorgenommen: Unter Annahme von normalverteilten Daten wurden Z-Werte und Prozentränge über die lineare Transformation berechnet; für den Fall von nicht normalverteilten Daten wurde zusätzlich ihr Berechnungsweg über kumulierte Häufigkeiten gewählt. Die weitgehende Übereinstimmung der erhaltenen Werte beweist endgültig die Normalverteilung der Daten, sodass an dieser Stelle lediglich die Berechnung der z-, Z- und Prozentrangwerte über die lineare Transformation aufgeführt wird. (Die Ergebnisse der Berechnung über Flächentransformation sind im Anhang Anh. 1: B 3.8.4, Tab. 55a-i-B 3.8.4 aufgeführt.) Wesentliche Informationen über die Wortschatzkenntnis einer Testperson können vor allem aus der Interpretation der Subtestteilergebnisse gewonnen werden. Zwar lassen extreme Gesamtwerte Rückschlüsse auf ein homogenes Leistungsvermögen in allen Subtestteilen zu, Gesamtwerte im mittleren Bereich können jedoch unterschiedlich (durch durchschnittliche Leistung in allen Skalen, oder aber durch überdurchschnittliche Leistung in einigen, unterdurchschnittliche Leistung in anderen Subtestteilen) bedingt sein. Gesamtwerte verlieren somit an Aussagekraft; interessanter sind Aussagen dazu, in welchen Subtestteilen eine Testperson über- oder unterdurchschnittliche Leistungen erbracht hat. In den folgenden Normierungstabellen können demzufolge die z-, Z- und Prozentrangwerte für jeden Subtestteil gesondert abgelesen werden. (Für die übersichtliche und anwenderfreundliche Darstellung der Ergebnisse werden dennoch alternative Normierungstabellen der Gesamtwerte der Subtests im Anwenderhandbuch (s. D 5.1.1, Tab. 105a-c-D 5.1.1) aufgeführt.) Da der Mittelwert und die Standardabweichung innerhalb der Skalen variieren, ändert sich auch die Höhe des Erwartungswerts. Für den z-Wert liegt er dagegen immer bei 0, für Z-Werte bei 100 und für Prozentränge bei 50. Jeweils eine Standardabweichung ober- und unterhalb des Mittelwerts liegen die Leistungen von etwa 66% aller Testpersonen, zwei Standardabweichungen ober- und unterhalb des Mittelwertes die Leistungen von 95% aller Testpersonen.

Tab. 56a-B 3.8.4: Normierung bei Normalverteilung Subtestteil Gebärdensprache Nomen (arithm. Mittel = 13,1; Standardabw. s = 2,87)

RW	f	x	z	Z	PR
0	0	- 13	-	-	0
1	0	- 12	-	-	0
2	0	- 11	-	-	0
3	0	- 10	-	-	0
4	0	-	9	-	-
5	0	-	8	-	-
6	3	- 7	- 2,44	76	1
7	1	- 6	- 2,09	79	2
8	0	- 5	- 1,74	83	5
9	8	- 4	- 1,39	86	8
10	8	- 3	- 1,05	89	14
11	15	- 2	- 0,70	93	24
12	11	- 1	- 0,35	96	35
13	11	0	0,00	100	50
14	16	1	0,35	104	66
15	12	2	0,70	107	76
16	8	3	1,05	111	87
17	13	4	1,39	114	92
18	4	5	1,74	117	96

0
0

Tab. 56b-B 3.8.4: Normierung bei Normalverteilung Subtestteil Gebärdensprache Verben (arithm. Mittel = 15,2; Standardabw. s = 2,21)

RW	f	x	z	Z	PR
0	0	- 15	-	-	0
1	0	- 14	-	-	0
2	0	- 13	-	-	0
3	0	- 12	-	-	0
4	0	- 11	-	-	0
5	0	- 10	-	-	0
6	0	-	9	-	-
7	1	-	8	-	-
8	1	-	7	-	-
9	1	-	6	-	-
10	1	- 5	- 2,26	77	1
11	5	- 4	- 1,81	82	4
12	4	- 3	- 1,36	86	8
13	5	- 2	- 0,90	91	18
14	12	- 1	- 0,45	95	31
15	21	0	0,00	100	50
16	23	1	0,45	105	70
17	26	2	0,90	109	82
18	10	3	1,36	114	92

0
0
0
0

**Tab. 56c-B 3.8.4: Normierung bei Normalverteilung Subtestteil Gebärden-
sprache Adjektive (arithm. Mittel = 12,4; Standardabw. s = 2,89)**

RW	f	x	z	Z	PR
0	0	- 12	-	-	0
1	0	- 11	-	-	0
2	0	- 10	-	-	0
3	0	-	9	-	-
4	0	-	8	-	-
5	0	- 7	- 2,43	76	1
6	0	- 6	- 2,08	79	2
7	6	- 5	- 1,73	83	4
8	6	- 4	- 1,39	86	8
9	7	- 3	- 1,04	90	16
10	12	- 2	- 0,69	93	24
11	13	- 1	- 0,35	96	34
12	14	0	0,00	100	50
13	10	1	0,35	104	66
14	12	2	0,69	107	76
15	10	3	1,04	110	84
16	13	4	1,39	114	92
17	5	5	1,73	117	96
18	2	6	2,08	121	98

0
0

**Tab. 56d-B 3.8.4: Normierung bei Normalverteilung Subtestteil Schrift-
sprache Nomen (arithm. Mittel = 11,9; Standardabw. s = 4,21)**

RW	f	x	z	Z	PR
0	0	- 12	-	-	0
1	0	- 11	-	-	0
2	0	- 10	- 2,38	76	1
3	0	- 9	- 2,14	79	2
4	4	- 8	- 1,90	81	3
5	3	- 7	- 1,66	83	4
6	6	- 6	- 1,43	86	8
7	6	- 5	- 1,19	88	12
8	9	- 4	- 0,95	90	16
9	9	- 3	- 0,71	93	24
10	5	- 2	- 0,48	95	31
11	11	- 1	- 0,24	98	42
12	7	0	0,00	100	50
13	5	1	0,24	102	58
14	7	2	0,48	105	69
15	8	3	0,71	107	76
16	8	4	0,95	110	84
17	11	5	1,19	112	88
18	11	6	1,43	114	92

Tab. 56e-B 3.8.4: Normierung bei Normalverteilung Subtestteil Schrift-**sprache Verben (arithm. Mittel = 9,8; Standardabw. s = 4,54)**

RW	f	x	z	Z	PR
0	1	- 10	- 2,21	78	1
1	1	- 9	- 1,98	80	2
2	5	- 8	- 1,76	82	3
3	3	- 7	- 1,54	85	7
4	6	- 6	- 1,32	87	10
5	7	- 5	- 1,10	89	13
6	9	- 4	- 0,88	91	18
7	6	- 3	- 0,66	93	24
8	4	- 2	- 0,44	95	31
9	6	- 1	- 0,22	98	42
10	5	0	0,00	100	50
11	9	1	0,22	102	58
12	9	2	0,44	104	66
13	10	3	0,66	107	76
14	7	4	0,88	109	82
15	7	5	1,10	111	86
16	10	6	1,32	113	90
17	2	7	1,54	115	93
18	1	8	1,76	118	96

Tab. 56f-B 3.8.4: Normierung bei Normalverteilung Subtestteil Schrift-**sprache Adjektive (arithm. Mittel = 10,1; Standardabw. s = 4,69)**

RW	f	x	z	Z	PR
0	0	- 10	- 2,13	79	2
1	0	- 9	- 1,92	81	3
2	2	- 8	- 1,70	83	4
3	4	- 7	- 1,49	85	7
4	6	- 6	- 1,28	87	10
5	8	- 5	- 1,07	89	13
6	11	- 4	- 0,85	91	18
7	8	- 3	- 0,64	94	27
8	9	- 2	- 0,43	96	34
9	5	- 1	- 0,21	98	42
10	6	0	0,00	100	50
11	8	1	0,21	102	58
12	8	2	0,43	104	66
13	2	3	0,64	106	73
14	3	4	0,85	109	82
15	6	5	1,07	111	86
16	5	6	1,28	113	90
17	13	7	1,49	115	93
18	4	8	1,70	117	96

Tab. 56g-B 3.8.4: Normierung bei Normalverteilung Subtestteil Lautsprache**Nomen (arithm. Mittel = 12,1; Standardabw. s = 3,86)**

RW	f	x	z	Z	PR
0	0	- 12	-	-	0
1	0	- 11	-	-	0
2	1	- 10	-	-	0
3	2	- 9	- 2,33	77	1
4	3	- 8	- 2,07	79	2
5	2	- 7	- 1,81	82	3
6	2	- 6	- 1,55	84	5
7	8	- 5	- 1,29	87	10
8	4	- 4	- 1,04	89	13
9	2	- 3	- 0,78	92	21
10	8	- 2	- 0,52	95	31
11	5	- 1	- 0,26	97	38
12	11	0	0,00	100	50
13	12	1	0,26	103	62
14	16	2	0,52	105	69
15	11	3	0,78	108	79
16	9	4	1,04	110	84
17	5	5	1,29	113	90
18	5	6	1,55	116	95

Tab. 56h-B 3.8.4: Normierung bei Normalverteilung Subtestteil Lautsprache**Verben (arithm. Mittel = 11,6; Standardabw. s = 3,66)**

RW	f	x	z	Z	PR
0	0	- 12	-	-	0
1	1	- 11	-	-	0
2	0	- 10	-	-	0
3	1	- 9	- 2,46	75	1
4	2	- 8	- 2,19	78	1
5	0	- 7	- 1,91	81	3
6	5	- 6	- 1,64	84	5
7	8	- 5	- 1,37	86	8
8	7	- 4	- 1,09	89	13
9	7	- 3	- 0,82	92	21
10	8	- 2	- 0,55	94	27
11	12	- 1	- 0,27	97	38
12	10	0	0,00	100	50
13	5	1	0,27	103	62
14	11	2	0,55	106	73
15	12	3	0,82	108	79
16	12	4	1,09	111	86
17	3	5	1,37	114	92
18	2	6	1,64	116	95

Tab. 56i-B 3.8.4: Normierung bei Normalverteilung Subtestteil Lautsprache					
Adjektive (arithm. Mittel = 10,5; Standardabw. s = 4,46)					
RW	f	x	z	Z	PR
0	1	- 11	- 2,47	75	1
1	0	- 10	- 2,24	78	1
2	1	- 9	- 2,02	80	2
3	4	- 8	- 1,80	82	3
4	4	- 7	- 1,57	84	5
5	5	- 6	- 1,35	86	8
6	9	- 5	- 1,12	89	13
7	5	- 4	- 0,90	91	18
8	12	- 3	- 0,67	93	24
9	5	- 2	- 0,45	95	31
10	10	- 1	- 0,22	98	42
11	5	0	0,00	100	50
12	3	1	0,22	102	58
13	9	2	0,45	105	69
14	7	3	0,67	107	76
15	6	4	0,90	109	82
16	10	5	1,12	111	86
17	8	6	1,35	114	92
18	2	7	1,57	116	95

Die Ergebnisse jeder einzelnen Testperson in den neun Subtestteilen bilden ein individuelles Leistungsprofil. Starke Schwankungen (in Größenordnung der kritischen Differenzen, s. B 3.9) in dem Profil einer Testperson sollten untersucht werden. Es muss dabei unterschieden werden zwischen Auffälligkeiten, die nur in einem einzelnen Subtestteil auftreten und solchen, die in mehreren Subtestteilen gleichzeitig deutlich werden. Besonderheiten in einzelnen Subtestteilen sollten nicht überbewertet werden, da sie durch zufällige Begebenheiten wie Konzentration oder Störungen während der Testsituation entstehen können und somit mehrdeutig sind. Gehäufte Auffälligkeiten, denen die gleiche Ursache zugrunde liegt, können dagegen möglicherweise auf besondere Probleme oder Fähigkeiten in bestimmten Modalitäten oder Wortarten hindeuten. Es sollte demnach im individuellen Profil untersucht werden, ob eine Testperson bei einer Teilgruppe von Subtestteilen, deren richtige Beantwortung die gleiche Fähigkeit voraussetzt, im Vergleich zu ihrer sonstigen Leistung besonders gut oder besonders schlecht abgeschnitten hat. Diese betreffenden Subtestteile sollten dann zusammenfassend unter Berücksichtigung ihrer Gemeinsamkeiten interpretiert werden. Folgende Subtestteilkombinationen bilden für diagnostische Zwecke sinnvolle Zusammenfassungen: Die drei Subtestteile einer Modalität (GSN, GSV, GSA), (SSN, SSV, SSA), (LSN, LSV, LSA) sowie die drei Subtestteile einer Wortart (GSN, SSN, LSN), (GSV, SSV, LSV), (GSA, SSA, LSA).

3.8.5 Normierung aufgrund modifizierter Fehlerbewertung

Zusätzlich zur regulären Normierung wurde für den Wortschatztest eine Normierung aufgrund einer modifizierten Fehlerbewertung vorgenommen. Dabei wurde davon ausgegangen, dass ein Teil der Testpersonen zwar einen Begriff vom Testitem hat, diese Kenntnis jedoch durch Schwierigkeiten in der Modalität nicht unter Beweis stellen konnte (vgl. C 1.2 und C 1.4). Demzufolge wurde in einer neuen Fehlerbewertung genau dann ein Punkt gegeben, wenn die Testperson das Item in Gebärden- oder in Schrift- oder in Lautsprache richtig beantworten konnte. Als falsch wurde die Aufgabe nur dann bewertet, wenn die Testperson sie durch alle drei Modalitäten hinweg falsch beantwortet hat. Daraus ergeben sich neue Rohwerte und demzufolge eine neue Normierung. Diese wurde über Flächentransformation vorgenommen, da keine Normalverteilung der Daten vorlag. Es ergeben sich nur drei Normierungstabellen, nämlich eine für Nomen, eine für Verben und eine für Adjektive, in denen sowohl Prozentränge als auch Z-Werte dargestellt sind:

Tab. 57a-B 3.8.5: Normierung aufgrund modifizierter Fehlerbewertung				
Nomen				
RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	0	0	0	-
7	0	0	0	-
8	0	0	0	-
9	1	1	0,4	73
10	0	1	0,9	74
11	2	3	1,8	78
12	4	7	4,5	84
13	6	13	8,9	86
14	8	21	15,2	90
15	8	29	22,3	92
16	14	43	32,1	95
17	24	67	49,1	100
18	45	112	79,9	108

Tab. 57b-B 3.8.5: Normierung aufgrund modifizierter Fehlerbewertung**Verben**

RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	0	0	0	-
7	0	0	0	-
8	0	0	0	-
9	0	0	0	-
10	2	2	0,9	74
11	1	3	2,2	79
12	0	3	2,7	80
13	2	5	3,6	82
14	7	12	7,6	86
15	8	20	14,3	89
16	26	46	29,5	95
17	27	73	53,1	101
18	39	112	82,6	109

Tab. 57c-B 3.8.5: Normierung aufgrund modifizierter Fehlerbewertung**Adjektive**

RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	0	0	0	-
7	0	0	0	-
8	0	0	0	-
9	2	2	0,9	74
10	4	6	3,6	82
11	6	12	8,0	86
12	3	15	12,1	88
13	6	21	16,1	90
14	12	33	24,1	93
15	13	46	35,3	96
16	14	60	47,3	99
17	19	79	62,1	103
18	33	112	98,7	121

3.9 Möglichkeiten zur Beurteilung von intra- und interindividuellen Leistungsunterschieden

3.9.1 Standardmessfehler

Wenn die Ergebnisse einer Testperson in verschiedenen Subtestteilen des Wortschatztests stark variieren, kann dies eine Folge des Messfehlers, also zufallsbedingt sein. Der Standardmessfehler gibt die Genauigkeit an, mit der aufgrund des Testergebnisses Aussagen über die wahre Merkmalsausprägung einer Testperson getroffen werden können. Mit einer Wahrscheinlichkeit von 2:3 liegt dieser wahre Wert einer Testperson in einem Bereich von einem Standardmessfehler ober- bzw. und unterhalb des gemessenen Wertes, mit einer Wahrscheinlichkeit von 19:20 liegt er in dem Streubereich von zwei Standardmessfehlern ober- bzw. und unterhalb des gemessenen Wertes. Der Standardmessfehler berechnet sich aus der Standardabweichung und dem Reliabilitätsindex einer Skala (vgl. Lienert/Raatz 1994, 365). Sein Ausmaß im Wortschatztest muss bekannt sein, um die Interpretation des Testergebnisses danach auszurichten:

Tab. 58-B 3.9.1: Standardmessfehler der Subtestteile			
Subtestteil	Standardabweichung	Reliabilität	Standardmessfehler
GSN	2,87	.79	1,31
GSV	2,21	.83	0,92
GSA	2,89	.79	1,32
SSN	4,21	.87	1,52
SSV	4,54	.86	1,69
SSA	4,69	.87	1,67
LSN	3,86	.85	1,49
LSV	3,66	.87	1,43
LSA	4,46	.88	1,58

Je geringer der Standardmessfehler einer Skala ausfällt, desto höher ist ihre Messgenauigkeit zu bewerten. Die Standardmessfehler des Wortschatztests weisen bei Werten zwischen 0,92 und 1,69 Punkten eine gute Messgenauigkeit der einzelnen Subtestteile aus. Aus diesem Standardmessfehler der Skalen lassen sich zudem für jeden einzelnen Kennwert die Vertrauensgrenzen (vgl. Lienert/Raatz 1994, 365) berechnen, in denen mit 5%iger Wahrscheinlichkeit der wahre Rohwert einer Testperson mit diesem gemessenen Rohwert liegt.

3.9.2 Kritische Differenz für die Beurteilung intraindividueller Unterschiede

Starke Leistungsschwankungen zwischen den Subtestteilen einer Testperson können auch überzufällig sein. Um diese Schwankungen genau analysieren zu können, muss jedes Subtestteilergebnis mit jedem anderen verglichen und müssen jeweils paarweise kritische Differenzen berechnet werden. Bei der Überschreitung dieser kritischen Differenz wird ein statistisch bedeutsamer, im Falle des Wortschatztests auf dem 5%-Niveau gesicherter Leistungsunterschied einer Testperson zwischen zwei Subtestteilen angezeigt. Erst wenn die Punktdifferenz über dem angegebenen Wert liegt, kann mit einer Irrtumswahrscheinlichkeit von 5% davon ausgegangen werden, dass die Testperson tatsächlich in einer Modalität oder in einer Wortart über eine bessere Wortschatzkenntnis verfügt als in der anderen.

Die Darstellung kritischer Differenzen schafft jedoch nur Scheingenauigkeit. Da die Reliabilität des Wortschatztests nicht auf einen exakten Wert festgelegt werden kann, werden auch die Werte der Differenzen schwanken. Zusätzlich verringert die Willkür in der Wahl des Signifikanzniveaus die Vergleichsmöglichkeit von Standardmessfehlern und kritischen Differenzen verschiedener Tests. Für den HAWIK-R beispielsweise werden Punktdifferenzen zwischen Skalen bereits als signifikant angesehen, wenn eine Zufallswahrscheinlichkeit auf dem 15%-Niveau gesichert wurde. Ein Ansetzen eines so geringen Niveaus hätte auch für den Wortschatztest niedrigere kritische Differenzen ergeben. Dennoch wurde in Anlehnung an Lienert/Raatz (vgl. 1994, 369) die kritische Differenz auf der 5%-Stufe gewählt.

Während die Tabellen zur mittleren Standardabweichung je zweier Subtestteile (Tab. 59-B 3.9.2) sowie zur Reliabilität je zweier Subtestteile (Tab. 60-B 3.9.2) im Anhang (Anh. 1: B 3.9.2a und Anh. 1: B 3.9.2b) einzusehen sind, werden die kritischen Differenzen je zweier Subtestteile im folgenden aufgeführt:

Tab. 61-B 3.9.2: Kritische Differenz je zweier Subtestteile für die Beurteilung intraindividueller Unterschiede									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	0	5,18	4,94	5,52	5,87	6,13	5,08	5,52	6,12
GSV		0	5,46	7,42	7,76	8,18	7,41	6,73	7,58
GSA			0	5,14	5,99	5,51	5,84	5,13	5,45
SSN				0	4,73	5,09	6,09	5,71	6,08
SSV					0	4,87	5,60	5,88	6,82
SSA						0	6,43	6,46	6,13
LSN							0	4,80	5,60
LSV								0	5,18
LSA									0

3.9.3 Kritische Differenz für die Beurteilung interindividueller Unterschiede

Auch für die Beurteilung interindividueller Unterschiede werden kritische Differenzen herangezogen. Bei ihrer Überschreitung wird ein statistisch bedeutsamer, auf dem 5%-Niveau gesicherter Leistungsunterschied zwischen zwei Testpersonen angezeigt. Erst wenn die Punktdifferenz über dem angegebenen Wert liegt, kann mit einer Irrtumswahrscheinlichkeit von 5% davon ausgegangen werden, dass die eine Testperson tatsächlich über eine bessere passive Wortschatzkenntnis verfügt als die andere.

In der folgenden Tabelle sind die Werte der Standardabweichung, der Reliabilität (nach Kri-
stof) und der kritischen Differenzen jedes Subtestteils sowie des Gesamttests abzulesen:

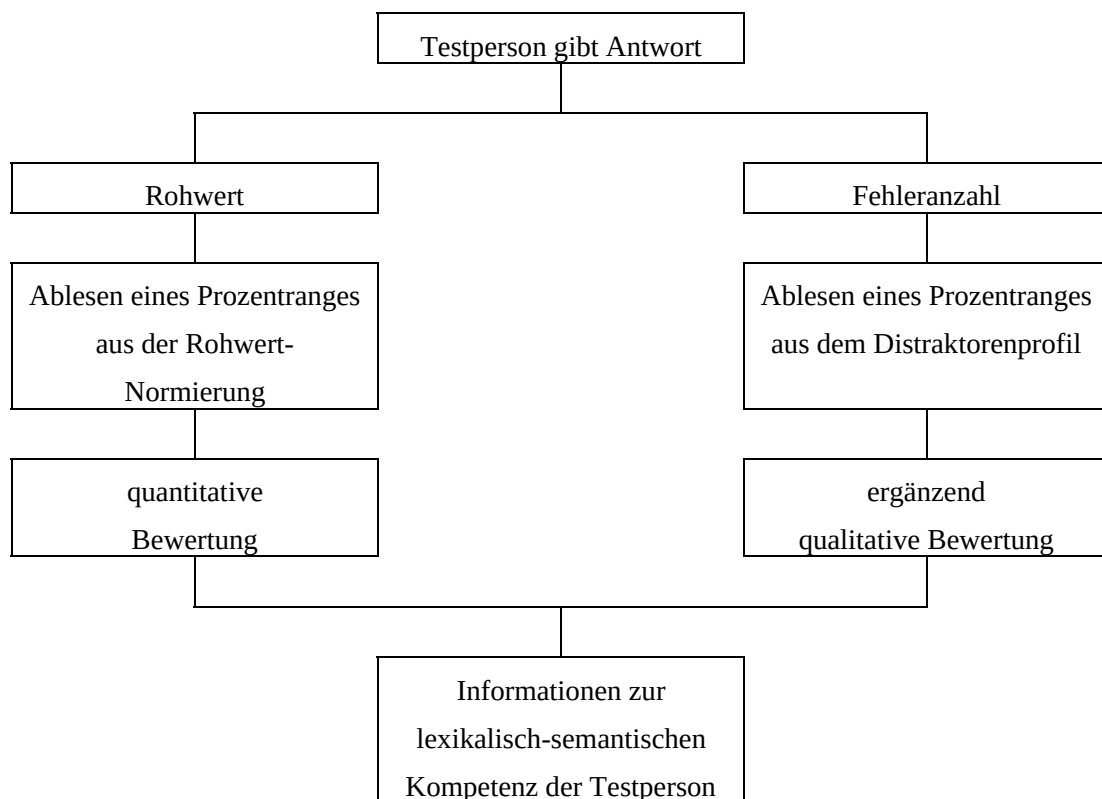
Tab. 62-B 3.9.3: Standardabweichung, Reliabilität und kritische Differenz für die Beurteilung interindividueller Unterschiede			
Subtestteil	Standard-abwei- chung	Reliabilität	kritische Differenz
GSN	2,87	.80	3,53
GSV	2,21	.80	2,75
GSA	2,89	.81	3,50
SSN	4,21	.87	4,16
SSV	4,54	.91	3,87
SSA	4,69	.86	4,90
LSN	3,86	.85	4,15
LSV	3,66	.84	4,06
LSA	4,46	.88	4,24
GESAMT	28,06	.85	30,32

3.10 Distraktorenprofil

Neben der normierten Rohwertskala stellt der Wortschatztest ergänzende qualitative Interpretationsmöglichkeiten der gemessenen Leistung bereit. Wie bereits bemerkt soll zusätzlich zu der ausschließlichen Betrachtung und Auszählung der richtigen Itemantworten auch der Wahl der Distraktoren bei falsch beantworteten Aufgaben Bedeutung zukommen. Da die Distraktoren der Items qualitativ abgestuft konzipiert sind, indem sie jeweils andere Ähnlichkeit zum Item aufweisen, soll ergänzend zur Auszählung des Gesamtpunktwertes die Art der Fehler untersucht werden. Es wird davon ausgegangen, dass damit Aufschluss über die Herangehensweise einer Testperson an sprachliche Anforderungen gegeben sowie Stärken oder Schwächen in einzelnen Modalitäten aufgezeigt werden können. Durch die differenzierte

Betrachtung eines individuellen Distraktorenprofils einer Testperson soll dem Anspruch der Differentialdiagnose genüge geleistet werden. Nur das Gesamtergebnis der quantitativen und der qualitativen Fehlerbewertung kann Grundlage für die differenzierte Bewertung einer Leistung sein. Angaben über die Homogenität oder Heterogenität des Distraktorenprofils können die Bedeutung des quantitativen Testergebnisses im Einzelfall relativieren. Die beiden parallel verlaufenden Wege der Informationsgewinnung über die sprachliche Kompetenz einer Testperson sind im folgenden Schaubild dargestellt:

Abb. 16-B 3.10: Schaubild zur Informationsgewinnung



Jedes Item weist in jeder Sprachmodalität Distraktoren auf, die bei entsprechender Vorgabe des Items plausibler sind als andere. Wählt eine Testperson beispielsweise in der Modalität Gebärdensprache bei Vorgabe eines gebärdeten Items einen gebärdensprachlichen Ablenker, so lässt dies darauf schließen, dass gebärdensprachlich orientierte Denkprozesse zu der spezifischen Fehlerwahl geführt haben könnten. Dabei wird davon ausgegangen, dass ein überdurchschnittlich häufiges Wählen einer Ablenkerart eine bestimmte Strategie bei der Aufgabenbearbeitung erkennen lässt. Entscheidet sich eine Testperson beispielsweise in allen Modalitäten überdurchschnittlich oft für den gebärdensprachlichen Distraktor, so kann dies bedeuten, dass sie sogar bei einer Itemvorgabe, die nicht in Gebärdensprache erfolgt, eine gedankliche Übersetzung in Gebärdensprache vornimmt, um die Vorgabe zu verstehen.

3.10.1 Überprüfung der qualitativen Distraktorenabstufung

Eine Überführung der qualitativen Abstufung der Distraktoren in eine quantitative Verteilung, bei der für die Wahl verschiedener Ablenker unterschiedlich viele Teilpunkte vergeben werden, wurde für den Wortschatztest verworfen. Ein derartiges Vorgehen stünde der Konzeption einer qualitativen Bewertung entgegen. Statt dessen wurde versucht, die bisher unbewiesenen Annahmen der qualitativen Abstufung mit Hilfe einer empirischen Untersuchung zu bestätigen. Diese Untersuchung soll die vermutete Qualität der einzelnen Ablenker in unterschiedlichen Modalitäten sichern, sodass durch eine zweigliedrige Bewertung mit quantitativem Rohwert und qualitativem Distraktorenprofil ein differenziertes Bild über die lexikalisch-semantische Kompetenz einer Testperson gewonnen werden kann.

In allen Subtestteilen existieren vier Ablenkerarten: Für jede Aufgabe wurde mindestens ein Inhaltsablenker (I) konstruiert. In einigen Fällen weist die Wahl dieser Ablenker auf eine elaborierte Sprachkenntnis (HRP) hin, da dieser Distraktor eine beinahe ebenso plausible Antwort wie das Item selber darstellt. Nicht alle Ablenker inhaltlicher (I) Ähnlichkeit entsprechen dieser Kategorie (HRP), sondern nur einige von ihnen; alle Ablenker dieser Art (HRP) sind jedoch gleichzeitig Ablenker mit einfacher inhaltlicher (I) Ähnlichkeit zum Item. Zusätzlich sollte jede Aufgabe mindestens einen Ablenker gebärdenähnlicher (G) und mundbildähnlicher (M) Art bereitstellen. Darüber hinaus wurde eine Kategorie der modalitätenentsprechenden Ablenkerart entwickelt. Diese Distraktorenart stellt in den Subtests Gebärden- bzw. Lautsprache ein Pendant zu gebärden- (G) bzw. mundbildähnlichen (M) Ablenkern dar, bildet in der Modalität Schriftsprache jedoch eine eigenständige Gruppe, die sich aus schriftbildähnlichen (S) Ablenkern zusammensetzt. Mundbild- (M) und schriftbildähnliche (S) Ablenker überschneiden sich häufig, aber nicht immer.

Es wird angenommen, dass die Wahl einiger plausibler Ablenker auf ein höheres Wortschatzniveau hindeutet als die Wahl anderer Ablenker, die in dieser Modalität keinen bedeutsamen Bezug zum Item aufweisen. So wird dem HRP-Ablenker sowie dem jeweilige Modalitäten-Ablenker eine höhere Plausibilität unterstellt als den jeweils anderen Distraktoren. Die hohe Plausibilität des HRP-Distraktoren ergibt sich aus seiner Konzeption, dem Item in inhaltlicher Hinsicht besonders ähnlich zu sein. Die Annahme der hohen Plausibilität der modalitätenähnlichen Distraktoren begründet sich in dem Umstand, dass eine Testperson bei der Wahl dieser Ablenker die angebotene Vorgabe des Items offensichtlich aufgenommen, jedoch nicht vollständig richtig auf die Zeichnung bzw. das geschriebene Wort angewendet hat. Eine Testperson mit einer derartigen Aufgabenbeantwortungs-Strategie hat damit im Hinblick auf das Erlernen und Beherrschen einer Sprachmodalität ein höheres Niveau erreicht als eine Testperson, die sich von modalitätenfremden Ähnlichkeiten oder von einfacher inhaltlicher Ähnlichkeit hat ablenken lassen.

Um diese Annahme anhand der vorliegenden Daten zu überprüfen, wurde die Hilfsannahme getroffen, dass diejenigen Testpersonen, die in der quantitativen Bewertung am besten abgeschnitten haben, auch mit großer Wahrscheinlichkeit bei falsch beantworteten Aufgaben nur knapp daneben liegen, also diese plausiblen Fehler machen, wenn sie überhaupt welche machen. Dagegen wird angenommen, dass die quantitativ schlechtesten Testpersonen unspezifisch oder sogar entgegen der mutmaßlichen qualitativen Abstufung Distraktoren wählen, also mit ihrer Antwort nicht nur knapp die richtige verfehlen.

Die folgenden Tabellen stellen die Fehlerantworten der 20 quantitativ besten (Max) und der fünf quantitativ schlechtesten (Min) Testpersonen der Analysestichprobe in je einer Modalität dar. (Da sich Testpersonen in verschiedenen Modalitäten als unterschiedlich leistungstark bzw. -schwach erwiesen, ergibt sich für die drei Modalitäten jeweils eine neue Zusammensetzung der besten und schlechtesten Testpersonengruppe.) Für beide Gruppen wird für jede Distraktorenart die Wahlhäufigkeit der Fehler (Anz.) angegeben. Dabei wird davon ausgegangen, dass ein Distraktorentyp wie beispielsweise der gebärdenähnliche Distraktor, der für jede Aufgabe existiert, von jeder Testperson in jedem Subtestteil maximal 18 Mal, in jeder Modalität maximal 54 Mal gewählt werden kann. Von den 20 besten Testpersonen kann diese Ablenkerart demnach $20 \times 54 = 1080$ Mal, von den fünf schlechtesten maximal $5 \times 54 = 270$ Mal gewählt werden. Einige Distraktorenarten existieren jedoch nicht für jedes Item, sodass sie potentiell seltener gewählt werden können. So setzt sich für den HRP-Distraktor die maximale Gesamtanzahl der Fehler in der Modalität Gebärdensprache aus $20 \times 33 = 660$ (Max) und $5 \times 33 = 165$ (Min) zusammen. Die Anzahl der Fehler in diesen beiden Distraktorenarten ist demnach nicht unmittelbar vergleichbar. Um die Relation der tatsächlichen Fehler zu der maximal möglichen Fehleranzahl einer Modalität zu berücksichtigen, wird neben der Fehleranzahl (Anz.) der prozentuale Anteil dieser Fehler an der Gesamtsumme der potentiell zu wählenden Fehler dieser Ablenkerart (%Pot) angegeben. In einer zweiten Spalte erfolgt zu jeder Distraktorenart eine Angabe darüber, wieviel Prozent (%) die Fehleranzahl dieser Distraktorenart von der Gesamtsumme der gemachten Fehler über alle Distraktorenarten hinweg ausmacht.

Tab. 63a-B 3.10.1: Gesamtfehlerverteilung der besten und schlechtesten Testpersonen im Subtest Gebärdensprache								
	G		HRP		M		I	
	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%
Max	58 / 5,4	30,9	49 / 7,4	42,3	29 / 2,7	15,4	22 / 2,0	11,4
Min	88 / 32,6	33,5	50 / 32,1	33,0	49 / 18,1	18,6	39 / 14,4	14,8

Tab. 63b-B 3.10.1: Gesamtfehlerverteilung der besten und schlechtesten Testpersonen im Subtest Schriftsprache										
	G		I		HRP		S		M	
	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%
Max	45/4,2	26,1	36/3,3	20,5	26/6,5	40,4	22/2,0	12,4	1/0,1	0,6
Min	79/29,3	22,7	95/35,2	27,3	32/32	24,8	65/24,1	18,7	23/8,5	6,6

Tab. 63c-B 3.10.1: Gesamtfehlerverteilung der besten und schlechtesten Testpersonen im Subtest Lautsprache								
	M		G		HRP		I	
	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%
Max	56 / 5,2	29,7	52 / 4,8	27,4	32 / 5,0	28,6	27 / 2,5	14,3
Min	102 / 37,8	30,1	90 / 33,3	22,9	36 / 22,5	17,9	86 / 31,9	25,4

Die Tabellen bestätigen die Annahme, dass der HRP-Ablenker für die quantitativ besten Testpersonen in allen Modalitäten tatsächlich den plausibelsten, da meistgewählten Ablenkertyp darstellte. Die auf die Wahl dieses Ablenkens entfallende Prozentangabe der quantitativ schlechtesten Testpersonen unterschreitet die der besten dagegen deutlich, nämlich um 9,3% (GS), 15,6% (SS) und 10,7% (LS). Die Plausibilität der modalitätenähnlichen Distraktoren bestätigt sich für die gute Testpersonengruppe nur für den Gebärdensprache- sowie für den Lautspracheteil. Erwartungsgemäß werden die jeweils nicht so plausiblen Distraktorenarten von der schlechten Testpersonengruppe deutlich häufiger gewählt als von der guten. Auffallend häufig wurde in allen Modalitäten von den quantitativ besten Testpersonen der gebärdenspracheähnliche Distraktor gewählt. Dieser Umstand unterläuft in gewisser Weise die Annahme, dass neben dem HRP-Distraktor in der Schriftsprache vor allem der schriftbild- und in der Lautsprache vor allem der mundbildähnliche Distraktor plausible Alternativantworten darstellen. Er lässt sich jedoch möglicherweise mit dem Umstand rechtfertigen, dass fast alle

Testpersonen in der Gebärdensprache die besten Leistungen erzielen und somit auf diese Modalität als die ihnen vertrauteste fixiert sind.

Eine zusätzliche Schwierigkeit bei der Interpretation der Ergebnisse entstand durch die Tatsache, dass ein Distraktor oft mehrere Ähnlichkeitsarten aufweist und deshalb nicht nachvollziehbar ist, aufgrund welcher Ähnlichkeit zum Item er gewählt wurde. Soll die obige Hilfsannahme als angenommen gelten, so kann zusätzlich unterstellt werden, dass die guten Testpersonen einen Ablenker aufgrund seiner plausiblen, schlechte Testpersonen ihn aufgrund seiner weniger plausiblen Ähnlichkeit zum Item gewählt haben. Unter dieser Voraussetzung kommt man für die drei Modalitäten zu folgenden Ergebnissen:

Tab. 64a-B 3.10.1: Gesamtfehlerverteilung Gebärdensprache – ausgewählte Ablenkerarten entsprechend dem quantitativen Testergebnis								
	G		HRP		M		I	
	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%
Max 20	58 / 5,4	36,7	49 / 7,4	50,3	12 / 1,1	7,5	9 / 0,8	5,4
Min 5	49 / 18,1	30,3	15 / 9,1	15,2	40 / 14,8	24,7	48 / 17,8	29,8

Tab. 64b-B 3.10.1: Gesamtfehlerverteilung Schriftsprache – ausgewählte Ablenkerarten entsprechend dem quantitativen Testergebnis										
	G		I		HRP		S		M	
	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%
Max 20	22 / 2,0	13,2	14 / 1,3	8,6	39 / 9,8	64,5	22 / 2,0	13,2	1 / 0,1	0,7
Min 5	79 / 29,3	26,9	95 / 35,2	32,4	21 / 21,0	19,3	40 / 14,8	13,6	23 / 8,5	7,8

Tab. 64c-B 3.10.1: Gesamtfehlerverteilung Lautsprache – ausgewählte Ablenkerarten entsprechend dem quantitativen Testergebnis								
	M		G		HRP		I	
	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%	Anz./ %Pot	%
Max 20	56 / 5,2	40,3	13 / 1,2	9,3	32 / 5,0	38,8	16 / 1,5	11,6
Min 5	60 / 22,2	24,3	90 / 33,3	36,5	6 / 3,8	4,2	86 / 31,9	35,0

Mit Hilfe der zusätzlichen Hilfsannahme bestätigen diese Tabellen die angenommene qualitative Abstufung der Distraktoren deutlicher, als es ohne die Annahmen in den vorhergegangenen Tabellen der Fall war.

Man kann demnach vorläufig (vergl. C. 5 Ausblick) davon ausgehen, dass sich die höhere Plausibilität einiger Ablenker gegenüber anderen bewahrheitet hat. Auf dieser Grundlage kann man mit Hilfe eines individuellen Distraktorenprofils ergänzend zum quantitativen Testergebnis einer Testperson differenziertere Aussagen zum Niveau ihrer Sprachkompetenz erhalten. Zudem kann ein derartiges Ergebnis als Stütze der Konstruktvalidität (vgl. B 3.7.3) des Tests gewertet werden.

3.10.2 Distraktoren-Normierungstabellen

Für die Erstellung eines individuellen Distraktorenprofils werden Normierungen der Ablenker-Wahlhäufigkeiten benötigt, um die Fehlerwahl einer Testperson mit denen vergleichbarer Testpersonen in Beziehung setzen zu können. Die Berechnung der Prozentränge für einzelne Ablenkerarten erfolgte dabei über Flächentransformation, da Normalverteilung nicht nachgewiesen wurde. Die Prozentrangwerte müssen in diesem Zusammenhang besonders interpretiert werden. Der Prozentrang einer Testperson in einer Ablenkerart in einem Subtestteil gibt nach Definition den Prozentanteil aller vergleichbaren getesteten Testpersonen an, den diese Person mit ihrer 'Leistung' übertroffen hat. Da es sich bei der gemessenen Größe in diesem Fall aber nicht um eine 'Leistung', sondern um die Anzahl der gemachten Fehler handelt, bedeutet ein höherer Rohwert (viele Fehler) eine schlechtere Leistung als ein niedriger Rohwert (wenige Fehler). Dementsprechend repräsentiert der Prozentrang einer Testperson in einer Ablenkerart in einem Subtestteil den Prozentanteil aller vergleichbaren Testpersonen, der weniger Fehler gemacht hat, als diese Testperson. Ein hoher Prozentrang bedeutet also, dass viele Testpersonen besser, ein niedriger Prozentrang, dass wenige Testpersonen besser waren als diese Testperson.

Mit Hilfe der Prozentränge dieser Normierungstabellen lässt sich feststellen, ob eine Testperson mit ihrer fehlerhaften Aufgabenbeantwortung im Durchschnitt der Normierungsstichprobe liegt, also durchschnittlich oft oder auffallend selten oder häufig bestimmte Distraktoren wählt. Dieser Bezug zur Vergleichsstichprobe ist Grundlage für die Untersuchung, ob eine Testperson im Vergleich zu ihren sonstigen Leistungen in einzelnen Subtestteilen auffallend gut oder auffallend schlecht abschneidet. So werden für eine Testperson zusätzlich zu dem Wert aus der Normierungstabelle im Rahmen der klassischen quantitativen Bewertung vier (Gebärden- und Lautsprache) bzw. fünf (Schriftsprache) Werte aus folgenden Tabellen abgelesen.

In Klammern unter der jeweiligen Ablenkerart ist die Gesamtanzahl der entsprechenden Fehler innerhalb aller Distraktoren des Subtestteils angegeben. Da mehrere Distraktoren eines einzelnen Items die gleiche Ablenkerart aufweisen können, also z. B. das Item ‘Angst’ zwei gebärdenähnliche und drei inhaltsähnliche Ablenker besitzt, kann die Angabe in Klammern die Aufgabenanzahl von 18 Items pro Subtestteil überschreiten. Trotzdem kann eine Testperson jede Ablenkerart nur maximal 18 Mal wählen, da sie sich pro Aufgabe nur für eine Antwort entscheiden kann. Aus diesem Grund bestehen die folgenden Tabellen aus jeweils 18 und nicht der Anzahl der vorkommenden Ablenkerarten entsprechend vielen Zeilen.

Die Normierung der Distraktorenwahl wurde für jeden Subtestteil gesondert vorgenommen. Die Tabellen zu Gebärdens- und Lautsprache enthalten jeweils vier Distraktorenspalten, da in diesen Fällen der Distraktorentyp des modalitätenähnlichen Ablenkers mit dem des gebärden- bzw. mundbildähnlichen Ablenkers übereinstimmt. In der Modalität Schriftsprache wird für die Ablenkerart mit Modalitätenähnlichkeit zum Item eine gesonderte Spalte mit schriftbildähnlichen Distraktoren aufgeführt, die Tabellen zur Schriftsprache weisen aus diesem Grund fünf Distraktorenspalten auf.

Tab. 65a-B 3.10.2: Distraktoren-Normierung Subtestteil Gebärdensprache								
Nomen								
Fehler- anzahl	HRP (16)		G = MOD (24)		M (24)		I (29)	
	f	PR	f	PR	f	PR	f	PR
0	8	4	13	6	15	7	6	3
1	27	20	19	21	35	30	17	14
2	28	45	21	39	20	55	20	30
3	22	68	18	57	19	73	20	49
4	17	85	18	73	11	86	21	67
5	7	96	11	86	5	94	14	83
6	1	100	5	94	5	98	8	93
7	0	100	3	97	0	98	3	98
8	0	100	1	99	0	98	1	100
9	0	100	1	100	0	98	0	100
10	0	100	0	100	0	98	0	100
11	0	100	0	100	0	98	0	100
12	0	100	0	100	0	98	0	100
13	0	100	0	100	0	98	0	100
14	0	100	0	100	0	98	0	100
15	0	100	0	100	0	98	0	100
16	0	100	0	100	0	98	0	100
17	-		0	100	0	98	0	100
18	-		0	100	0	98	0	100

Tab. 65b-B 3.10.2: Distraktoren-Normierung Subtestteil Gebärdensprache
Verben

Fehler- anzahl	HRP (11)		G = MOD (23)		M (27)		I (26)	
	f	PR	f	PR	f	PR	f	PR
0	31	15	16	8	32	15	16	8
1	48	50	36	31	58	56	40	33
2	23	83	26	60	12	88	28	64
3	5	95	18	80	5	95	12	82
4	2	99	5	90	2	99	8	91
5	1	100	7	95	0	99	2	96
6	0	100	2	100	1	100	3	98
7	0	100	0	100	0	100	0	98
8	0	100	0	100	0	100	1	100
9	0	100	0	100	0	100	0	100
10	0	100	0	100	0	100	0	100
11	0	100	0	100	0	100	0	100
12	-		0	100	0	100	0	100
13	-		0	100	0	100	0	100
14	-		0	100	0	100	0	100
15	-		0	100	0	100	0	100
16	-		0	100	0	100	0	100
17	-		0	100	0	100	0	100
18			0	100	0	100	0	100

Tab. 65c-B 3.10.2: Distraktoren-Normierung Subtestteil Gebärdensprache
Adjektive

Fehler- anzahl	HRP (6)		G = MOD (22)		M (26)		I (35)	
	f	PR	f	PR	f	PR	f	PR
0	6	3	3	2	27	13	2	1
1	37	23	11	8	28	38	13	8
2	39	57	18	21	31	65	11	19
3	22	85	17	37	18	87	21	34
4	5	97	9	49	5	97	14	50
5	1	100	21	63	1	100	21	65
6	0	100	16	80	0	100	9	79
7	-		8	90	0	100	10	88
8	-		5	96	0	100	6	95
9	-		2	100	0	100	3	99
10	-		0	100	0	100	0	99
11	-		0	100	0	100	0	99
12	-		0	100	0	100	0	99
13	-		0	100	0	100	0	99
14	-		0	100	0	100	0	99
15	-		0	100	0	100	0	99
16	-		0	100	0	100	0	99
17	-		0	100	0	100	0	99
18			0	100	0	100	0	99

Tab. 65d-B 3.10.2: Distraktoren-Normierung Subtestteil Schriftsprache

Nomen										
Fehler- anzahl	HRP (9)		S = MOD (14)		G (24)		M (24)		I (29)	
	f	PR	f	PR	f	PR	f	PR	f	PR
0	25	12	39	18	23	11	34	16	12	6
1	24	34	32	50	21	31	20	40	11	16
2	26	57	18	73	18	49	14	56	13	27
3	23	79	9	85	20	66	13	68	15	40
4	8	93	7	93	15	82	13	80	12	52
5	4	99	4	98	7	92	6	89	15	65
6	0	99	1	100	4	97	7	95	12	77
7	0	99	0	100	2	100	2	99	12	88
8	0	99	0	100	0	100	0	99	5	95
9	0	99	0	100	0	100	1	100	3	99
10	-		0	100	0	100	0	100	0	99
11	-		0	100	0	100	0	100	0	99
12	-		0	100	0	100	0	100	0	99
13	-		0	100	0	100	0	100	0	99
14	-		0	100	0	100	0	100	0	99
15	-		-		0	100	0	100	0	99
16	-		-		0	100	0	100	0	99
17	-		-		0	100	0	100	0	99
18	-		-		0	100	0	100	0	99

Tab. 65e-B 3.10.2: Distraktoren-Normierung Subtestteil Schriftsprache

Verben										
Fehler- anzahl	HRP (8)		S = MOD (26)		G (23)		M (28)		I (28)	
	f	PR	f	PR	f	PR	f	PR	f	PR
0	9	5	9	5	3	2	9	5	1	1
1	38	26	17	17	6	6	15	16	6	4
2	35	60	12	30	21	19	13	29	18	15
3	21	86	18	44	15	35	16	42	11	29
4	5	98	13	58	17	50	8	53	13	40
5	4	99	15	71	15	65	19	66	14	52
6	0	99	12	84	15	79	12	80	11	64
7	0	99	6	92	10	90	8	89	14	75
8	0	99	4	97	5	97	6	96	8	86
9	-		1	99	1	100	0	96	5	92
10	-		1	100	0	100	1	99	5	96
11	-		0	100	0	100	1	100	1	99
12	-		0	100	0	100	0	100	1	100
13	-		0	100	0	100	0	100	0	100
14	-		0	100	0	100	0	100	0	100
15	-		0	100	0	100	0	100	0	100
16	-		0	100	0	100	0	100	0	100
17	-		0	100	0	100	0	100	0	100
18	-		0	100	0	100	0	100	0	100

Tab. 65f-B 3.10.2: Distraktoren-Normierung Subtestteil Schriftsprache

Adjektive										
Fehler- anzahl	HRP (3)		S = MOD (18)		G (22)		M (27)		I (35)	
	f	PR	f	PR	f	PR	f	PR	f	PR
0	73	34	24	12	19	9	21	10	6	3
1	27	81	19	31	13	24	16	27	16	13
2	5	95	17	48	11	35	7	38	5	23
3	3	99	14	63	18	49	7	44	9	30
4	-		11	74	17	65	9	52	7	37
5	-		17	87	9	77	12	62	14	47
6	-		2	96	13	87	13	73	15	60
7	-		3	98	5	95	13	85	9	71
8	-		1	100	2	99	4	93	10	80
9	-		0	100	1	100	3	96	11	90
10	-		0	100	0	100	3	99	4	97
11	-		0	100	0	100	1	100	1	99
12	-		0	100	0	100	0	100	1	100
13	-		0	100	0	100	0	100	0	100
14	-		0	100	0	100	0	100	0	100
15	-		0	100	0	100	0	100	0	100
16	-		0	100	0	100	0	100	0	100
17	-		0	100	0	100	0	100	0	100
18	-		0	100	0	100	0	100	0	100

Tab. 65g-B 3.10.2: Distraktoren-Normierung Subtestteil Lautsprache Nomen

Fehler- anzahl	HRP (16)		G (24)		M = MOD (24)		I (29)	
	f	PR	f	PR	f	PR	f	PR
0	20	10	18	9	10	5	11	6
1	34	35	25	29	11	15	26	23
2	22	62	22	51	26	33	18	44
3	18	81	15	69	20	54	16	60
4	7	92	8	80	13	70	14	74
5	3	97	5	86	11	81	7	84
6	2	100	6	91	7	90	4	89
7	0	100	4	96	3	94	3	92
8	0	100	3	99	2	97	4	96
9	0	100	1	100	3	99	2	99
10	0	100	0	100	0	99	1	100
11	0	100	0	100	0	99	0	100
12	0	100	0	100	0	99	0	100
13	0	100	0	100	0	99	0	100
14	0	100	0	100	0	99	0	100
15	0	100	0	100	0	99	0	100
16	0	100	0	100	0	99	0	100
17	-		0	100	0	99	0	100
18	-		0	100	0	99	0	100

Tab. 65h-B 3.10.2: Distraktoren-Normierung Subtestteil Lautsprache Verben								
Fehler- anzahl	HRP (10)		G (23)		M = MOD (27)		I (26)	
	f	PR	f	PR	f	PR	f	PR
0	21	10	18	9	6	3	6	3
1	42	40	25	29	21	16	27	19
2	19	69	22	51	15	33	13	38
3	18	86	15	69	18	49	9	48
4	5	97	8	80	14	64	17	60
5	1	100	5	86	13	76	12	74
6	0	100	6	91	10	87	7	83
7	0	100	4	96	4	94	12	92
8	0	100	3	99	4	98	2	99
9	0	100	0	99	0	98	1	100
10	0	100	0	99	0	98	0	100
11	-		0	99	0	98	0	100
12	-		0	99	0	98	0	100
13	-		0	99	0	98	0	100
14	-		0	99	0	98	0	100
15	-		0	99	0	98	0	100
16	-		0	99	0	98	0	100
17	-		0	99	0	98	0	100
18	-		0	99	0	98	0	100

Tab. 65i-B 3.10.2: Distraktoren-Normierung Subtestteil Lautsprache Adjektive								
Fehler- anzahl	HRP (6)		G (22)		M = MOD (26)		I (35)	
	f	PR	f	PR	f	PR	f	PR
0	23	11	10	5	13	7	3	2
1	30	36	17	18	11	18	15	10
2	28	64	10	31	16	31	14	24
3	19	86	19	44	16	46	10	35
4	5	97	13	59	12	59	7	43
5	1	100	15	73	15	72	12	52
6	0	100	11	85	9	83	9	62
7	-		7	93	7	91	11	72
8	-		3	98	5	96	10	82
9	-		1	100	1	99	8	90
10	-		0	100	1	100	6	97
11	-		0	100	0	100	1	99
12	-		0	100	0	100	1	100
13	-		0	100	0	100	0	100
14	-		0	100	0	100	0	100
15	-		0	100	0	100	0	100
16	-		0	100	0	100	0	100
17	-		0	100	0	100	0	100
18	-		0	100	0	100	0	100

3.10.3 Standardmessfehler und kritische Differenzen im Distraktorenprofil

Auch für das Distraktorenprofil soll mit Hilfe des Standardmessfehlers die Genauigkeit festgelegt werden, mit der aufgrund des Profilergebnisses Aussagen über die wahre Merkmalsausprägung einer Testperson getroffen werden können. Zusätzlich wurden auch in diesem Zusammenhang kritische Differenzen zur Beurteilung von Leistungsunterschieden mehrerer Testpersonen in demselben Subtestteil berechnet. (Die im Anhang Anh. 1: B 3.10.3a und Anh. 1: B 3.10.3b aufgeführten Tabellen Tab. 66a-e-B 3.10.3 und Tab. 67a-e-B 3.10.3 geben Standardabweichung und Reliabilität je zweier Subtestteile wieder, die zur Berechnung des Standardmessfehlers sowie der kritischen Differenzen benötigt werden.) Die folgenden Tabellen stellen für jeden Distraktorentyp Standardabweichung, Reliabilität, Standardmessfehler und kritische Differenzen in allen Subtestteilen dar:

Tab. 68a-B 3.10.3: Standardmessfehler und kritische Differenzen HRP				
Subtestteil	Standard- abweichung	Reliabilität	Standard- messfehler	Kritische Differenz
GSN	1,40	.15	1,30	3,59
GSV	0,99	.28	0,84	2,32
GSA	1,01	.19	0,90	2,50
SSN	1,39	.40	1,07	2,97
SSV	1,01	-.15	1,08	1,53
SSA	0,71	.29	0,60	1,66
LSN	1,43	.29	1,20	3,33
LSV	1,18	.17	1,08	2,98
LSA	1,20	.29	1,01	2,80

Tab. 68b-B 3.10.3: Standardmessfehler und kritische Differenzen MOD				
Subtestteil	Standard- abweichung	Reliabilität	Standard- messfehler	Kritische Differenz
GSN	1,98	.52	1,37	3,81
GSV	1,46	.47	1,07	2,96
GSA	2,19	.57	1,43	3,97
SSN	1,46	.49	1,04	2,88
SSV	2,34	.58	1,52	4,22
SSA	2,04	.57	1,34	3,70
LSN	2,15	.51	1,50	4,15
LSV	2,23	.55	1,50	4,15
LSA	2,43	.62	1,50	4,15

Tab. 68c-B 3.10.3: Standardmessfehler und kritische Differenzen G				
Subtestteil	Standard-abweichung	Reliabilität	Standard-messfehler	Kritische Differenz
GSN	1,98	.52	1,37	3,81
GSV	1,46	.47	1,07	2,96
GSA	2,19	.57	1,43	3,97
SSN	1,82	.51	1,28	3,55
SSV	2,10	.42	1,59	4,41
SSA	2,30	.59	1,47	4,08
LSN	2,12	.59	1,36	3,78
LSV	1,70	.31	1,41	3,91
LSA	2,27	.56	1,50	4,16

Tab. 68d-B 3.10.3: Standardmessfehler und kritische Differenzen M				
Subtestteil	Standard-abweichung	Reliabilität	Standard-messfehler	Kritische Differenz
GSN	1,61	.43	1,22	3,37
GSV	0,99	.45	0,74	2,05
GSA	1,21	.28	1,03	2,85
SSN	2,11	.63	1,29	3,57
SSV	2,49	.62	1,53	4,23
SSA	2,96	.73	1,53	4,23
LSN	2,15	.51	1,50	4,15
LSV	2,23	.55	1,50	4,15
LSA	2,43	.62	1,50	4,15

Tab. 68e-B 3.10.3: Standardmessfehler und kritische Differenzen I				
Subtestteil	Standard-abweichung	Reliabilität	Standard-messfehler	Kritische Differenz
GSN	1,83	.34	1,49	4,12
GSV	1,50	.48	1,08	3,00
GSA	2,20	.54	1,50	4,16
SSN	2,50	.63	1,53	4,23
SSV	2,68	.60	1,69	4,68
SSA	3,05	.69	1,70	4,71
LSN	2,32	.62	1,43	3,95
LSV	2,32	.56	1,54	4,26
LSA	3,04	.70	1,68	4,65

Auch im Distraktorenprofil wurden kritische Differenzen zur Beurteilung intraindividuel­ler Unterschiede einer Testperson in verschiedenen Subtestteilen berechnet. Die folgen­den Tabellen geben die Differenzwerte aller Subtestteilkombinationen zu jedem Distrakto­rentyp wieder.

Tab. 69a-B 3.10.3: Kritische Differenz je zweier Subtestteile HRP

	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	0	2,80	2,82	3,37	3,09	2,86	3,27	3,35	3,15
GSV		0	2,45	2,92	2,33	2,35	3,21	2,83	2,82
GSA			0	2,67	2,32	2,12	3,07	2,68	2,49
SSN				0	2,63	2,50	2,99	3,23	2,92
SSV					0	2,13	2,99	2,88	2,57
SSA						0	2,60	2,48	2,62
LSN							0	2,95	2,93
LSV								0	3,07
LSA									0

Tab. 69b-B 3.10.3: Kritische Differenz je zweier Subtestteile MOD

	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	0	3,54	4,26	3,61	4,45	3,99	4,05	4,17	3,79
GSV		0	4,27	3,62	4,21	4,40	4,46	4,38	4,52
GSA			0	3,23	4,01	3,45	4,67	3,83	3,93
SSN				0	3,60	3,06	4,06	3,54	3,53
SSV					0	3,90	4,28	4,50	4,24
SSA						0	4,48	4,00	4,19
LSN							0	3,99	3,70
LSV								0	3,90
LSA									0

Tab. 69c-B 3.10.3: Kritische Differenz je zweier Subtestteile G

	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	0	3,54	4,26	3,61	3,96	4,47	3,60	4,41	4,54
GSV		0	4,27	3,76	3,75	4,78	4,09	3,97	4,65
GSA			0	3,58	3,82	4,13	3,98	4,56	3,78
SSN				0	3,61	3,76	3,95	3,84	3,96
SSV					0	3,90	4,23	4,13	4,97
SSA						0	4,56	4,60	4,65
LSN							0	4,28	4,29
LSV								0	4,59
LSA									0

Tab. 69d-B 3.10.3: Kritische Differenz je zweier Subtestteile M

	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	0	2,95	3,16	3,61	3,93	4,44	3,70	3,94	4,17
GSV		0	2,60	3,96	4,21	4,82	4,27	3,96	4,24
GSA			0	3,49	3,97	4,00	3,59	3,26	3,36
SSN				0	3,67	3,93	4,47	4,18	3,95
SSV					0	4,53	4,29	4,64	4,26
SSA						0	5,50	4,68	4,71
LSN							0	3,99	3,70
LSV								0	3,90
LSA									0

Tab. 69e-B 3.10.3: Kritische Differenz je zweier Subtestteile I									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	0	5,54	3,90	4,36	4,40	4,66	4,19	4,32	5,02
GSV		0	4,19	4,39	4,53	4,85	4,86	4,46	5,23
GSA			0	4,23	4,37	4,61	4,90	4,18	4,55
SSN				0	3,75	4,15	4,38	4,25	4,62
SSV					0	4,17	4,41	4,43	5,33
SSA						0	4,59	5,16	5,09
LSN							0	4,09	4,53
LSV								0	4,40
LSA									0

C. Auswertung der Untersuchungsergebnisse

1. Überprüfung pädagogischer Fragestellungen

Im Rahmen der Überprüfung pädagogischer Fragestellungen ist zu zeigen, inwiefern die vorliegenden Daten eine Grundlage bilden für die differenzierte Einschätzung der Kompetenzen gehörloser Dritt- und Viertklässler sowie für die Bewertung gehörlosenpädagogischer Methodik. Diese Auswertung der Daten wird in zwei Etappen vorgenommen. So werden zunächst die aus der Pilotstudie hervorgegangenen Daten der Erhebung mit den Hamburger bilingualen sowie den hörenden Testpersonen bearbeitet (C 1.1 und C 1.2). Teile dieser Bearbeitung wurden bereits im Zwischenbericht zum Hamburger bilingualen Schulversuch vorgelegt (Günther 1999 a, 167-173) und werden an dieser Stelle, z. T. modifiziert und erweitert, wiederholt. In einem weiteren Schritt werden die Daten der Hauptuntersuchung ausgewertet (C 1.3 ff). Diese Ergebnisse werden teilweise auf die der Pilotstudie bezogen. Aus diesem Grund ist deren Wiederholung notwendig. Die zentrale Stellung im Rahmen der vorliegenden Arbeit kommt der Auswertung der Hauptuntersuchung zu.

1.1 Schwierigkeitsgefälle der Modalitäten im Rahmen der Pilotstudie

Angenommen wird, dass die gehörlosen Testpersonen in der Modalität Gebärdensprache die vergleichsweise besten Ergebnisse erreichen würden, während ihnen die Lautsprache am schwersten fallen und die Schriftsprache sich im Mittelfeld bewegen würde. Den Hintergrund dieser Annahme bildet der Umstand, dass die Gebärdensprache Gehörlosen durch ihren visuellen Charakter am leichtesten und vollständigsten zugänglich ist. Im Folgenden wird geprüft, ob sich dieses Schwierigkeitsgefälle unter den sprachlichen Modalitäten bzw. Subtests bestätigt. Neben der Überprüfung dieses Gefälles wird die Annahme, dass die Gebärdensprache die Leistungsspitze darstellt, besonders fokussiert.

Die in den Leistungs-Tabellen in B 2.3.1, B 2.3.3 und B 2.4.1 (Tab. 7a/b-B 2.3.1, Tab. 9a/b-B 2.3.3 und Tab. 11-B 2.4.1) zusammengefassten Ergebnisse bestätigen die Annahme, dass die gehörlosen, bilingual erzogenen Testpersonen in der Gebärdensprache vergleichsweise zu den anderen beiden Modalitäten gut abschneiden. Im Schriftsprachtest kamen insgesamt etwas schlechtere, zum Teil deutlich schlechtere Ergebnisse zustande. Die größten Schwierigkeiten bereitete erwartungsgemäß der lautsprachliche Subtest. Insgesamt müssen die Ergebnisse bei den verheerenden Prognosen, die im Allgemeinen über den Wortschatz Gehörloser vorgenommen werden, als erstaunlich gut eingeschätzt werden.

Die erreichten Scores bewegen sich zwischen höchstens 100% und niedrigstens 11%. Dabei stellt das einmalig auftretende Ergebnis von 11% das mit einigem Abstand niedrigste Ergebnis dar, welches nur bedingt aussagekräftig ist, weil die betreffende Testperson deutliche Verhaltensauffälligkeiten in der Testsituation zeigte und Konzentrationsschwierigkeiten aufwies. Das zweitschlechteste Ergebnis von 17% wurde insgesamt sechsmal erreicht. In drei dieser Fälle entfällt das Ergebnis von 17% pro Subtest und Wortart auf die Testperson Orhan. Das Kind verfügte bei Eintritt in die Schulversuchs-Klasse über keinerlei Sprache und auch seine Eltern verfügen über keine Gebärdensprachkenntnisse. Er ist demnach gegenüber seinen Mitschülern, welche mehrheitlich mit LBG frühgefördert wurden und schon im Kindergarten DGS angeboten bekamen, deutlich benachteiligt. Dies relativiert den größeren Anteil dieser sehr schlechten Ergebnisse.

Die Gesamtergebnisse beider Klassen ebnen zwar die extremen Werte etwas ein, spiegeln jedoch trotzdem das benannte Schwierigkeitsgefälle der Modalitäten. Dabei ist der Einebnungseffekt maßgeblich auf die beschriebenen schwachen Schüler, die statistisch sog. Ausreißer darstellen, zurückzuführen und relativiert das Ergebnis in gleicher Weise wie bei den Betrachtungen der Einzelergebnisse.

Festzuhalten ist, dass die Kompetenzen der bilingualen gehörlosen Testpersonen in der Modalität Gebärdensprache am stärksten sind. Es wurde bereits erläutert aus welchen Gründen die Gebärdensprache für diese Testpersonen als Mutter- bzw. Basissprache fungiert. Auf diesem Hintergrund stellt sich anhand der Daten heraus, dass diese ihnen am leichtesten und natürlichsten zugängliche Modalität ihr Sprachverständnis am vollständigsten abbildet. Die gleiche vollständige Zugänglichkeit gilt für Hörende in ihrer Muttersprache, der deutschen Lautsprache. Daher drängt sich ein Vergleich der Ergebnisse Hörender im Subtest Lautsprache mit den Ergebnissen bilingualer Gehörloser im Subtest Gebärdensprache auf, der im Folgenden vorgenommen wird.

1.2 Vergleich der Kompetenzen mit neuer Fehlerbewertung im Rahmen der Pilotstudie

Es wird angenommen, dass die Testpersonen vielfach zwar über die mentale Repräsentation eines Items verfügen, dies bei einseitiger Abfrage in ausschließlich lautsprachlicher Modalität jedoch nicht unter Beweis stellen können. Wenn eine gehörlose Testperson ein Item in einer der Modalitäten nicht in Zielrichtung beantwortet, so lässt dies noch keinen Rückschluss darauf zu, dass sie keinen Begriff von ihm hat, sondern legitimiert zunächst nur die Feststellung, dass sie ihre mögliche Kenntnis des Begriffs in dieser Modalität nicht unter Beweis stellen kann. Es wird vermutet, dass in der überwiegenden Anzahl der Fälle die Items von

den Testpersonen zumindest in einer Modalität richtig zugeordnet und damit verstanden werden. Das vorliegende Instrument bietet alle drei Modalitäten mit dem jeweils gleichen Itemkorpus an. Diese Konzeption ermöglicht eine Übersicht über die tatsächliche Fehlerstruktur der gehörlosen Testpersonen.

Gleichzeitig wird angenommen, dass die gehörlosen Kinder, für die die DGS die Muttersprache darstellt, das gleiche Wortschatzniveau wie hörende Kinder gleichen Alters in ihrer Muttersprache, der Lautsprache, erreichen würden. Die gehörlosen Testpersonen, für die die DGS die Basissprache darstellt, würden, so die Annahme, zumindest vergleichbare Scores wie die hörenden Kinder erreichen, d. h. relativ nah an das Niveau der Muttersprachler heranreichen.

1.2.1 Überprüfung I

In diesem Zusammenhang wurde ausgewertet, ob die Testpersonen – hörende wie gehörlose – die Items, die sie in bestimmten Subtests falsch beantworteten, in den jeweils anderen Subtests richtig auswählten. Die Tabellen Tab. 69a/b–B 2.10 stellen die Fehlerscores der ursprünglichen Bewertung denen dieser neuen Bewertung gegenüber. Angegeben werden zunächst die Fehler insgesamt, d. h. die Prozentangabe der ursprünglichen Fehlerbewertung, in der jede fehlerhafte Antwort mit null Punkten bewertet wurde, ungeachtet der Frage, ob die betreffende Testperson das Item möglicherweise in einer der anderen angebotenen Modalitäten richtig auswählte. Demgegenüber finden sich die Angaben der Fehleranzahl in Prozent, die von der betreffenden Testperson durchgehend durch alle Subtests gemacht wurden. D. h., als Fehlerpunkt wurde eine falsche Antwort in dieser Sparte erst dann gewertet, wenn die Testperson tatsächlich in keiner Sprachmodalität Zugriff auf einen entsprechenden Begriff bewies.

Tab. 70a-B 2.10: Fehlervergleich Klasse 3								
Nomen			Verben			Adjektive		
Fehler insge- samt %	Fehler durch- gehend in allen Subtests %	Na- me	Fehler insge- samt %	Fehler durch- gehend in allen Subtests %	Na- me	Fehler insge- samt %	Fehler durch- gehend in allen Subtests %	Na- me
9,3	0	F	20,4	5,6	F	30	5,6	F
13	0	A	26	0	M	40,7	22,2	M
13	0	M	27,8	0	A	42,6	22,2	A
26	0	V	27,8	5,6	V	57,4	27,8	T
27,8	0	T	37	0	T	59	22,2	V
55,6	11,1	O	53,7	16,7	O	72,2	38,9	O

Tab. 70b-B 2.10: Fehlervergleich Klasse 4								
Nomen			Verben			Adjektive		
Fehler insge- samt %	Fehler durch- gehend in allen Subtests %	Na- me	Fehler insge- samt %	Fehler durch- gehend in allen Subtests %	Na- me	Fehler insge- samt %	Fehler durch- gehend in allen Subtests %	Na- me
10,2	0	K	14,4	5,6	K	26,4	16,7	K
20,4	0	L	22,2	11,1	L	30	11,1	L
25,5	5,6	S	29,2	0	S	63	27,8	F
35,2	5,6	E	38,9	0	F	66,7	27,8	S
50 %	5,6	F	44,4	5,6	E	69	22,2	E

Diese Tabellen der Klassen 3 und 4 geben ein differenziertes Bild darüber ab, wie gut die Testpersonen zum großen Teil abschneiden. So ist zunächst der Wert der durchgängig in allen Subtests fehlerhaft beantworteten Items in allen Fällen geringer, in aller Regel deutlich geringer, als bei der ursprünglichen Fehlerbewertung. Die Differenz zwischen den beiden Werten bewegt sich für Klasse 3 zwischen dem höchsten Wert von 44,5% und dem niedrigsten Wert von 9,3%. In Klasse 4 beträgt der höchste Differenzwert 46,8% und der niedrigste 8,8%. Dies bestätigt bereits die Annahme, dass die Differenz zwischen den beiden Modalitäten der Fehlerbewertung prägnant ausfällt.

Dabei entfallen in Klasse 3 die jeweils höchsten Differenzwerte pro Wortart auf die durchgängig beste Testperson Florian mit 44,5% bei den Nomen, 37% bei den Verben und 24,4% bei den Adjektiven. Die jeweils niedrigsten Differenzwerte entfallen auf die durchgängig schlechteste Testperson Orhan mit 9,3% bei den Nomen, 14,8% bei den Verben und 33,3% bei den Adjektiven. In Klasse 4 entfallen die jeweils höchsten Differenzwerte pro Wortart bei den Nomen mit 44,4% auf die Testperson Farid, bei den Verben mit 38,8% und bei den Adjektiven mit 46,8% auf die Testperson Erik. Beide Schüler gehören zu den schlechtesten der Klasse. Die jeweils niedrigsten Differenzwerte entfallen auf die durchgängig beste Testperson Katharina mit 10,2% bei den Nomen, 8,8% bei den Verben und 9,7% bei den Adjektiven.

In die Betrachtung der Ergebnisse muss einbezogen werden, dass sich, wie oben bereits erwähnt, unter den Gehörlosen zwei Testpersonen befinden, welche bis vor wenigen Jahren über keine Sprache verfügten und die Werte entsprechend negativ beeinflussen. Sprachschwierigkeiten bzw. -mangel dieser Prägnanz treten in der hörenden Vergleichsstichprobe nicht auf. Selbst wenn sich in der hörenden Gruppe einzelne sprachschwache Schüler befanden, fällt dies wesentlich weniger ins Gewicht als in der Stichprobe von nur elf gehörlosen Kindern. Dieser Umstand unterstreicht die Ergebnisse.

1.2.2 Überprüfung II

Tab. 71a-C 1.2.2: Fehlervergleich Gesamtergebnisse Gehörlose					
Fehler (insgesamt)			Fehler (durchgehend in allen Subtests)		
Nomen	Verben	Adjektive	Nomen	Verben	Adjektive
26 %	31,1 %	50,6 %	2,6 %	4,6 %	22,2 %

Tab. 71b-C 1.2.2: Fehlervergleich Gesamtergebnisse Hörende					
Fehler (insgesamt)			Fehler (durchgehend in allen Subtests)		
Nomen	Verben	Adjektive	Nomen	Verben	Adjektive
4,6 %	6 %	15,5 %	1,1 %	1,2 %	5,8 %

Die Werte der Gegenüberstellung der Gesamtergebnisse Gehörloser und Hörender zeigen, dass sich die Differenz zwischen ihnen deutlich relativiert. Erscheint bei der ursprünglichen Fehlerbewertung der Abstand zwischen Gehörlosen und Hörenden mit 21,4% Differenz bei den Nomen, 25,1% bei den Verben und 35,1% bei den Adjektiven noch hoch, so gleichen sich die Werte in der Gegenüberstellung der durchgehend in allen Subtests begangenen Fehler einander deutlich an. Hier beträgt die Differenz bei den Nomen nur noch 1,5%, bei den Verben 2,4% und bei den Adjektiven 16,4%.

1.2.3 Überprüfung III

Tab. 72-C 1.2.3: Vergleich Hörende Lautsprache vs. Gehörlose Gebärdensprache		
Hörende Lautsprache		
Fehler Nomen	Fehler Verben	Fehler Adjektive
4,2 %	5,6 %	20,9 %
Gehörlose Gebärdensprache (insgesamt)		
Fehler Nomen	Fehler Verben	Fehler Adjektive
6,1 %	11,1 %	35,9 %

Katharina Gebärdensprache (muttersprachlich gebärdend)		
Fehler Nomen	Fehler Verben	Fehler Adjektive
0 %	5,6 %	16,8 %
Gehörlose Gebärdensprache (ohne Erik, Orhan, Simon)		
Fehler Nomen	Fehler Verben	Fehler Adjektive
0,5 %	1,5 %	5,6 %

Am prägnantesten präsentieren sich schließlich die Werte des Vergleichs zwischen dem Subtest Lautsprache bei hörenden Kindern und dem Subtest Gebärdensprache bei gehörlosen Kindern. Betrachtet man zunächst die oberen beiden Spalten der Tabelle, so stellt sich gemäß der Annahmen der Autoren heraus, dass die bilingual erzogenen Kinder des Hamburger Schulversuchs, besonders in den Wortarten Nomen und Verben, dicht hinter den hörenden Testpersonen liegen. Dies obgleich sie die Gebärdensprache zum weit überwiegenden Anteil nur als ihre Basissprache beherrschen.

Die Werte der einzigen Testperson, die muttersprachlich gebärdend aufwuchs (Katharina), finden sich in der dritten Spalte. Sie weist in den Wortarten Nomen und Adjektive sogar bessere Ergebnisse auf als die hörenden Vergleichspersonen. Nimmt man nun die drei in der gebärdensprachlichen Modalität schwächsten Schüler Erik und Orhan und Simon (jeweils 72% richtige Antworten) aus der Bewertung heraus (letzte Spalte), um die Werte nicht ungleichgewichtig zu den hörenden Testpersonen zu belasten, so zeigt sich ein noch besseres Ergebnis. Die Annahme, die gehörlosen bilingualen Kinder verfügen – zumindest in der Gebärdensprache – über einen angemessenen Wortschatz, wird mit diesen Werten bestätigt, denn sie sind deutlich besser als die der hörenden in der Lautsprache.

Es lässt sich zusammenfassend sagen, dass – bei aller Vorläufigkeit der mit einer Test-Rohversion erhobenen Daten – die Ergebnisse der bilingualen Kinder auffällig gut ausfallen. Die weitere Betrachtung von Testergebnissen in dieser Auswertungsmodalität wird zeigen, ob sich diese Aussage auch auf dem Hintergrund des deutlich erweiterten Datenpools aufrecht erhalten lässt.

1.3 Schwierigkeitsgefälle der Modalitäten im Rahmen der Hauptuntersuchung

Hier soll mit den Daten der Hauptuntersuchung genau die gleiche Annahme untersucht werden, die bereits weiter oben im Rahmen der Pilotstudie unter der gleichen Überschrift for-

muliert wurde. Angenommen wird also, dass der überwiegende Anteil der Testpersonen in der Modalität Gebärdensprache die besten Ergebnisse erzielt.

1.3.1 Überprüfung I

Zunächst wurde überprüft, wie viel Prozent aller Testpersonen in der Gebärdens- bzw. in der Schrift- und in der Lautsprache die besten Ergebnisse erzielten. Von den 112 Testpersonen konnten 111 an dieser Untersuchung zur maximalen Leistung innerhalb der Modalitäten teilnehmen, da eine Testperson nur den Subtest Schriftsprache bearbeitet hat und deshalb kein Vergleich möglich ist. Von den 111 haben 74 Testpersonen (66,7%) die beste Leistung von drei Modalitäten in der Gebärdensprache, zwölf Testpersonen (10,8%) die beste Leistung von drei Modalitäten in der Schriftsprache und neun Testpersonen (8,1%) die beste Leistung von drei Modalitäten in der Lautsprache erzielt. Darüber hinaus haben zwei Testpersonen (1,8%) in allen Subtests gleich gute Leistungen erbracht, zwei Testpersonen (1,8%) haben sowohl in der Gebärdens- als auch in der Lautsprache die gleiche, beste Leistung erbracht. Eine Testperson (0,9%) hat sowohl in Gebärdens- als auch in Lautsprache, eine Testperson (0,9%) sowohl in Schrift- als auch in Lautsprache die gleich gute, beste Leistung erzielt. Zehn Testpersonen haben nur zwei der drei Modalitäten bearbeitet. Von ihnen haben acht Testpersonen (7,2%) die bessere Leistung in der Gebärdensprache (von drei dieser Testpersonen fehlt die Leistung in der Schriftsprache, von fünf Testpersonen die in der Lautsprache) und zwei Testpersonen (1,8%) die bessere Leistung in der Modalität Lautsprache (von einer dieser Testpersonen fehlt die Leistung in der Gebärdensprache, von einer Testperson die in der Schriftsprache) erzielt.

Von 111 Testpersonen haben 74 die beste Leistung aus drei Subtests in der Gebärdensprache, zusätzlich 8 die beste Leistung aus zwei Subtests in der Gebärdensprache erzielt. Diese Anzahl entspricht 66,7% bzw. 73,9% der Gesamtstichprobe. Damit haben fast drei Viertel aller Testpersonen ihre stärkste Leistung in der Gebärdensprache gezeigt.

1.3.2 Überprüfung II

Daraufhin wurde untersucht, ob Korrelationen zwischen den Testpersonen, die in der Gebärdensprache bzw. in der Schrift- oder Lautsprache die beste Leistung erzielten, und bestimmten Begleitmerkmalen bestehen. Diese Merkmale sind das Konzept, nach welchem die Testpersonen unterrichtet werden, die Klassenstufe und das Geschlecht dem sie angehören, Alter und Intelligenzquotient und ihre Testleistung im Vergleich zur Gesamtstichprobe.

Zunächst wurden die einflussnehmenden Faktoren derjenigen Testpersonen untersucht, die in der Gebärdensprache am besten abschnitten. Insgesamt erzielten 82 Testpersonen bei zwei

oder drei Subtests die beste Leistung in dieser Modalität. Darunter waren fünf (83,3%) der sechs bilingual erzogenen, 57 (81,4%) der 70 gebärdenorientiert und 20 (57,1%) der 35 hörgerichtet erzogenen Testpersonen. Es handelt sich dabei um 36 (80%) der Dritt-, 17 (55%) der Viert- und 29 (81%) der Fünftklässler. Es waren 42 Mädchen und 40 Jungen. Sie waren mit einem durchschnittlichem Alter von 10;6 Jahren um einen Monat älter und erreichten mit durchschnittlich 97 Intelligenzpunkten 1 Punkt mehr als der Gesamtdurchschnitt. Im Verhältnis zur gesamten Stichprobe erzielten sie die Leistungsrangplätze 1 bis 110, durchschnittlich 60. Auch in ihrer Gesamtleistung über alle drei Modalitäten rangieren die betreffenden Testpersonen mit dem durchschnittlichen Leistungsrangplatz 60 zwischen dem 4. und 101. Rangplatz.

In einem zweiten Schritt wurden diejenigen Testpersonen überprüft, die in der Schriftsprache bzw. in der Lautsprache am besten abschnitten. Von den zwölf Testpersonen, die ihre beste Leistung in der Schriftsprache zeigten, war eine Testperson (16,7%) nach dem bilingualen Konzept, sieben (10,0%) nach gebärdenorientierten Konzepten und vier (11,4%) nach hörgerichteten Konzepten erzogen. Es waren zwei (4,4%) der Drittklässler, fünf (16,1%) der Viertklässler und fünf (13,9%) aller Fünftklässler. Die Gruppe besteht aus sechs Mädchen und sechs Jungen. Diese waren um durchschnittlich zwei Monate älter als der Gesamtdurchschnitt der Stichprobe und erzielten durchschnittlich einen um 4,2 Intelligenzpunkte höheren Intelligenzquotienten als der Gesamtdurchschnitt. Die zwölf Testpersonen schneiden insgesamt in der Modalität Schriftsprache sehr gut ab. Im Verhältnis zur gesamten Stichprobe erzielten sie die Leistungsrangplätze 3 bis 30, durchschnittlich 15. Auch in ihrer Gesamtleistung (in allen drei Modalitäten) bewegen sich die zwölf Testpersonen im obersten Leistungsdrittel, sie rangieren zwischen dem 7. und 44., durchschnittlich auf dem 21. Rangplatz. Darüber hinaus fällt auf, dass fünf dieser zwölf Testpersonen aus derselben Schule, vier davon sogar aus derselben Klasse kommen.

Unter den elf Testpersonen, die in der Lautsprache die stärkste Leistung erzielten, wurden fünf (7,1%) gebärdenorientiert und sechs (17,1%) hörgerichtet erzogen. Es waren drei (6,6%) aller Dritt-, sechs (19,4%) aller Viert- und zwei (5,6%) aller Fünftklässler; vier Mädchen und sieben Jungen. Sie waren im Durchschnitt zwei Monate älter und 6,7 Intelligenzpunkte besser als der Durchschnitt der gesamten Stichprobe. Diese elf Testpersonen sind unter den leistungsstärksten in der Lautsprache, sie befinden sich leistungsmäßig auf Rangplatz 1 bis 36, durchschnittlich auf Platz 18. Auch von diesen Testpersonen besuchen fünf dieselbe, nach hörgerichtetem Konzept unterrichtende Schule. Eine dieser elf Testpersonen hat lediglich die Subtests Schrift- und Lautsprache bearbeitet und wird aus diesem Grund im weiteren Verlauf der Untersuchung nicht weiter berücksichtigt.

In der Gruppe derjenigen Testpersonen, die ihr bestes Ergebnis im Subtest Gebärdensprache erzielten, waren 83,3% der bilingual erzogenen, 81,4% der gebärdenorientiert und 57,1% der hörgerichtet erzogenen Testpersonen. Neben den Testpersonen, die konzeptionell bereits mit der Gebärdensprache vertraut sind (bilingual, gebärdenorientiert) haben demnach auch mehr als die Hälfte der primär hörgerichtet erzogenen Testpersonen einen guten Zugang zur Gebärdensprache gezeigt. Die übrigen Merkmale fielen in dieser Gruppe unauffällig aus.

Bei der detaillierten Betrachtung derjenigen Testpersonen, die in der Schrift- oder Lautsprache die besten Ergebnisse erzielten, ist ihr insgesamt gutes Abschneiden in der Modalität Schriftsprache das einzig auffällige Ergebnis. Da jedoch ein Drittel der betreffenden Schüler aus einer Klasse stammen, in der womöglich seitens des Klassenlehrers besonderes Gewicht auf die Schriftsprache gelegt wird, wird hier mit verallgemeinernden Schlüssen zurückhaltend umgegangen.

In der Gruppe schließlich, die in der Lautsprache am stärksten abgeschnitten haben, fällt auf, dass kein bilingual erzogener Schüler darunter ist.

1.3.3 Überprüfung III

Schließlich wurde überprüft, wie die Testpersonen, die in anderen Modalitäten als in der Gebärdensprache ihre beste Leistung erzielten, im Verhältnis zu der gesamten Stichprobe abschnitten und ob sie bestimmte Schwierigkeiten in der Gebärdensprache hatten. Die zwölf Testpersonen, die in der Schriftsprache am besten abschnitten, haben dort eine Leistung erzielt, die zwischen 1 und 7 Punkten, durchschnittlich 3,7 Punkte über ihrer Leistung in der Modalität Gebärdensprache lag. Die zwölf Testpersonen lagen in ihrer durchschnittlichen Leistung im Subtestteil Gebärdensprache Nomen um 1,5 Punkte, im Subtestteil Gebärdensprache Verben um 0,5 Punkte und im Subtestteil Gebärdensprache Adjektive um 2,1 Punkte über der durchschnittlichen Leistung der Gesamtstichprobe. In der gesamten Modalität Gebärdensprache (max. 54 Punkte) lag ihre Leistung um durchschnittlich 4,1 Punkte über der Leistung der Stichprobe. Sie haben in vier Aufgaben des Subtestteils Gebärdensprache Nomen durchschnittlich mehr Fehler gemacht als der Gesamtdurchschnitt. Es handelt sich um die Items 'Straße', 'Zahn', 'Zahl' und 'Welt', die nach neuer Reihenfolge für den Subtest Gebärdensprache auf den Rangplätzen 17, 4, 9 und 6 durchschnittlich auf Platz 9 rangieren. Es wurden demnach sowohl bei leichten als auch bei schweren Items mehr Fehler gemacht. Innerhalb dieser vier Aufgaben wählten diese Testpersonen in Relation zu ihrer Gruppengröße 24,1% mehr Distraktoren mit besonderer inhaltlicher Ähnlichkeit, 14,2% mehr Distraktoren mit Gebärdenähnlichkeit, 12,1% mehr Distraktoren mit Mundbildähnlichkeit und 1,8% weniger Distraktoren mit einfach inhaltlicher Ähnlichkeit. Es handelt sich demnach überwiegend um die qualitativ besseren Fehler dieses Subtests.

Im Subtestteil Gebärdensprache Verben haben die zwölf Testpersonen in sechs Aufgaben durchschnittlich mehr Fehler gemacht, als die gesamte Stichprobe aufweist. Dies waren die Items ‘springen’, ‘schlafen’, ‘denken’, ‘verstecken’, ‘brennen’ und ‘stinken’, die auf den neuen Rangplätzen 4, 11, 18, 10, 12 und 3 rangieren. Eine überdurchschnittliche Fehlerwahl erfolgte demnach bei mittelschweren Aufgaben des durchschnittlichen Rangplatzes 10. Die entsprechenden Testpersonen wählten innerhalb dieser Aufgaben 4,9% mehr Distraktoren mit besonderer inhaltlicher Ähnlichkeit, 27,4% mehr Distraktoren mit Gebärdenähnlichkeit, 0,2% mehr Distraktoren mit Mundbildähnlichkeit und 12,0% mehr Distraktoren mit einfach inhaltlicher Ähnlichkeit. Besonders häufig wurden demnach die in diesem Subtest besonders plausiblen Distraktoren mit Ähnlichkeit in der Gebärde gewählt.

Im Subtestteil Gebärdensprache Adjektive machten die zwölf Testpersonen in nur zwei Aufgaben durchschnittlich mehr Fehler, als die gesamte Stichprobe aufweist. Es handelt sich dabei um die Items ‘schief’ und ‘schmutzig’, die auf Rangplatz 12 und 4, durchschnittlich auf Rangplatz 8 rangieren. Dabei waren 2,3 % mehr Distraktoren mit besonderer inhaltlicher Ähnlichkeit, 2,3% mehr Distraktoren mit Gebärdenähnlichkeit und 1,4% mehr Distraktoren mit einfach inhaltlicher Ähnlichkeit gewählt. Es sind in diesem Subtestteil keine besonderen Häufungen in der Fehlerwahl ersichtlich, die Fehleranzahl übersteigt nur gering die der gesamten Stichprobe.

Die 10 Testpersonen mit Bestleistung in der Lautsprache haben dort ihre Leistung in der Gebärdensprache um 1 bis 10, durchschnittlich um 3,5 Punkte übertroffen. Die zehn Testpersonen, die ihre beste Leistung in der Modalität Lautsprache erzielten, lagen in ihrer durchschnittlichen Leistung im Subtestteil Gebärdensprache Nomen um 0,3 und im Subtestteil Gebärdensprache Adjektive um 1,4 Punkte über der durchschnittlichen Leistung der Gesamtstichprobe, im Subtestteil Gebärdensprache Verben um 0,4 Punkte unter der durchschnittlichen Leistung der Gesamtstichprobe. In der gesamten Modalität Gebärdensprache (max. 54 Punkte) lag ihre Leistung um durchschnittlich 1,3 Punkte über der Leistung der Stichprobe. Sie haben im Subtestteil Gebärdensprache Nomen in 6 Aufgaben durchschnittlich mehr Fehler gemacht, als die gesamte Stichprobe aufweist. Dies war bei den Items ‘Mutter’, ‘Garten’, ‘Zahn’, ‘Leute’, ‘Geld’ und ‘Rose’ der Fall. Diese Items rangieren nach der neuen Reihenfolge auf den Rangplätzen 16, 18, 4, 10, 14 und 8, durchschnittlich auf Platz 12. Die entsprechenden Testpersonen wählten innerhalb dieser Aufgaben 30,9% mehr Distraktoren mit besonderer inhaltlicher Ähnlichkeit, 61,9% mehr Distraktoren mit Gebärdenähnlichkeit, 54,6% mehr Distraktoren mit Mundbildähnlichkeit und 27,4% mehr Distraktoren mit einfach inhaltlicher Ähnlichkeit, als die gesamte Stichprobe aufweist..

Im Subtestteil Gebärdensprache Verben machten die zehn Testpersonen in zehn Aufgaben durchschnittlich mehr Fehler als die gesamte Stichprobe. Dies betrifft die Items ‘fliegen’, ‘denken’, ‘hinsetzen’, ‘warten’, ‘werfen’, ‘schreiben’, ‘singen’, ‘brennen’, ‘schieben’ und ‘stinken’. Diese Items befinden sich in der neuen Schwierigkeitsrangfolge auf den Plätzen 2, 18, 17, 15, 8, 1, 14, 12, 9 und 3, durchschnittlich auf Rangplatz 10. Die zehn Testpersonen wählten im Vergleich zu der gesamten Stichprobe 74,7% mehr Distraktoren mit besonderer inhaltlicher Ähnlichkeit, 28,4% mehr Distraktoren mit Gebärdenähnlichkeit, 32,9% mehr Distraktoren mit Mundbildähnlichkeit und 20,1% mehr Distraktoren mit einfach inhaltlicher Ähnlichkeit. Eine deutliche Häufung der Wahl von Distraktoren mit besonderer inhaltlicher Ähnlichkeit ist neben einer an sich bereits hohen Überschreitung der Fehleranzahl der Gesamtstichprobe in diesen Aufgaben ersichtlich.

Im Subtestteil Gebärdensprache Adjektive machten diese Testpersonen in fünf Aufgaben durchschnittlich mehr Fehler, als die gesamte Stichprobe aufweist. Es handelt sich dabei um die Items ‘voll’, ‘blond’, ‘spitz’, ‘orange’ und ‘schief’, die sich auf den Schwierigkeitsrangplätzen 15, 14, 2, 5 und 12, durchschnittlich auf Platz 10 befinden. Dabei haben die 10 Testpersonen 20,9% mehr Distraktoren mit besonderer inhaltlicher Ähnlichkeit, 31,8% mehr Distraktoren mit Gebärdenähnlichkeit, 7,2% mehr Distraktoren mit Mundbildähnlichkeit und 1,8% weniger Distraktoren mit einfach inhaltlicher Ähnlichkeit gewählt als in der gesamten Stichprobe auftreten. Eine Häufung in der Wahl von plausiblen gegenüber weniger plausiblen Distraktoren wird in diesem Subtestteil deutlich.

Zwölf Testpersonen haben ihre beste Leistung in der Schriftsprache, elf in der Lautsprache erzielt. Dabei lagen diese 23 Testpersonen mit ihren Ergebnissen in der Gebärdensprache über dem Gesamtdurchschnitt. Auffällig ist, dass die Testpersonen mit Bestleistung in der Schriftsprache die Leistung der gesamten Stichprobe in der Modalität Gebärdensprache deutlicher überschritten als die Testpersonen, die ihre beste Leistung in der Lautsprache erzielt haben. Sie machten zudem insgesamt weniger und unter diesen mehr qualitativ bessere Fehler in der Modalität Gebärdensprache als die Testpersonen, die am besten in der Lautsprache abschnitten.

Bei den beiden Gruppen handelt es sich demnach um insgesamt leistungsstarke Testpersonen. Ein Rückgriff auch auf die Überprüfung II belegt, dass zum Erreichen guter Kenntnisse in diesen beiden Modalitäten gute allgemeine Fähigkeiten nötig sind, wie z. B. Intelligenz. Denn die Quotienten lagen um 4,2% bzw. 6,7% durchschnittlich höher als bei der Gesamtstichprobe.

1.4 Vergleich der Kompetenzen mit modifizierter Fehlerbewertung

Der Vergleich der Kompetenzen zwischen gehörlosen und hörenden Testpersonen bei Anwendung einer modifizierter Fehlerbewertung mit den Daten der Hauptuntersuchung entspricht der im Rahmen der Pilotstudie unter C 1.2 geschilderten Untersuchung. So werden in der modifizierten Version die Fehler als solche bewertet, die durch alle Modalitäten hindurch begangen wurden.

1.4.1 Überprüfung I

Zur Überprüfung der Annahme wurde ausgezählt, wie oft Nomen, Verben oder Adjektive von den Testpersonen in mindestens einer Modalität richtig zugeordnet wurden. Die folgenden Tabellen geben wieder, wie viele Items nur in einer, in zwei oder in allen drei Modalitäten richtig beantwortet und wie viele Items in mindestens einer Modalität richtig zugeordnet wurden. In der ersten Tabelle werden dabei die richtigen Antworten der 103 Testpersonen dargestellt, die in allen Modalitäten den Subtestteil Nomen bearbeitet haben, in der zweiten und dritten Tabelle sind es jeweils 101 Testpersonen, die in allen Subtestteilen die Wortart Verben bzw. Adjektive bearbeitet haben:

Tab. 73a-C 1.4.1: Ergebnisse Nomen durch drei Modalitäten (n = 103)					
Aufgabe	nie richtig	in einer Modalität richtig	in zwei Modalitäten richtig	in drei Modalitäten richtig	mindestens einmal richtig
1	-	1	15	87	103
2	3	6	16	78	100
3	1	10	60	32	102
4	1	12	32	58	102
5	5	15	35	48	98
6	9	17	31	9	94
7	14	17	24	48	89
8	5	31	24	43	98
9	-	29	38	36	103
10	7	34	40	22	96
11	1	15	36	51	102
12	6	25	32	40	97
13	7	23	40	33	96
14	9	23	30	41	94
15	14	24	28	37	89
16	17	28	34	24	86
17	17	19	34	33	86
18	29	20	23	31	74

Tab. 73b-C 1.4.1: Ergebnisse Verben durch drei Modalitäten (n = 101)					
Aufgabe	nie richtig	in einer Modalität richtig	in zwei Modalitäten richtig	in drei Modalitäten richtig	mindestens einmal richtig
1	2	8	23	68	99
2	-	13	16	72	101
3	-	14	48	39	101
4	-	11	38	52	101
5	1	13	23	64	100
6	1	13	21	66	100
7	1	18	31	51	100
8	1	19	24	57	100
9	1	33	42	25	100
10	5	18	37	41	96
11	1	18	31	51	100
12	2	32	30	37	99
13	3	21	41	36	98
14	14	27	26	34	87
15	9	26	28	38	92
16	19	42	24	16	82
17	34	39	20	8	67
18	32	46	15	8	69

Tab. 73c-C 1.4.1: Ergebnisse Adjektive durch drei Modalitäten (n = 101)					
Aufgabe	nie richtig	in einer Modalität richtig	in zwei Modalitäten richtig	in drei Modalitäten richtig	mindestens einmal richtig
1	1	13	43	44	100
2	4	21	18	58	97
3	7	8	31	55	94
4	7	15	24	55	94
5	5	17	40	39	96
6	-	30	28	43	101
7	8	22	28	43	93
8	4	22	35	40	97
9	2	21	37	41	99
10	12	40	38	11	89
11	11	29	26	35	90
12	11	36	32	22	90
13	21	33	18	29	80
14	15	30	20	36	86
15	22	25	23	31	78
16	21	29	27	24	80
17	29	34	27	11	72
18	31	33	28	9	70

Vergleicht man die Anzahl der Items, die durchgehend richtig gelöst wurden, mit der Anzahl derer, die mindestens einmal richtig beantwortet wurden, so fällt ein deutlicher Anstieg der Anzahl richtiger Lösungen auf. In der Wortart Nomen wurden Aufgaben zwischen 9- und

87-mal, durchschnittlich 42-mal durchgehend richtig gelöst, dagegen waren Items zwischen 74- und 103-mal, durchschnittlich 95-mal in mindestens einer Modalität bekannt. In der Wortart Verben sind die Items zwischen 8- und 72-, durchschnittlich 42-mal durchgehend bekannt, wohingegen Aufgaben zwischen 67- bis 101-mal, durchschnittlich 94-mal mindestens in einer Modalität bekannt waren. In der Wortart Adjektive waren die Items zwischen 9- und 58-mal, durchschnittlich 35-mal durchgehend bekannt, zwischen 70- und 101-mal, durchschnittlich 89-mal mindestens in einer Modalität bekannt.

Die Werte belegen überdeutlich die Differenz zwischen der jeweiligen neuen und der modifizierten Fehlerbewertung. Ein Großteil der Items ist den Testpersonen in mindestens einer Modalität bekannt, d. h. sie verfügen über den mit dem Wortschatztest erfassten Grundwortschatz. So wurden von 92,2% der Testpersonen alle Nomen, von 93,1% der Testpersonen alle Verben und von 88,1% der Testpersonen alle Adjektive mindestens in einer Modalität richtig zugeordnet. Für alle drei Wortarten mit geringer Einschränkung bei den Adjektiven gilt, dass der durchschnittliche Wert in der modifizierten Version an die 100% heranreicht.

1.4.2 Überprüfung II

Darüber hinaus wurde untersucht, wie häufig die Items, die nur einmal richtig beantwortet wurden, gerade in der Modalität Gebärdensprache richtig zugeordnet wurden. Die folgenden Tabellen veranschaulichen diesen Zusammenhang für jede Wortart getrennt:

Tab. 74a-C 1.4.2: Anteil der Nomen, die nur in Gebärdensprache richtig beantwortet wurden			
Aufgabe	in einer Modalität richtig	in GS richtig	%
1	1	1	100
2	6	2	33,3
3	10	9	90
4	12	7	58,3
5	15	7	46,7
6	17	7	41,2
7	17	8	47,1
8	31	14	45,2
9	29	24	82,8
10	34	27	79,4
11	15	1	6,7
12	25	7	28
13	23	13	56,5
14	23	4	17,4
15	24	5	20,8
16	28	14	50
17	19	5	26,3
18	20	6	30

Tab. 74b-C 1.4.2: Anteil der Verben, die nur in Gebärdensprache richtig beantwortet wurden			
Aufgabe	in nur einer Modalität richtig	in GS richtig	%
1	8	7	87,5
2	13	13	100
3	14	11	78,6
4	11	10	90,9
5	13	12	92,3
6	13	11	84,6
7	18	16	88,9
8	19	18	94,7
9	33	30	90,9
10	18	14	77,8
11	18	6	33,3
12	32	30	93,8
13	21	17	81
14	27	26	96,3
15	26	18	69,2
16	42	32	76,2
17	39	15	38,5
18	46	26	56,5

Tab. 74c-C 1.4.2: Anteil der Adjektive, die nur in der Gebärdensprache richtig beantwortet wurden			
Aufgabe	in einer Modalität richtig	in GS richtig	%
1	13	11	84,6
2	21	17	81
3	8	2	25
4	15	8	53,3
5	17	13	76,5
6	30	24	80
7	22	12	54,5
8	22	19	86,4
9	21	11	52,4
10	40	29	72,5
11	29	8	27,6
12	36	23	63,9
13	33	7	21,2
14	30	3	10
15	25	3	12
16	29	15	51,7
17	34	14	41,2
18	33	3	9,1

Es zeigt sich, dass überdurchschnittlich häufig in Fällen, in denen ein Item nur in einer Modalität richtig zugeordnet wurde, dieses in der gebärdensprachlichen Modalität erfolgte. Bei den Nomen war dies bei 47,8%, bei den Verben bei 79,5% und bei den Adjektiven bei

50,2% der nur einmal richtig beantworteten Items der Fall. Bei gleichmäßiger Verteilung auf alle drei Sprachmodalitäten wären dagegen nur 33,3% dieser Items gerade in der Gebärdensprache richtig zugeordnet worden. So wird deutlich, wie wenig man gehörlosen Testpersonen gerecht wird, wenn man sie nur in der Lautsprache abfragt.

1.5 Bilinguale Schüler an der Leistungsspitze

Es wird angenommen, dass Kinder, die nach dem bilingualen Konzept unterrichtet werden, kompetenter in Gebärdens- und Laut- sowie in Schriftsprache abschneiden als vergleichbare Testpersonen, die nach anderen Konzepten unterrichtet werden. Diese Annahme basiert auf Ergebnissen des Zwischenberichts zum bilingualen Schulversuch Hamburg (Günther u. a. 1999 a) sowie auf Erfahrungen mit der ersten Version des Wortschatztests in den beiden Schulversuchsklassen

1.5.1 Überprüfung I

Zunächst wurde überprüft, in welchen Subtests und Subtestteilen die bilingual erzogenen Testpersonen am leistungsstärksten abschnitten. Daraufhin wurde untersucht, wie gut die Testpersonen, die nach anderen gehörlosenspezifischen Unterrichtskonzepten erzogen werden, in den verschiedenen Sprachmodalitäten abschnitten.

Dazu wurden zunächst die Leistungen der bilingual erzogenen Testpersonen in der Modalität Gebärdensprache überprüft. Die sechs bilingual erzogenen Testpersonen haben im Subtestteil Gebärdensprache Nomen eine durchschnittliche Leistung von 17,7 Punkten erreicht. Der Durchschnitt der gesamten Stichprobe in diesem Subtestteil liegt bei 13,1 Punkten. Die sechs Testpersonen befinden sich in diesem Subtestteil auf den Leistungsrangplätzen 5, 6, 7, 8, 9 und 19, durchschnittlich auf Platz 9 von 110. Im Subtestteil Gebärdensprache Verben haben die sechs Testpersonen eine durchschnittliche Leistung von 16,2 Punkten gegenüber einer Gesamtleistung von 15,2 Punkten erzielt. Die betreffenden Testpersonen befinden sich mit dieser Leistung auf den Rangplätzen 2, 3, 13, 14, 81 und 93, durchschnittlich auf Platz 34. Im Subtestteil Gebärdensprache Adjektive haben die sechs Testpersonen 15,8, die gesamte Stichprobe 12,4 Punkte erreicht. Damit rangieren sie auf den Rangplätzen 1, 4, 14, 21, 22 und 33, durchschnittlich auf Platz 16. Für den gesamten Subtest Gebärdensprache haben die bilingual erzogenen Testpersonen eine durchschnittliche Leistung von 48,8 gegenüber einer Leistung von 40,4 der gesamten Stichprobe erzielt. Sie befinden sich mit ihrer Leistung in der Gebärdensprache auf den Rangplätzen 1, 5, 11, 14, 28 und 33, durchschnittlich auf dem Rangplatz 15.

Daraufhin wurde die Leistung der bilingualen Testpersonen in der schriftsprachlichen Modalität untersucht. Im Subtestteil Schriftsprache Nomen haben die bilingual erzogenen Testpersonen eine durchschnittliche Leistung von 17,0 Punkten gegenüber einer Gesamtleistung von 11,9 Punkten erzielt. Sie befinden sich mit dieser Leistung auf den Rangplätzen 1, 10, 13, 20, 23 und 25, durchschnittlich auf Platz 14. Im Subtestteil Schriftsprache Verben haben die sechs Testpersonen eine durchschnittliche Leistung von 14,8, die gesamte Stichprobe von 9,8 Punkten erzielt. Damit rangieren die bilingual erzogenen Testpersonen auf den Leistungsrangplätzen 4, 7, 11, 20, 22 und 33, durchschnittlich auf Platz 16. Im Subtestteil Schriftsprache Adjektive haben die 6 Testpersonen eine durchschnittliche Leistung von 15,7 Punkten gegenüber einer Leistung der gesamten Stichprobe von 10,1 Punkten erbracht. Sie befinden sich mit dieser Leistung auf den Rangplätzen 1, 8, 12, 16, 25 und 50, durchschnittlich auf Platz 19. In der Gesamtmodalität Schriftsprache haben die sechs bilingual erzogenen Testpersonen durchschnittlich 47, 5 Punkte erreicht, die gesamte Stichprobe hat 32,0 Punkte erzielt. Sie befinden sich damit auf den Leistungsrangplätzen 1, 7, 11, 17, 24 und 37, durchschnittlich auf Platz 16.

Auch die Ergebnisse der bilingual erzogenen Testpersonen im Subtest Lautsprache wurden überprüft. Die sechs Testpersonen haben im Subtestteil Lautsprache Nomen eine durchschnittliche Leistung von 15,2 gegenüber einer Leistung von 12,1 Punkten der gesamten Stichprobe erzielt. Diese Leistung bringt sie auf die Rangplätze 8, 11, 13, 22, 38 und 47, durchschnittlich auf Platz 23. Im Subtestteil Lautsprache Verben haben die sechs Testpersonen eine durchschnittliche Leistung von 14,8, die gesamte Stichprobe von 11,6 Punkten erzielt. Damit erreichen sie die Leistungsrangplätze 15, 18, 20, 22, 26 und 41, durchschnittlich Platz 24. Im Subtestteil Lautsprache Adjektive haben sie eine durchschnittliche Leistung von 13,7 Punkten gegenüber 10,5 Punkten der gesamten Stichprobe erbracht. Sie befinden sich mit dieser Leistung auf den Rangplätzen 7, 15, 22, 23, 48 und 67, durchschnittlich auf Platz 30. In der gesamten Modalität Lautsprache erzielten die sechs bilingualen Testpersonen eine durchschnittliche Leistung von 43,7 Punkten, die gesamte Stichprobe erzielte 34,1 Punkte. Mit dieser Leistung rangieren die sechs Testpersonen auf den Plätzen 8, 15, 17, 23, 43 und 48, durchschnittlich 26.

Schließlich wurden die Ergebnisse der bilingualen Testpersonen im Gesamttest untersucht. Im gesamten Test erzielten die sechs Testpersonen eine durchschnittliche Leistung von 140 gegenüber 108 Punkten der gesamten Stichprobe. Damit erreichen sie die Rangplätze 8, 10, 12, 14, 24 und 36, durchschnittlich Platz 17.

Die bilingual erzogenen Testpersonen liegen mit ihrer Leistung im Wortschatztest in allen Subtestteilen deutlich über der Leistung der gesamten Stichprobe (vgl. auch Ergebnisse der einfaktoriellen Varianzanalyse in C 1.6). Im Subtestteil Gebärdensprache Nomen erreichen

sie durchschnittlich 4,6 Punkte mehr, im Subtestteil Gebärdensprache Verben 1,0 Punkte mehr, im Subtestteil Gebärdensprache Adjektive 3,4 Punkte mehr als der Durchschnitt der Gesamtstichprobe. Im Subtestteil Schriftsprache Nomen erzielen sie durchschnittlich 5,1 Punkte mehr, im Subtestteil Schriftsprache Verben 5,0 Punkte mehr und im Subtestteil Schriftsprache Adjektive 5,6 Punkte mehr als der Durchschnitt der gesamten Stichprobe. Im Subtestteil Lautsprache Nomen haben sie durchschnittlich 3,1 Punkte mehr, in den Subtestteilen Lautsprache Verben und Adjektive jeweils 3,2 Punkte mehr, als der Durchschnitt der Gesamtstichprobe aufweist. Die auffallendste Leistungssteigerung gegenüber der gesamten Stichprobe erreichen die bilingual erzogenen Testpersonen dabei in der Modalität Schriftsprache mit durchschnittlich 15,5 Punkten über der Leistung der gesamten Stichprobe. (Dieses Ergebnis wird allerdings durch den Umstand relativiert, dass sich die Gruppe der bilingual erzogenen Kinder ausschließlich aus Fünftklässlern zusammensetzt, die in ihrer Sprachentwicklung vor allem den Drittklässlern selbstverständlich voraus sind.) In den Modalitäten Gebärden- und Lautsprache überragen sie die durchschnittliche Leistung der gesamten Stichprobe um 8,4 bzw. 9,6 Punkte.

Damit erreichen die bilingual erzogenen Testpersonen generell eine durchschnittlich deutlich bessere Leistung als die Gesamtstichprobe. Sie befinden sich mit ihrer Leistung in allen Subtests und Subtestteilen sowie im Gesamttest im obersten Leistungsdrittel, in fünf Subtestteilen sowie im Gesamttest rangieren sie sogar im obersten Leistungssechstel.

Daraufhin wurde untersucht, wie gut die Testpersonen, die nach anderen gehörlosenspezifischen Unterrichtskonzepten erzogen werden, in den verschiedenen Sprachmodalitäten abschnitten. Dafür wurden zunächst die Ergebnisse der 36 oral erzogenen Testpersonen analysiert. Die oral erzogenen Testpersonen erzielten im Subtestteil Gebärdensprache Nomen eine durchschnittliche Leistung von 13,0 Punkten, während die gesamte Stichprobe 13,1 erreichte. Die 36 Testpersonen rangieren mit dieser Leistung durchschnittlich auf dem Rangplatz 59. Im Subtestteil Gebärdensprache Verben erreichen sie eine durchschnittliche Leistung von 14,9 Punkten gegenüber einer Leistung von 15,2 Punkten der gesamten Stichprobe und erreichen damit den durchschnittlichen Leistungsrangplatz 63. Im Subtestteil Gebärdensprache Adjektive erzielen sie durchschnittlich 12,7 Punkte im Gegensatz zu 12,4 Punkten der gesamten Stichprobe. Damit rangieren sie auf Platz 53 der Leistungsrangfolge. In der gesamten Modalität Gebärdensprache haben die 36 Testpersonen eine durchschnittliche Leistung von 40,6 Punkten, die Gesamtstichprobe von 40,4 Punkten erreicht. Damit befinden sie sich auf dem durchschnittlichen Rangplatz 57.

Im Subtestteil Schriftsprache Nomen erzielten die 36 oral erzogenen Testpersonen eine durchschnittliche Leistung von 12,8 Punkten, während die Gesamtstichprobe eine Leistung von 11,9 Punkten erreichten. Damit rangieren sie auf dem durchschnittlichen Rangplatz 49.

Im Subtestteil Schriftsprache Verben erreichten sie durchschnittlich 10,5 Punkte, die Gesamtleistung der Stichprobe in diesem Subtestteil lag bei 9,8 Punkten. Die 36 Testpersonen befinden sich damit auf Platz 50 der Leistungsrangfolge. Im Subtestteil Schriftsprache Adjektive erreichten sie durchschnittlich 11,1 Punkte, die gesamte Stichprobe 10,1 Punkte. Diese Leistung bringt die 36 Testpersonen auf Rangplatz 48. Auch in der Gesamtmodalität Schriftsprache befinden sich die 36 Testpersonen mit einer durchschnittlichen Leistung von 34,6 gegenüber einer Leistung der gesamten Stichprobe von 32,0 Punkten auf Rangplatz 48.

Im Subtestteil Lautsprache Nomen haben die 36 Testpersonen eine durchschnittliche Leistung von 13,1 gegenüber 12,1 Punkten der gesamten Stichprobe erzielt. Damit befinden sie sich auf dem durchschnittlichen Rangplatz 46. Im Subtestteil Lautsprache Verben erreichten sie durchschnittlich 12,4 Punkte, während die gesamte Stichprobe durchschnittlich 11,6 Punkte erreichte. Die 36 Testpersonen erreichen damit den durchschnittlichen Leistungsrangplatz 47. Im Subtestteil Lautsprache Adjektive erbrachten sie eine durchschnittliche Leistung von 12,1 Punkte, der Durchschnitt der Gesamtstichprobe erreichte in diesem Subtestteil 10,5 Punkte. Damit rangieren die betreffenden Testpersonen durchschnittlich auf Rangplatz 43. In der gesamten Modalität Lautsprache erzielten die oral erzogenen Testpersonen durchschnittlich 37,6 Punkte, während die gesamte Stichprobe 34,1 Punkte erreichte. Sie befinden sich mit dieser Leistung auf dem durchschnittlichen Rangplatz 44.

Im Gesamtest rangieren diese Testpersonen mit einer durchschnittlichen Leistung von 113,9 Punkten gegenüber 108 Punkten der Gesamtstichprobe auf Rangplatz 45.

Während die oral erzogenen Testpersonen in den Subtestteilen Gebärdensprache Nomen und Verben etwas unter der durchschnittlichen Leistung der gesamten Stichprobe liegen, erreichen sie in den übrigen Subtestteilen durchgehend etwa einen Punkt mehr als der Durchschnitt. Sie erreichen damit Rangplätze, die etwas über dem Mittel liegen.

Entsprechend wurden auch die Ergebnisse der 70 gebärdensorientiert erzogenen Testpersonen untersucht. Die gebärdensorientiert erzogenen Testpersonen erzielten in dem Subtestteil Gebärdensprache Nomen durchschnittlich 12,8 Punkte, während die durchschnittliche Leistung der gesamten Stichprobe bei 13,1 Punkten lag. Sie erreichten damit den durchschnittlichen Leistungsrangplatz 58. Im Subtestteil Gebärdensprache Verben erbrachten sie eine durchschnittliche Leistung von 15,3 Punkten gegenüber einer Leistung von 15,2 Punkten in der Gesamtstichprobe. Mit dieser Leistung liegen sie auf dem durchschnittlichen Rangplatz 54. Im Subtestteil Gebärdensprache Adjektive erreichten die 70 Testpersonen durchschnittlich 11,9 Punkte, die gesamte Stichprobe dagegen 12,4 Punkte. Sie rangieren damit auf dem durchschnittlichen Rangplatz 60. In der gesamten Modalität Gebärdensprache erreichten sie

mit einer Leistung von 40,1 Punkten gegenüber einer Leistung von 40,4 Punkten in der Gesamtstichprobe den durchschnittlichen Rangplatz 58.

Im Subtestteil Schriftsprache Nomen erbrachten die 70 Testpersonen eine Leistung von 11,0 Punkten gegenüber 11,9 Punkten in der gesamten Stichprobe. Damit rangieren sie durchschnittlich auf Rangplatz 63. Im Subtestteil Schriftsprache Verben erreichen sie 9,1 Punkte, während die gesamte Stichprobe 9,8 Punkte erreicht. Damit befinden sich die gebärdenorientiert erzogenen Testpersonen durchschnittlich auf Rangplatz 62. Im Subtestteil Schriftsprache Adjektive erlangen die 70 Testpersonen ebenfalls 9,1 Punkte, die Gesamtstichprobe dagegen 10,1 Punkte. Sie befinden sich mit dieser Leistung auf dem durchschnittlichen Rangplatz 63. In der gesamten Modalität Schriftsprache erzielen diese Testpersonen eine durchschnittliche Leistung von 29,3 gegenüber 34,6 Punkten der Gesamtstichprobe und erreichen damit den durchschnittlichen Rangplatz 63.

Im Subtestteil Lautsprache Nomen erbringen die 70 Testpersonen eine durchschnittliche Leistung von 11,2 Punkten gegenüber 12,1 Punkten der gesamten Stichprobe. Sie rangieren damit auf dem Leistungsrangplatz 63. Im Subtestteil Lautsprache Verben erlangen sie 10,8 Punkte, während die gesamte Stichprobe 11,6 Punkte erreicht. Diese Leistung bringt sie auf den durchschnittlichen Rangplatz 62. Sie erreichen im Subtestteil Lautsprache Adjektive mit einer Leistung von 9,4 Punkten gegenüber 10,5 Punkten der Gesamtstichprobe den durchschnittlichen Rangplatz 64. Auch in der gesamten Modalität Lautsprache erreichen sie mit durchschnittlich 31,5 Punkten im Gegensatz zu 34,1 Punkten der gesamten Stichprobe den Rangplatz 64.

Im Gesamttest haben die 70 gebärdenorientiert erzogenen Testpersonen 102,1 Punkte erzielt, die gesamte Stichprobe dagegen 108 Punkte. Damit rangieren diese Testpersonen im gesamten Test auf dem durchschnittlichen Rangplatz 61.

Die Ergebnisse der Überprüfung I sind in der folgenden Tabelle zusammengefasst:

Tab. 75-C 1.5.1: Testergebnisse der Testpersonen nach Konzepten sortiert				
Subtest(teil)	alle (112)	bilingual (6)	mit Gebärden (70)	oral (36)
GSN	13,1	17,7	12,8	13,0
GSV	15,2	16,2	15,3	14,9
GSA	12,4	15,8	11,9	12,7
GS	40,4	48,8	40,1	40,6

Subtest(teil)	alle (112)	bilingual (6)	mit Gebärden (70)	oral (36)
SSN	11,9	17,0	11,0	12,8
SSV	9,8	14,8	9,1	10,5
SSA	10,1	15,7	9,1	11,1
SS	32,0	47,5	29,3	34,6
LSN	12,1	15,2	11,2	13,1
LSV	11,6	14,8	10,8	12,4
LSA	10,5	13,7	9,4	12,1
LS	34,1	43,7	31,5	37,6
gesamt	108,0	140,0	102,1	113,9

Die Tabelle veranschaulicht, dass die bilingualen Testpersonen in allen Subtests, in jedem Subtestteil sowie demnach im Gesamtergebnis mit jeweils einigem Abstand leistungsmäßig vor denen anderer Konzepte liegen. Die Überprüfung I bestätigt damit uneingeschränkt die vorangestellte Annahme der Vorrangstellung der bilingual erzogenen Schüler gegenüber gleichaltrigen alternativ erzogenen Vergleichspersonen.

1.5.2 Überprüfung II

Weiterhin wurde untersucht, von welchen Testpersonen die zehn leistungsbesten sowie die zehn leistungsschwächsten Rangplätze belegt wurden und nach welchen Konzepten sie unterrichtet werden. In diesem Zusammenhang wurden zunächst die zehn leistungsstärksten Testpersonen der Modalität Gebärdensprache ermittelt. Im Subtestteil Gebärdensprache Nomen haben fünf (83,3%) der sechs bilingual erzogenen, drei (4,3%) der 70 gebärdenorientiert erzogenen und zwei (5,9%) der 34 oral erzogenen Testpersonen die obersten 10 Rangplätze erreicht. Dabei haben sie 17 oder 18 Items richtig beantwortet. Weitere sieben Testpersonen haben ebenfalls 17 Punkte erzielt, unter ihnen waren sechs (8,6%) der gebärdenorientiert und eine (2,9%) der oral erzogenen Testpersonen. Insgesamt haben demnach 83,3% der bilingual, neun (12,9%) der gebärdenorientiert und drei (8,8%) der oral erzogenen Testpersonen eine Leistung von 17 oder 18 Punkten erbracht.

Im Subtestteil Gebärdensprache Verben haben zwei (33,3%) der bilingual, sieben (10,0%) der gebärdenorientiert und eine (2,9%) der oral erzogenen Testpersonen mit einer Leistung von 18 Punkten die obersten 10 Rangplätze erreicht.

Im Subtestteil Gebärdensprache Adjektive sind es zwei (33,3%) der bilingual, drei (4,3%) der gebärdenorientiert und fünf (14,7%) der oral erzogenen Testpersonen, die mit einer Leistung zwischen 16 und 18 Punkten die obersten 10 Rangplätze einnehmen. Weitere zehn

Testpersonen erzielten ebenfalls eine Leistung von 16 Punkten in diesem Subtestteil. Dies waren eine (16,7%) der bilingual, sieben (10,0%) der gebärdenorientiert und zwei (2,9%) der oral erzogenen Testpersonen. Damit haben drei (50%) der bilingual, zehn (14,3%) der gebärdenorientiert und sieben (20,6%) der oral erzogenen Testpersonen eine Leistung von 16 bis 18 Punkten erzielt.

In der gesamten Modalität Gebärdensprache haben zwei (33,3%) der bilingual, fünf (7,1%) der gebärdenorientiert und drei (8,8%) der oral erzogenen Testpersonen eine Bestleistung zwischen 50 und 53 Punkten erreicht. Eine weitere (16,7%) der bilingualen Testpersonen hat ebenfalls 50 Punkte erreicht. Damit haben drei (50%) der bilingual, fünf (7,1%) der gebärdenorientiert und drei (8,8%) der oral erzogenen Testpersonen eine Leistung von 50 bis 53 Punkten in der Modalität Gebärdensprache erzielt.

Auch für die Leistungen in der Schriftsprache wurden die besten 10 ermittelt. Im Subtestteil Schriftsprache Nomen haben zwei (33,3%) der sechs bilingual erzogenen, fünf (7,2%) der 69 gebärdenorientiert erzogenen und drei (8,6%) der 35 oral erzogenen Testpersonen die obersten 10 Rangplätze erreicht. Dabei haben sie alle 18 Items richtig beantwortet. Eine (2,9%) weitere oral erzogene Testperson hat ebenfalls volle Punktzahl erreicht, sodass insgesamt zwei (33,3%) der bilingual, fünf (7,2%) der gebärdenorientiert und vier (11,4%) der oral erzogenen Testpersonen eine Leistung von 18 Punkten erbracht haben.

Im Subtestteil Schriftsprache Verben haben zwei (33,3%) der bilingual, vier (5,9%) der gebärdenorientiert und vier (11,8%) der oral erzogenen Testpersonen mit einer Leistung von 16 bis 18 Punkten die obersten 10 Rangplätze erreicht. Weitere drei Testpersonen erzielten ebenfalls eine Leistung von 16 Punkten in diesem Subtestteil. Dies waren eine (16,7%) der bilingual und zwei (2,9%) der gebärdenorientiert erzogenen Testpersonen. Damit haben drei (50%) der bilingual, sechs (8,8%) der gebärdenorientiert und vier (11,8%) der oral erzogenen Testpersonen eine Leistung von 16 bis 18 Punkten erzielt.

Im Subtestteil Schriftsprache Adjektive sind es zwei (33,3%) der bilingual, vier (5,9%) der gebärdenorientiert und vier (11,8%) der oral erzogenen Testpersonen, die mit einer Leistung von 17 oder 18 Punkten die obersten 10 Rangplätze einnehmen. Weitere sieben Testpersonen erzielten ebenfalls eine Leistung von 17 Punkten in diesem Subtestteil. Dies waren zwei (33,3%) der bilingual, drei (4,4%) der gebärdenorientiert und zwei (5,9%) der oral erzogenen Testpersonen. Damit haben vier (66,7%) der bilingual, sieben (10,3%) der gebärdenorientiert und sechs (17,6%) der oral erzogenen Testpersonen eine Leistung von 17 oder 18 Punkten erzielt.

In der gesamten Modalität Schriftsprache haben zwei (33,3%) der bilingual, vier (5,9%) der gebärdenorientiert und vier (11,8%) der oral erzogenen Testpersonen eine Bestleistung zwischen 50 und 52 Punkten erreicht.

Daraufhin wurde für die Modalität Lautsprache untersucht, welche Testpersonen die obersten 10 Rangplätze belegen. Im Subtestteil Lautsprache Nomen haben eine (16,7%) der bilingual, drei (4,5%) der gebärdenorientiert und sechs (17,6%) der oral erzogenen Testpersonen mit einer Leistung von 17 oder 18 Punkten die obersten 10 Rangplätze erreicht.

Im Subtestteil Lautsprache Verben sind es fünf (7,6%) der gebärdenorientiert und fünf (14,7%) der oral erzogenen Testpersonen, die mit einer Leistung zwischen 16 und 18 Punkten die obersten 10 Rangplätze einnehmen. Weitere sieben Testpersonen erzielten ebenfalls eine Leistung von 16 Punkten in diesem Subtestteil. Dies waren eine (16,7%) der bilingual, fünf (7,6%) der gebärdenorientiert und eine (2,9%) der oral erzogenen Testpersonen. Damit haben eine (16,7%) der bilingual, zehn (15,2%) der gebärdenorientiert und sechs (17,6%) der oral erzogenen Testpersonen eine Leistung von 16 bis 18 Punkten erzielt.

Im Subtestteil Lautsprache Adjektive haben eine (16,7%) der bilingual, zwei (3,0%) der gebärdenorientiert und sieben (20,6%) der oral erzogenen Testpersonen mit einer Leistung von 17 oder 18 Punkten die obersten 10 Rangplätze erreicht.

In der gesamten Modalität Lautsprache haben eine (16,7%) der bilingual, drei (4,5%) der gebärdenorientiert und sechs (17,6%) der oral erzogenen Testpersonen eine Bestleistung zwischen 49 und 53 Punkten erreicht.

Schließlich wurden auch für die Leistungen im Gesamttest die Top 10 ermittelt. Im gesamten Wortschatztest haben zwei (33,3%) der bilingual, vier (6,3%) der gebärdenorientiert und vier (12,9%) der oral erzogenen Testpersonen mit einer Leistung von 148 bis 155 Punkten die obersten 10 Rangplätze belegt.

Entsprechend wurden auch die Testpersonen ermittelt, die die jeweils 10 untersten Leistungsrangplätze belegen. Diese untersten Leistungsrangplätze sind in den meisten Fällen von Testpersonen gebärdenorientierter Konzepte und nie von bilingual erzogenen Testpersonen belegt. Im Subtestteil Gebärdensprache Nomen rangieren vier, im Subtestteil Gebärdensprache Verben drei und in der Gesamtmodalität Gebärdensprache zwei Testpersonen oral orientierter Konzepte auf den 10 untersten Rangplätzen. Im Subtestteil Schriftsprache Nomen rangiert eine, im Subtestteil Schriftsprache Verben zwei und in der Gesamtmodalität Schriftsprache eine Testperson oral orientierter Konzepte auf den 10 untersten Rangplätzen. Im Subtestteil Lautsprache Nomen handelt es sich um eine, im Subtestteil Lautsprache

Adjektive um zwei und in der Gesamtmodalität Lautsprache um eine Testperson oral orientierter Konzepte, die auf den 10 untersten Rangplätzen rangieren. Auf den untersten Leistungsrangplätzen des Gesamttests befindet sich eine Testperson, die nach oral orientiertem Konzept erzogen wird.

Erwartungsgemäß befindet sich ein Großteil der Testpersonen, die mit ihrer Leistung auf den obersten Rangplätzen liegen, in den Klassenstufen 4 und 5, während die Testpersonen auf den untersten Rangplätzen hauptsächlich dritten Klassen entstammen:

Tab. 76-C 1.5.2: Klassenzugehörigkeit der besten und schlechtesten Testpersonen in allen Subtestteilen						
Subtest (-teil)	Top 10			Flop 10		
	3. Klasse	4. Klasse	5. Klasse	3. Klasse	4. Klasse	5. Klasse
GSN	-	2	8	8	-	2
GSV	1	5	4	7	1	2
GSA	1	6	3	7	2	1
GS	1	5	4	7	1	2
SSN	1	5	4	9	1	-
SSV	-	5	5	8	1	1
SSA	-	6	4	8	1	1
SS	-	6	4	10	-	-
LSN	-	7	3	6	2	2
LSV	-	7	3	6	3	1
LSA	-	7	3	6	2	2
LS	-	6	4	5	2	3
gesamt	-	5	5	8	1	1

Mit einer Ausnahme im Subtestteil Lautsprache Verben rangiert immer mindestens eine der sechs bilingual erzogenen Testpersonen unter den zehn leistungsstärksten eines Subtestteils, im Subtestteil Gebärdensprache Nomen befinden sich sogar fünf der sechs Testpersonen auf den obersten zehn Leistungsrangplätzen. Mit ihren durchschnittlichen Leistungsrangplätzen bewegen sie sich stets im obersten Leistungsdrittel. Im Subtestteil Lautsprache Adjektive bewegen sie sich im obersten Leistungsviertel, in den Subtestteilen Lautsprache Nomen und Adjektive im obersten Leistungsfünftel, im Subtestteil Schriftsprache Adjektive im obersten Leistungssechstel, in den Subtestteilen Gebärdensprache Adjektive und Schriftsprache Verben im obersten Leistungssiebtel, im Subtestteil Schriftsprache Nomen im obersten Leistungsaachtel und im Subtestteil Gebärdensprache Nomen sogar im obersten Dreizehntel.

1.6 Korrelationen

Wie bereits aus den Ergebnissen der Validierung hervorgeht, bestehen mehr oder weniger signifikante Zusammenhänge zwischen der Leistung einer Testperson im Wortschatztest und bestimmten schulischen und familiären Sozialisationsbedingungen. Über die bisher in diesem Zusammenhang angewendete Überprüfungsmethode der Korrelation soll mit Hilfe von Varianzanalysen die Signifikanz des Einflusses der Faktoren 'gehörlosenspezifische Konzepte', 'Klassenstufe', 'Intelligenzquotient' sowie 'Hörstatus der Eltern' untersucht werden. Es wird angenommen, dass hohe Signifikanz in allen Fällen besteht.

Es liegt in der Natur der Sache begründet, dass Begleitumstände der Testpersonen, die aufgeführten Faktoren, Einfluss auf die sprachlichen Leistungen von Kindern ausüben. Weiterhin kann angenommen werden, dass diese möglichen Einflüsse die Objektivität der Testergebnisse beeinträchtigen oder auch den Horizont der Aussagekraft der Ergebnisse erweitern könnten. Daher sind entsprechende Überprüfungen ebenso notwendig wie interessant.

1.6.1 Überprüfung I

Um diese Zusammenhänge differenziert zu überprüfen, wurden mit Hilfe einfaktorieller Varianzanalysen die Gruppenmittelwerte und Streuungen der unterschiedlichen Faktoren überprüft. Die zugrunde liegende und auf Wahrscheinlichkeit zu prüfende Nullhypothese behauptet, dass alle Gruppenmittelwerte der Variablen gleich sind, dass also in der Grundgesamtheit keine Gruppendifferenzen vorliegen. Gemäß der Hypothese gilt jedoch zu zeigen, dass signifikante Gruppenunterschiede in den einflussnehmenden Faktoren 'gehörlosenspezifische Konzepte', 'Klassenstufe', 'Intelligenzquotient' sowie 'Hörstatus der Eltern' bestehen. Dafür wurde mit Hilfe des F-Tests die Nullhypothese überprüft, ob die zu vergleichenden Stichproben aus Grundgesamtheiten mit gleichen Varianzen stammen, d. h. ob mögliche Varianzunterschiede nur stichprobenbedingt bzw. zufällig sind. Weist diese Untersuchung hohe Signifikanz aus, so kann die Hypothese, es bestünden keine linearen Zusammenhänge zwischen der passiven Wortschatzentwicklung und dem jeweiligen Faktor zurückgewiesen werden. In Fällen mit mindestens dreigliedriger Abstufung wurde zusätzlich mit der Student-Newman-Keuls-Prozedur untersucht, ob signifikante Unterschiede in den Gruppenmittelwerten bestehen.

Zunächst wurde untersucht, ob sich die Leistungsmittelwerte in den zwei Klassenstufe 3 und 4 (+5) signifikant voneinander unterscheiden. Alle Subtestteile in der Modalität Gebärdensprache ($p_{GSN}=.0002$, $p_{GSV}=.0001$, $p_{GSA}=.0004$, $p_{GS}=.0000$), alle Subtestteile in der Modalität Schriftsprache ($p_{SSN}=.0000$, $p_{SSV}=.0000$, $p_{SSA}=.0000$, $p_{SS}=.0000$), alle Subtestteile in der Modalität Lautsprache ($p_{LSN}=.0013$, $p_{LSV}=.0023$, $p_{LSA}=.0003$, $p_{LS}=.0003$) sowie der Gesamttest

($p_{Ges}=.0000$) weisen sehr signifikante Unterschiede auf. Die Mittelwerte in aufsteigender Reihenfolge sind in der folgenden Tabelle für jeden Subtest(teil) und für den Gesamttest dargestellt, Angaben zu signifikanten Unterschieden nach Duncan und Student-Newman-Keuls sind bei einer zweistufigen Variablen (3./4. Klasse) nicht möglich:

Tab. 77-C 1.6.1: Leistungsmittelwerte der Klassenstufen							
GSN	Klasse	GSV	Klasse	GSA	Klasse	GS	Klasse
11,9	3	14,3	3	11,2	3	37,3	3
13,9	4	15,9	4	13,2	4	43,0	4
SSN	Klasse	SSV	Klasse	SSA	Klasse	SS	Klasse
9,7	3	7,6	3	7,8	3	25,2	3
13,5	4	11,3	4	11,7	4	36,5	4
LSN	Klasse	LSV	Klasse	LSA	Klasse	LS	Klasse
10,5	3	10,2	3	8,5	3	29,2	3
13,0	4	12,4	4	11,7	4	37,0	4
Ges.	Klasse						
93,0	3						
116,7	4						

Der Mittelwert der Klasse 3 liegt tatsächlich in allen Subtestteilen von 1,6 bis 3,9 Punkten, durchschnittlich 2,8 Punkte unter dem der Klasse 4. Die Unterschiede sind so signifikant, dass die Hypothese, es bestünden keine linearen Zusammenhänge zwischen der passiven Wortschatzentwicklung und der Klassenstufe, zurückgewiesen wird. Damit fallen Zusammenhänge bzgl. der Klassenstufe im Sinne der Annahme sowie der Konstruktbeschreibung aus. Es kristallisiert sich heraus, dass die Kenntnis des passiven Wortschatzes von der dritten zur vierten Klasse signifikant ansteigt.

In einer weiteren Varianzanalyse wurde untersucht, ob sich die Leistungsmittelwerte in den drei verschiedenen gehörlosenspezifischen Konzepten signifikant voneinander unterscheiden. Bis auf den Subtestteil Gebärdensprache Verben ($p_{GSV}=.3681$) weisen alle Subtestteile in der Modalität Gebärdensprache ($p_{GSN}=.0038$, $p_{GSA}=.0037$, $p_{GS}=.0061$), alle Subtestteile in der Modalität Schriftsprache ($p_{SSN}=.0009$, $p_{SSV}=.0060$, $p_{SSA}=.0011$, $p_{SS}=.0009$), alle Subtestteile in der Modalität Lautsprache ($p_{LSN}=.0080$, $p_{LSV}=.0092$, $p_{LSA}=.0023$, $p_{LS}=.0023$) sowie der Gesamttest ($p_{Ges}=.0019$) sehr signifikante Unterschiede auf. Die Mittelwerte in aufsteigender Ordnung sowie signifikante Unterschiede auf dem 0,05%-Niveau (x) sind in der folgenden Tabelle für jeden Subtest(teil) und für den Gesamttest dargestellt.

Tab. 78-C 1.6.1: Leistungsmittelwerte und signifikante Unterschiede der Konzepte				
GSN	Konzept	mit Gebärden	oral	bilingual
12,8	mit Gebärden			
13,0	oral			
16,8	bilingual	x	x	

GSV	Konzept			
14,9	oral			
15,3	mit Gebärden			
16,2	bilingual			
GSA	Konzept	mit Gebärden	oral	bilingual
11,9	mit Gebärden			
12,7	oral			
15,8	bilingual	x	x	
GS	Konzept	mit Gebärden	oral	bilingual
40,1	mit Gebärden			
40,6	oral			
48,8	bilingual	x	x	
SSN	Konzept	mit Gebärden	oral	bilingual
11,0	mit Gebärden			
12,8	oral	x		
17,0	bilingual	x	x	
SSV	Konzept	mit Gebärden	oral	bilingual
9,1	mit Gebärden			
10,5	oral			
14,8	bilingual	x	x	
SSA	Konzept	mit Gebärden	oral	bilingual
9,1	mit Gebärden			
11,1	oral	x		
15,7	bilingual	x	x	
SS	Konzept	mit Gebärden	oral	bilingual
29,3	mit Gebärden			
34,6	oral	x		
47,5	bilingual	x	x	
LSN	Konzept	mit Gebärden	oral	bilingual
11,2	mit Gebärden			
13,1	oral	x		
15,2	bilingual	x		
LSV	Konzept	mit Gebärden	oral	bilingual
10,8	mit Gebärden			
12,4	oral	x		
14,8	bilingual	x		
LSA	Konzept	mit Gebärden	oral	bilingual
9,4	mit Gebärden			
12,1	oral	x		
13,7	bilingual	x		
LS	Konzept	mit Gebärden	oral	bilingual
31,5	mit Gebärden			
37,6	oral	x		
43,7	bilingual	x		
Ges.	Konzept	mit Gebärden	oral	bilingual
102,1	mit Gebärden			
113,9	oral	x		
140,0	bilingual	x	x	

Hieraus geht hervor, dass der Mittelwert der bilingual erzogenen Kinder immer von 1,6 bis 4,6 Punkten, durchschnittlich 3,0 Punkte über dem der oral erzogenen Kinder liegt. Der Mittelwert der bilingual erzogenen Kinder liegt immer von 0,9 bis 6,6 Punkten, durchschnittlich 4,4 Punkte über dem der gebärdenorientiert erzogenen Kinder. Der Mittelwert der oral erzogenen Kinder liegt mit einer Ausnahme im Subtestteil Gebärdensprache Verben in allen Subtestteilen von 0,2 bis 2,7 Punkten, durchschnittlich 1,6 Punkte über dem der gebärdenorientiert erzogenen Kinder. Diese Zusammenhänge werden in der folgenden Tabelle noch einmal zusammengefasst. In den Spalten sind jeweils die positiven (+) oder negativen (-) Punktdifferenzen von einem Konzept zum anderen aufgeführt.

Tab. 79-C 1.6.1: Differenzen der Konzeptmittelwerte			
	Diff. von 'bilingual' zu 'oral'	Diff. von 'bilingual' zu 'mit Gebärden'	Diff. von 'oral' zu 'mit Gebärden'
GSN	+ 3,8	+ 4,0	- 0,2
GSV	+ 1,3	+ 0,9	+ 0,4
GSA	+ 3,1	+ 3,9	- 0,8
SSN	+ 4,2	+ 6,0	- 1,8
SSV	+ 4,3	+ 5,7	- 1,4
SSA	+ 4,6	+ 6,6	- 2,0
LSN	+ 2,1	+ 4,0	- 1,9
LSV	+ 2,4	+ 4,0	- 1,6
LSA	+ 1,6	+ 4,3	- 2,7

Die angegebenen Unterschiede sind so signifikant, dass die Hypothese, es bestünden keine linearen Zusammenhänge zwischen der passiven Wortschatzentwicklung und dem angewandten Erziehungskonzept, in allen Fällen mit Ausnahme im Subtestteil GSV zurückgewiesen wird. Auch die Student-Newman-Keuls-Prozedur weist in den übrigen acht Subtestteilen und im Gesamttest eindeutige Zusammenhänge aus. Diese fallen im Sinne der Annahme sowie der Konstruktsbeschreibung aus. Es bestätigt sich die Annahme, dass sich der passive Wortschatz bei bilingual erzogenen Kindern in allen Modalitäten signifikant besser entwickelt als bei oral- oder gebärdenorientierter Erziehung. Signifikante Leistungsunterschiede zwischen Testpersonen der beiden letztgenannten Konzepte sind eindeutig in der Modalität Lautsprache, teils in der Modalität Schriftsprache und nicht in der Modalität Gebärdensprache erkennbar.

Weiterhin wurde die Signifikanz der Differenzen der Leistungsmittelwerte in vier verschiedenen Intelligenzgruppen untersucht. (Hierfür wurde für die Variable 'Intelligenz' eine Einteilung in jeweils vier gleich große Gruppen, in leistungsgemäß aufsteigender Reihenfolge vorgenommen). Alle Subtestteile der Modalität Gebärdensprache ($p_{GSN}=.0343$, $p_{GSV}=.0417$, $p_{GSA}=.0000$, $p_{GS}=.0004$), alle Subtestteile der Modalität Schriftsprache ($p_{SSN}=.0060$, $p_{SSV}=.0014$, $p_{SSA}=.0001$, $p_{SS}=.0005$), bis auf Subtestteil Lautsprache Verben ($p_{LSV}=.0683$)

alle Subtestteile der Modalität Lautsprache ($p_{LSN}=.0180$, $p_{LSA}=.0008$, $p_{LS}=.0092$) sowie der Gesamttest ($p_{Ges.}=.0035$) weisen signifikante Unterschiede auf. Die Mittelwerte in aufsteigender Ordnung sowie signifikante Unterschiede auf dem 0,05%-Niveau (x) sind in der folgenden Tabelle für jeden Subtest(teil) und für den Gesamttest dargestellt.

Tab. 80-C 1.6.1: Leistungsmittelwerte und signifikante Unterschiede der IQ-Gruppen					
GSN	IQ-Gruppe	1	2	3	4
12,0	1				
12,8	2				
13,6	3	x			
14,1	4	x			
GSV	IQ-Gruppe	1	2	3	4
14,6	1				
15,1	2				
15,2	3				
16,2	4	x			
GSA	IQ-Gruppe	1	2	3	4
11,3	1				
11,5	2				
12,3	3				
14,6	4	x	x	x	
GS	IQ-Gruppe	1	2	3	4
37,9	1				
39,4	2				
41,1	3				
45,0	4	x	x	x	
SSN	IQ-Gruppe	1	2	3	4
10,7	1				
11,3	2				
11,5	3				
14,4	4	x	x	x	
SSV	IQ-Gruppe	1	2	3	4
8,2	1				
8,3	2				
9,3	3				
12,7	4	x	x	x	
SSA	IQ-Gruppe	1	2	3	4
7,7	1				
9,7	2				
10,0	3				
13,3	4	x	x	x	
SS	IQ-Gruppe	1	2	3	4
26,7	1				
30,3	2				
31,0	3				
40,4	4	x	x	x	

LSN	IQ-Gruppe	1	2	3	4
10,4	1				
11,7	2				
12,9	3	x			
13,4	4	x			
LSV	IQ-Gruppe				
11,3	1				
10,8	2				
11,0	3				
13,2	4				
LSA	IQ-Gruppe	1	2	3	4
8,5	1				
9,5	2				
11,0	3	x			
13,1	4	x	x		
LS	IQ-Gruppe	1	2	3	4
30,2	1				
31,9	2				
34,9	3				
39,7	4	x	x		
Ges.	IQ-Gruppe	1	2	3	4
96,5	1				
103,8	2				
107,5	3				
124,2	4	x	x	x	

Hieraus geht hervor, dass der Mittelwert der IQ-Gruppe 1 außer bei LSV von 0,2 bis 2,0 Punkten durchschnittlich 0,9 Punkte unter dem der IQ-Gruppe 2 liegt. Der Mittelwert der IQ-Gruppe 1 liegt außer bei LSV von 0,6 bis 2,5 Punkten, durchschnittlich 1,5 Punkte unter dem der IQ-Gruppe 3. Der Mittelwert der IQ-Gruppe 1 liegt tatsächlich immer von 1,6 bis 5,6 Punkten, durchschnittlich 3,4 Punkte unter dem der IQ-Gruppe 4. Der Mittelwert der IQ-Gruppe 2 liegt tatsächlich immer von 0,1 bis 1,5 Punkten, durchschnittlich 0,6 Punkte unter dem der IQ-Gruppe 3. Der Mittelwert der IQ-Gruppe 2 liegt tatsächlich immer von 1,1 bis 3,6 Punkten, durchschnittlich 2,6 Punkte unter dem der IQ-Gruppe 4. Der Mittelwert der IQ-Gruppe 3 liegt tatsächlich immer von 0,5 bis 3,4 Punkten, durchschnittlich 1,8 Punkte unter dem der IQ-Gruppe 4. Diese Zusammenhänge werden in der folgenden Tabelle noch einmal zusammengefasst. In den Spalten sind jeweils die positiven (+) oder negativen (-) Punktdifferenzen von einer IQ-Gruppe zur anderen aufgeführt.

Tab. 81-C 1.6.1: Differenzen der IQ-Gruppenmittelwerte						
	Diff. von 1 zu 2	Diff. von 1 zu 3	Diff. von 1 zu 4	Diff. von 2 zu 3	Diff. von 2 zu 4	Diff. von 3 zu 4
GSN	+ 0,8	+ 1,6	+ 2,1	+ 0,6	+ 1,3	+ 0,5
GSV	+ 0,5	+ 0,6	+ 1,6	+ 0,1	+ 1,1	+ 1,0
GSA	+ 0,2	+ 1,0	+ 3,3	+ 0,8	+ 3,1	+ 2,3

	Diff. von 1 zu 2	Diff. von 1 zu 3	Diff. von 1 zu 4	Diff. von 2 zu 3	Diff. von 2 zu 4	Diff. von 3 zu 4
SSN	+ 0,6	+ 0,8	+ 3,7	+ 0,2	+ 3,1	+ 2,9
SSV	+ 1,0	+ 1,1	+ 4,5	+ 0,1	+ 3,5	+ 3,4
SSA	+ 2,0	+ 2,2	+ 5,6	+ 0,2	+ 3,6	+ 3,4
LSN	+ 1,2	+ 2,5	+ 3,0	+ 1,3	+ 1,8	+ 0,5
LSV	- 0,5	- 0,3	+ 1,9	+ 0,2	+ 2,4	+ 2,2
LSA	+ 1,0	+ 2,5	+ 4,6	+ 1,5	+ 3,6	+ 2,1

Diese Unterschiede sind so signifikant, dass die Hypothese, es bestünden keine linearen Zusammenhänge zwischen der passiven Wortschatzentwicklung und dem Intelligenzniveau, in allen Fällen mit Ausnahme von Subtestteil LSV zurückgewiesen wird. Auch die Student-Newman-Keuls-Prozedur weist in den übrigen acht Subtestteilen und im Gesamttest eindeutige Zusammenhänge aus. Diese fallen im Sinne der Annahme sowie der Konstruktsbeschreibung aus. Es stellt sich heraus, dass sich der passive Wortschatz bei höherem Intelligenzniveau signifikant besser entwickelt.

In der letzten einfaktoriellen Varianzanalyse wurde untersucht, ob sich die Leistungsmittelwerte in den zwei Testpersonengruppen mit hörenden und mit gehörlosen Eltern signifikant voneinander unterscheiden. Alle Subtestteile der Modalität Gebärdensprache ($p_{GSN}=.0001$, $p_{GSV}=.0390$, $p_{GSA}=.0001$, $p_{GS}=.0000$), alle Subtestteile der Modalität Schriftsprache ($p_{SSN}=.0011$, $p_{SSV}=.0012$, $p_{SSA}=.0001$, $p_{SS}=.0002$), alle Subtestteile der Modalität Lautsprache ($p_{LSN}=.0064$, $p_{LSV}=.0127$, $p_{LSA}=.0019$, $p_{LS}=.0024$) sowie der Gesamttest ($p_{Ges.}=.0002$) weisen signifikante Unterschiede auf. Die Mittelwerte in aufsteigender Reihenfolge sind in den folgenden Tabellen für jeden Subtest(teil) und für den Gesamttest dargestellt, Angaben zu signifikanten Unterschieden nach Duncan und Student-Newman-Keuls sind bei einer zweistufigen Variablen (hörend/gehörlos) nicht möglich:

Tab. 82-C 1.6.1: Leistungsmittelwerte der Testpersonen mit hörenden / gehörlosen Eltern							
GSN	Hörstatus	GSV	Hörstatus	GSA	Hörstatus	GS	Hörstatus
12,8	hörend	15,1	hörend	12,0	hörend	39,9	hörend
16,5	gehörlos	16,6	gehörlos	15,7	gehörlos	48,8	gehörlos
SSN	Hörstatus	SSV	Hörstatus	SSA	Hörstatus	SS	Hörstatus
11,5	hörend	9,4	hörend	9,6	hörend	30,6	hörend
16,0	gehörlos	14,2	gehörlos	15,5	gehörlos	45,7	gehörlos
LSN	Hörstatus	LSV	Hörstatus	LSA	Hörstatus	LS	Hörstatus
11,7	hörend	11,3	hörend	10,1	hörend	33,1	hörend
15,2	gehörlos	14,3	gehörlos	14,6	gehörlos	44,1	gehörlos
Ges.	Hörstatus						
104,6	hörend						
138,6	gehörlos						

Der Mittelwert der gehörlosen Kinder gehörloser Eltern liegt tatsächlich immer von 1,5 bis 5,9 Punkten, durchschnittlich 3,9 Punkte über dem der gehörlosen Kinder hörender Eltern. Die Unterschiede sind so signifikant, dass die Hypothese, es bestünden keine linearen Zusammenhänge zwischen der passiven Wortschatzentwicklung und dem Hörstatus der Eltern, zurückgewiesen wird. Diese Zusammenhänge bzgl. des Hörstatus der Eltern fallen im Sinne der Annahme sowie der Konstruktbeschreibung aus. Es kristallisiert sich heraus, dass sich der passive Wortschatz in der Grundschule bei gehörlosen Kindern gehörloser Eltern in allen Modalitäten signifikant besser entwickelt als bei gehörlosen Kindern hörender Eltern.

Es bestätigt sich, dass die Faktoren 'gehörlosenspezifische Konzepte', 'Klassenstufe', 'Intelligenzquotient' sowie 'Hörstatus der Eltern' die passive Wortschatzentwicklung signifikant beeinflussen. Den wahrscheinlichsten Einzeleinfluss übt dabei die Variable 'Klasse' ($p=.0000$) aus, dicht gefolgt von der Variablen 'Hörstatus der Eltern' ($p=.0002$).

1.6.2 Überprüfung II

In einer einfach mehrfaktoriellen Varianzanalyse wurde darüber hinaus untersucht, ob Kombinationen von Faktoren signifikanten Einfluss auf die Ergebnisse des Wortschatztests nehmen und ob eine globale Signifikanz der Haupteffekte besteht.

Die einzelnen Gruppen in der Stichprobe weisen gemäß den Erwartungen und entgegen der Nullhypothese unterschiedliche Mittelwerte für den Wortschatztest auf. Ob sich diese Mittelwertdifferenzen zufällig durch die Zusammenstellung der Stichprobe ergeben haben, oder ob die Gruppen auch in der Grundgesamtheit unterschiedliche Mittelwerte besitzen, wurde mit Hilfe des F-Tests zum Gesamteinfluss aller Variablen überprüft. Dabei zeigte sich in allen Subtestteilen sowie im Gesamttest ein sehr signifikanter Einfluss ($p=.000$).

Zusätzlich wurden die Wirkungen einzelner Variablen auf das Testergebnis einander gegenübergestellt. Dabei wurden zwei Variablenkombinationen gewählt: Zum einen wurden die vier Variablen 'Klassenstufe', 'gehörlosenspezifische Konzepte', 'Hörstatus der Eltern' sowie 'Intelligenzquotient' untersucht, die bereits im Rahmen der Validierung signifikanten Einfluss aufwiesen, zum anderen wurde als fünfte Variable der Faktor 'Wohnort (Schule)' hinzugenommen, der erst im Zusammenhang mit der Annahmenuntersuchung an Bedeutung gewonnen hat.

Selbstverständlich soll mit diesen Aussagen (wie auch in C 1.7 und C 1.8) über Leistungen in bestimmten Schulen keine Bewertung vorgenommen werden. Abweichungen im Leistungsniveau sind auf einen Selektionseffekt der Stichprobe und damit – zu großen Anteilen über das zu messende Merkmal hinausgehend – auf Zufälligkeiten zurückzuführen.

Tab. 83a-C 1.6.2: Einfluss von Faktorenkombinationen mit 4 Variablen										
4 Variablen	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA	ges.
p Klasse	.007	.002	.024	.000	.001	.003	.021	.012	.007	.001
p Konzept	.002	.333	.002	.001	.011	.001	.007	.008	.002	.001
p Eltern	.000	.116	.000	.000	.000	.000	.002	.002	.000	.000
p IQ	.352	.199	.001	.175	.046	.014	.320	.154	.061	.106

Die drei Einflussgrößen ‘Klassenstufe’, ‘Erziehungskonzept’ und ‘Hörstatus der Eltern’ verursachen mit einer Ausnahme im Subtestteil GSV in allen Subtestteilen signifikante oder sogar sehr signifikante Mittelwertdifferenzen. Auch für den Gesamttest ist ein sehr signifikanter Zusammenhang zwischen dem Testergebnis und den drei Variablen erkennbar. Den höchstwahrscheinlichen Einfluss übt der Faktor ‘Hörstatus der Eltern’ vor den Faktoren ‘Klassenstufe’ und ‘Konzept’ aus. Eine etwas größere Irrtumswahrscheinlichkeit besteht bei der Annahme der Hypothese, es bestünden signifikante Unterschiede in der Wortschatzentwicklung je nach Intelligenzentwicklung.

Tab. 83b-C 1.6.2: Einfluss von Faktorenkombinationen mit 5 Variablen										
5 Variablen	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA	ges.
p Klasse	.042	.009	.443	.020	.092	.176	.878	.559	.177	.159
p Konzept	.030	.359	.000	.000	.000	.000	.000	.000	.000	.000
p Eltern	.000	.139	.000	.000	.000	.000	.002	.001	.000	.000
p IQ	.337	.247	.001	.170	.061	.015	.338	.069	.083	.040
p Schule	.010	.005	.000	.006	.002	.027	.000	.000	.000	.000

Durch Hinzunahme einer fünften Variablen verändern sich die Zusammenhänge zwischen den übrigen Faktoren und dem Testergebnis. Die drei Einflussgrößen ‘Erziehungskonzept’, ‘Hörstatus der Eltern’ und ‘Schule’ verursachen mit einer Ausnahme im Subtestteil GSV in

allen Subtestteilen signifikante oder sogar sehr signifikante Mittelwertdifferenzen. Auch für den Gesamttest ist ein sehr signifikanter Zusammenhang zwischen dem Testergebnis und den drei Variablen erkennbar. Die Variable 'Intelligenzniveau', die auch bei nur vier Variablen die größte Irrtumswahrscheinlichkeit birgt, weist mit einer Wahrscheinlichkeit von .040 im Gesamttest immerhin einen signifikanten Einfluss auf. Eine etwas größere Irrtumswahrscheinlichkeit ($p=.159$) besteht in dieser Variablenkombination bei der Annahme der Hypothese, es bestünden signifikante Unterschiede in der Wortschatzentwicklung je nach Klassenstufe.

In Form einer multiplen Klassifikationsanalyse wurde anhand spezieller Koeffizienten die Stärke des Zusammenhangs zwischen einem Faktor und der Testleistung geprüft. Neben den schon bekannten Korrelationen lassen sich daraus eventuelle Wechselwirkungen zwischen den Faktoren ablesen. Variablenkombinationen ließen jedoch keinen signifikanten Einfluss erkennen.

Demnach bestehen sehr signifikante Zusammenhänge zwischen der Variablen 'Konzept' ($p=.0019$) und der passiven Wortschatzentwicklung sowie zwischen der Variablen 'Intelligenzniveau' ($p=.0035$) und der passiven Wortschatzentwicklung. Signifikante Mittelwertunterschiede in der Testleistung werden zudem durch die Variable 'Wohnort (Schule)' ($p=.0266$) verursacht.

1.7 Typische Profilverläufe

Es wird angenommen, dass es typische Profilverläufe innerhalb der Teilleistungen der Subtests gibt, die besonders häufig vorkommen. Anhand solcher Häufungen von Verläufen könnten sich Tendenzen abzeichnen, die die Wissensstruktur der Testpersonen differenzierter erhellt.

Selbstverständlich soll mit diesen Aussagen (wie auch in C 1.62 und C 1.8) über Leistungen in bestimmten Schulen keine Bewertung vorgenommen werden. Abweichungen im Leistungsniveau sind auf einen Selektionseffekt der Stichprobe und damit – zu großen Anteilen über das zu messende Merkmal hinausgehend – auf Zufälligkeiten zurückzuführen.

1.7.1 Überprüfung I

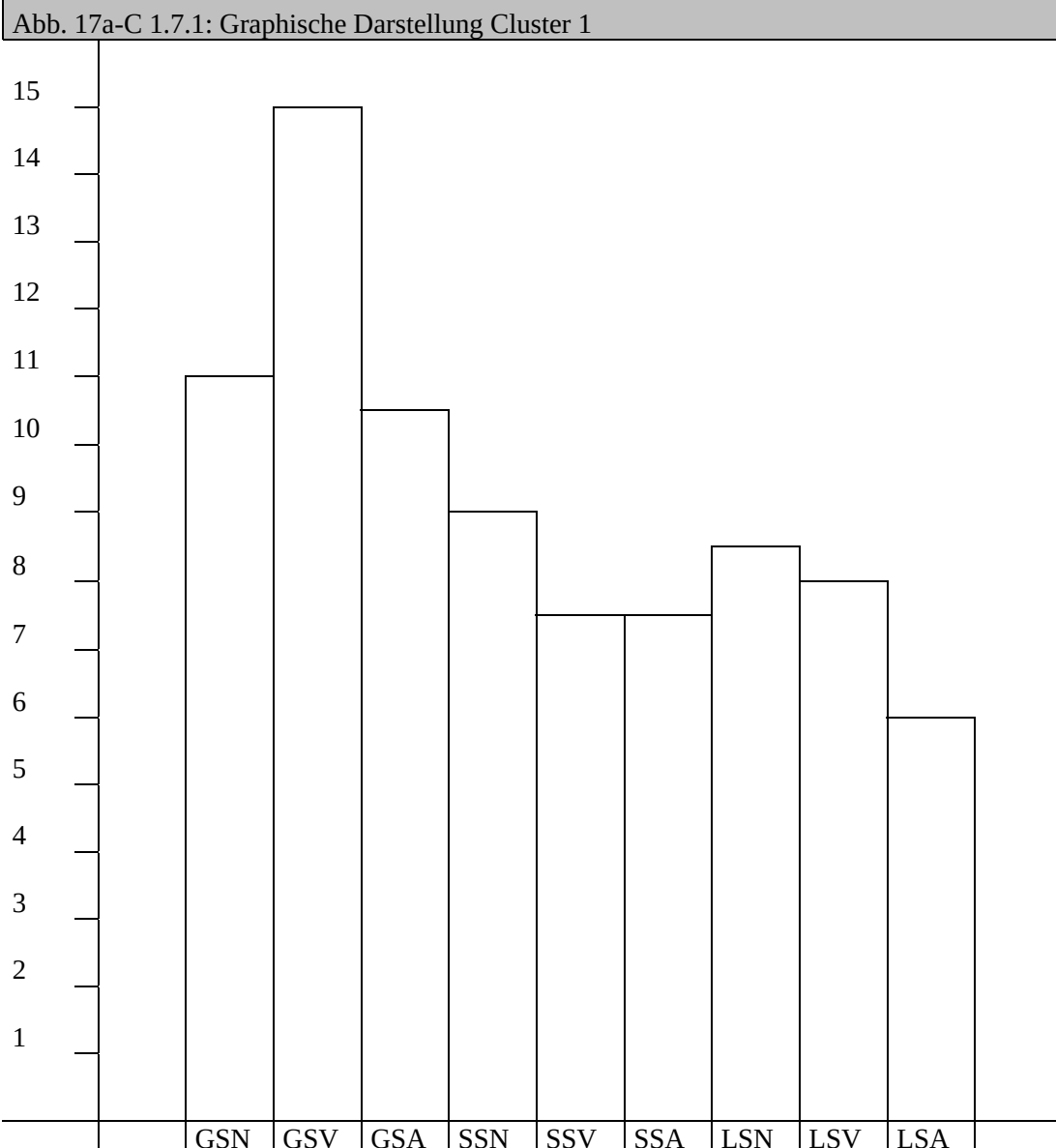
Mit Hilfe des Verfahrens der Clusteranalyse wurden Gruppen von Testpersonen gebildet, die möglichst ähnliche Variablenausprägungen der neun Subtestteile aufweisen. Dafür wurde zunächst im Rahmen einer hierarchischen Clusteranalyse festgestellt, dass in diesem Fall

eine Lösung mit sechs Clustern optimal ist. Dann wurden in einer weiteren Clusteranalyse für hohe Fallzahlen eine Zusammenfassung des Iterationsverlaufs und endgültige Clusterzentren untersucht.

An der Untersuchung haben die 101 Testpersonen teilgenommen, die alle Subtestteile bearbeitet haben. Sie verteilen sich auf die sechs Cluster wie folgt:

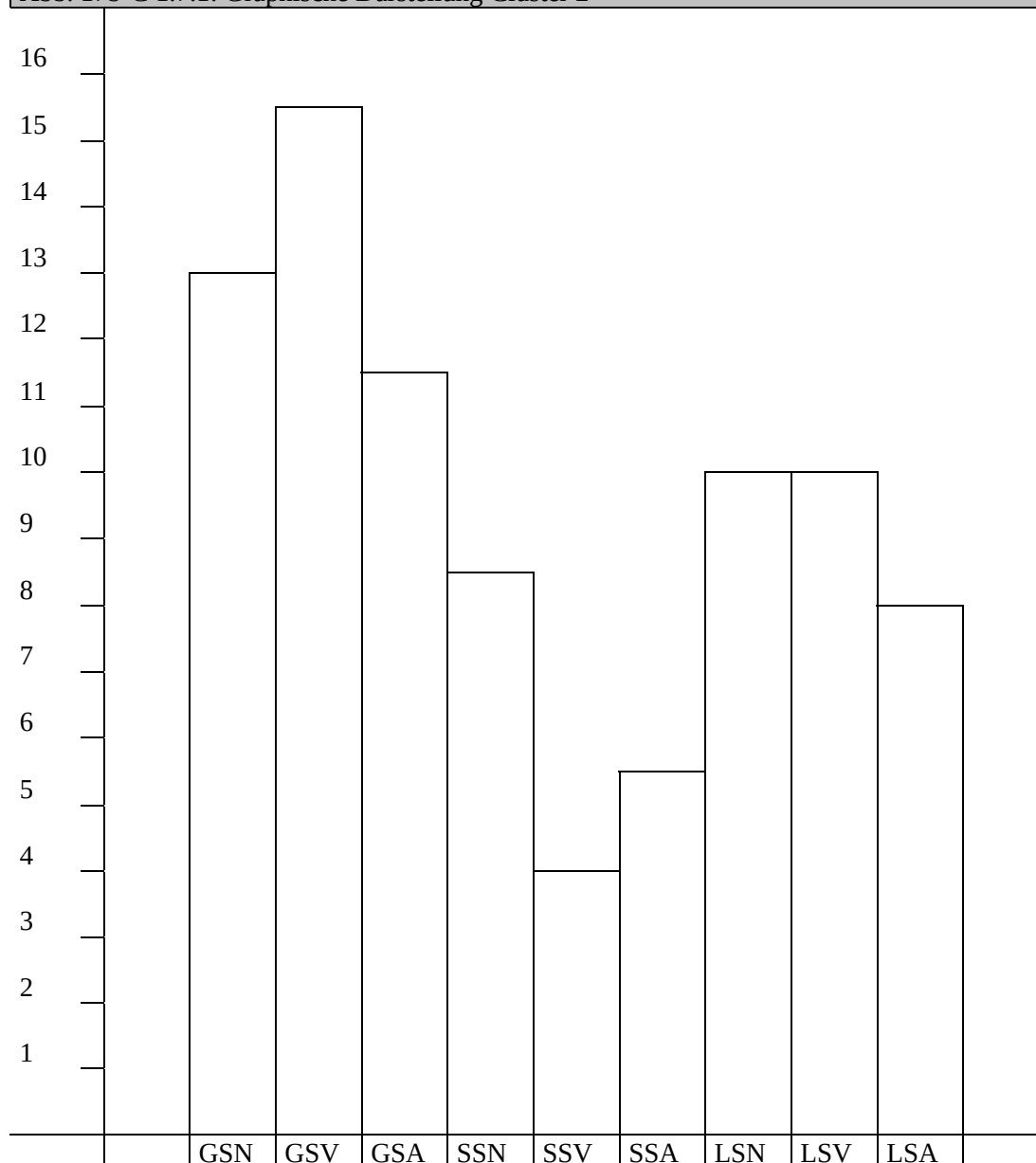
Tab. 84-C 1.7.1: Verteilung der Testpersonen auf sechs Cluster						
Cluster	1	2	3	4	5	6
Anzahl	18	11	29	23	9	11

Die Cluster werden im Folgenden graphisch dargestellt und ihre Zusammenstellung kurz analysiert:



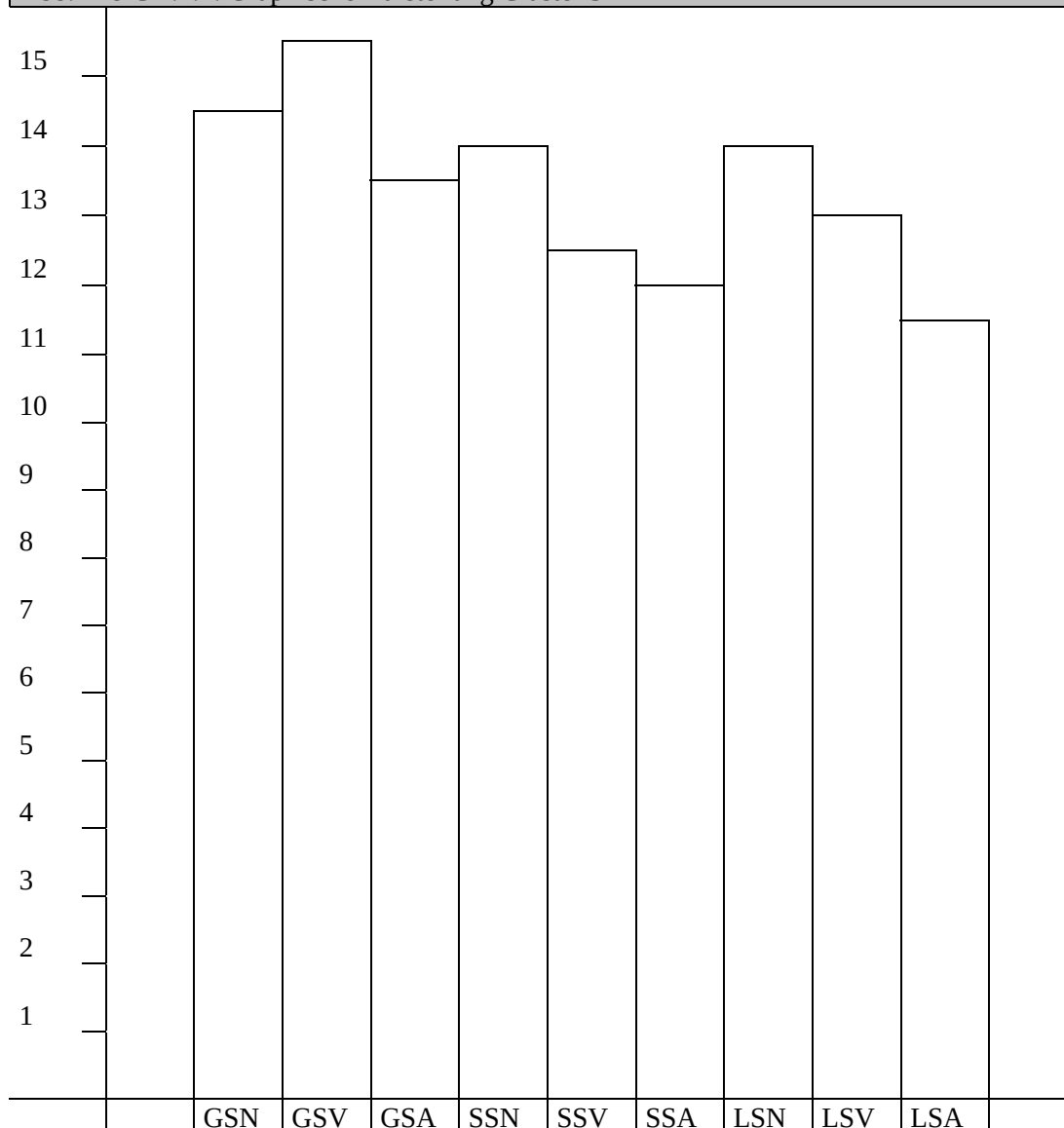
Cluster 1 beschreibt die Zusammenfassung von Ergebnissen, die von der Modalität Gebärdensprache bis zur Modalität Lautsprache immer mehr abnehmen. Bei den Mitgliedern des Clusters handelt es sich überwiegend um Testpersonen aus dritten Klassen (44%), die mit Gebärden unterrichtet werden (66%) und die vor allem in Schleswig und Köln zur Schule gehen. Testpersonen aus Hamburg und Münster passen mit ihren Testergebnissen nicht in dieses Cluster.

Abb. 17b-C 1.7.1: Graphische Darstellung Cluster 2



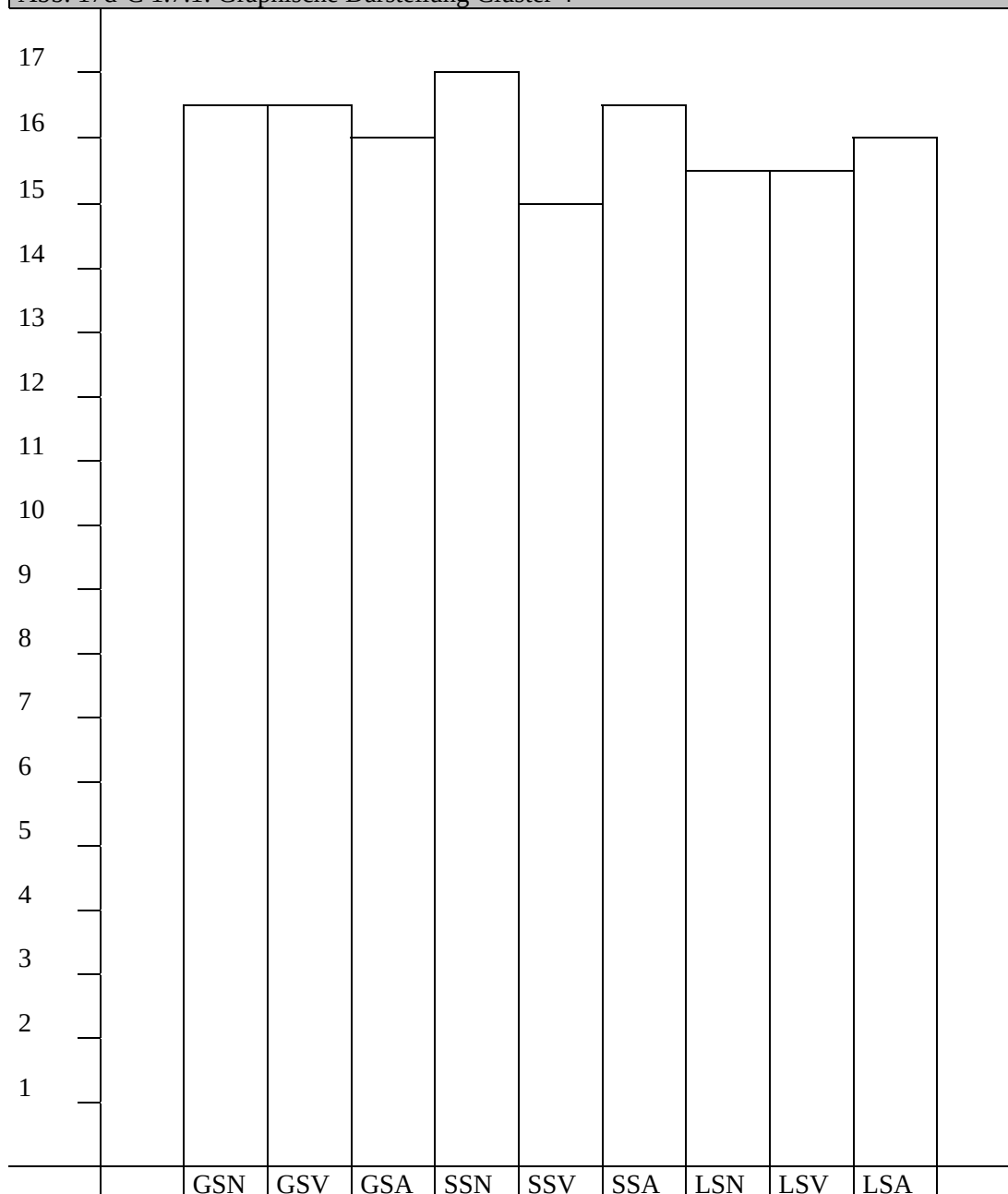
Die Ergebnisse dieses Clusters zeigen die mit Abstand beste Leistung in der Modalität Gebärdensprache und die schwächste Leistung in der Modalität Schriftsprache. Auch bei diesen Mitgliedern handelt es sich überwiegend um Testpersonen aus dritten Klassen (64%), die mit Gebärdensprache unterrichtet werden (82%) und die vor allem in Hamburg zur Schule gehen. Testpersonen aus Würzburg passen mit ihren Testergebnissen nicht in dieses Cluster.

Abb. 17c-C 1.7.1: Graphische Darstellung Cluster 3



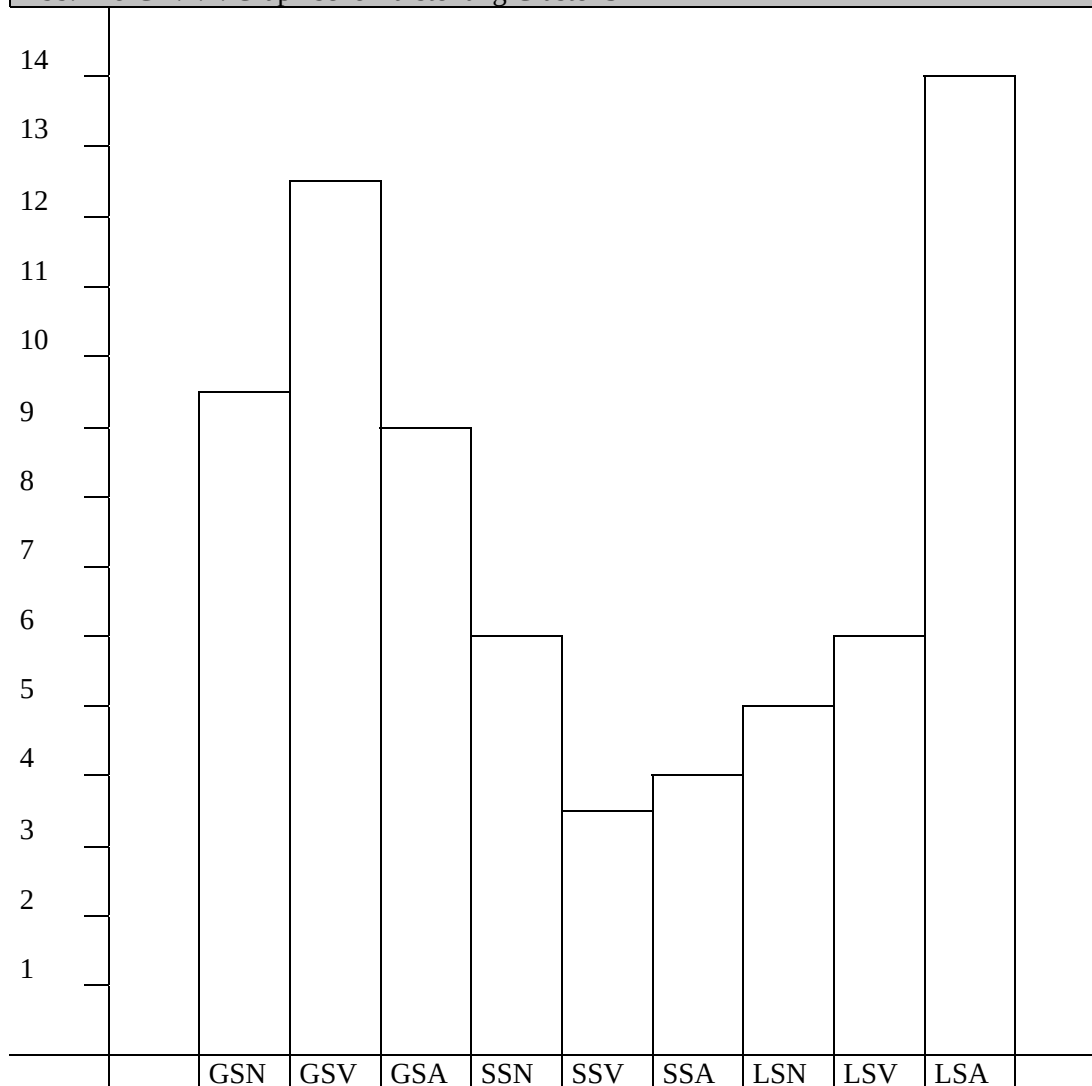
Mit etwas niedrigeren Leistungen als Cluster 4 zeigt dieses Cluster dennoch eine durchgehend gute Leistung in allen Subtestteilen. Bei den Mitgliedern dieses Clusters handelt es sich überwiegend um Testpersonen aus fünften Klassen (45%), die mit Gebärden unterrichtet werden (59%) und die vor allem in Schleswig, Bremen, Würzburg und Köln zur Schule gehen. Auch zwei der sechs bilingual erzogenen Testpersonen und drei der zehn Testpersonen mit gehörlosen Eltern sind mit ihren Testergebnissen Bestandteil dieses Clusters.

Abb. 17d-C 1.7.1: Graphische Darstellung Cluster 4



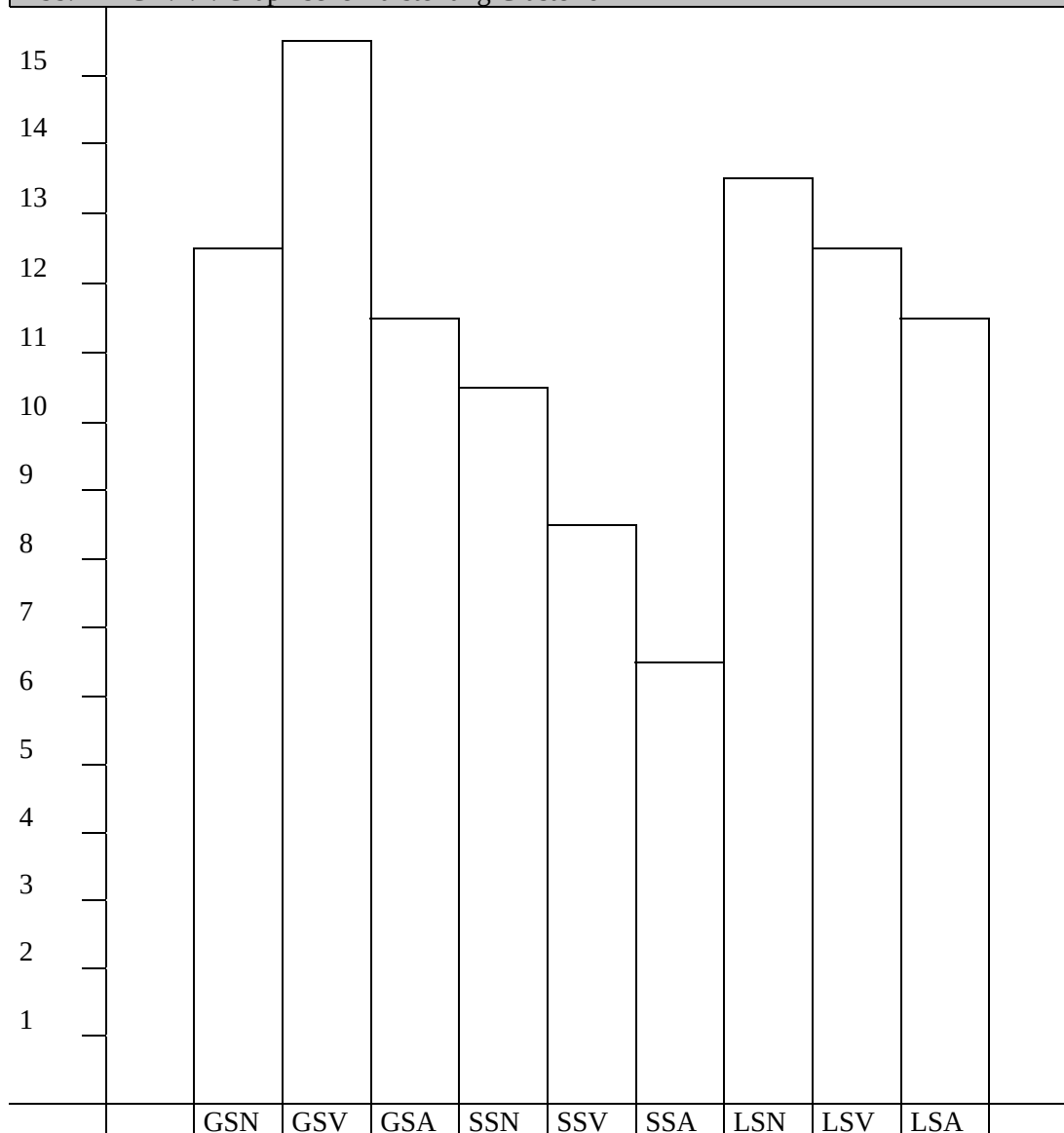
Dieses Cluster stellt eine Zusammenfassung der durchgehend besten Leistungen dar. Bei den Mitgliedern des Clusters handelt es sich fast ausschließlich um Testpersonen aus vierten und fünften Klassen (91%), die mit Gebärdensprache unterrichtet werden (48%) und die vor allem in Hamburg, Köln und Würzburg zur Schule gehen. Testpersonen aus Schleswig passen mit ihren Testergebnissen nicht in dieses Cluster. Vier der sechs Testpersonen, die bilingual erzogen werden, gehören mit ihren Ergebnissen in dieses Cluster.

Abb. 17e-C 1.7.1: Graphische Darstellung Cluster 5



Dieses Cluster zeigt die eindeutig schwächsten Leistungen in der Modalität Schriftsprache und auffallend gute Leistungen im Subtestteil Lautsprache Adjektive. Bei den Mitgliedern des Clusters handelt es sich fast ausschließlich um Testpersonen aus dritten Klassen (78%), die alle mit Gebärdensprache unterrichtet werden und die vor allem in Bremen und Köln zur Schule gehen. Testpersonen aus Schleswig, Münster, Zürich und Würzburg passen mit ihren Testergebnissen nicht in dieses Cluster. Es handelt sich bei den betreffenden Testpersonen überwiegend um Jungen (67%) und um Testpersonen, deren Ergebnisse im Raven zu 44% in der untersten Intelligenzgruppe anzusiedeln sind.

Abb. 17f-C 1.7.1: Graphische Darstellung Cluster 6



Cluster 6 zeigt eine Zusammenfassung von guten Leistungen Gebärden- und Lautsprache. Es handelt sich fast ausschließlich um Testpersonen aus dritten Klassen (78%), die alle mit Gebärdensprache unterrichtet werden und die vor allem in Bremen und Köln zur Schule gehen. Ergebnisse von Testpersonen aus Schleswig, Münster, Zürich und Würzburg passen nicht in dieses Cluster.

1.7.2 Überprüfung II

Über eine Clusteranalyse hinaus wurde untersucht, welche Profilverläufe innerhalb der drei Modalitäten häufig vorkommen. Dafür wurde ausgezählt, wie oft die Testpersonen ihre beste der drei Teilleistungen in Gebärdensprache bzw. in Schrift- oder Lautsprache erzielten. Von 101 Testpersonen haben 68 die beste Leistung in Gebärdensprache erzielt. Davon haben 43 Testpersonen ihre zweitbeste Leistung in Laut-, 25 ihre zweitbeste Leistung in Schriftsprache erzielt.

che erreicht. Von den zehn Testpersonen, die in Schriftsprache die stärkste von drei Leistungen aufweisen, haben sieben die zweitbeste Leistung in Gebärdensprache, drei in Lautsprache erzielt. Insgesamt haben sieben Testpersonen von den drei Leistungen ihre beste in Lautsprache erreicht, davon sind fünf in Gebärdensprache, zwei in Schriftsprache am zweitbesten. Zudem haben 16 Testpersonen in mindestens zwei Modalitäten die gleiche Punktzahl erreicht. Die so entstehenden sieben Profile werden nicht im einzelnen beschrieben. In der folgenden Tabelle sind die 13 Profile durch in Höhe variierende Kreuze (x) gekennzeichnet und ihrer Häufigkeit im Vorkommen nach geordnet:

Tab. 85-C 1.7.2: Teilleistungsprofile innerhalb der Modalitäten			
Gebärdensprache	Schriftsprache	Lautsprache	Anzahl
x			
		x	43
	x		
x			
	x		25
		x	
x			
	x	x	7
	x		
x			7
		x	
		x	
x			5
	x		
	x		
		x	3
x			
		x	
	x		2
x			
x		x	
	x		2
x	x	x	
			2
	x		
x		x	2
x	x		
		x	1
	x	x	
x			1
		x	
x	x		1

Innerhalb von drei Profilverläufen wird in 74,3% aller Fälle in der Gebärdensprache die eindeutig beste Leistung erzielt. Das häufigste (57,3%) dieser Profile zeigt die zweitbeste Leistung in der Lautsprache, das zweithäufigste (33,3%) in der Schriftsprache. Alle übrigen Profile kommen in der Stichprobe selten, d. h. maximal siebenmal (6,9%) vor. Dabei werden in 11,9% aller Fälle (drei Profile) in der Schriftsprache, in 7,9% aller Fälle (drei Profile) in der Lautsprache die eindeutig besten Leistungen erzielt.

Im Folgenden werden diejenigen Profilverläufe innerhalb der Modalitäten im Koordinatensystem differenziert dargestellt, die in mindestens 10% aller Fälle vorkommen:

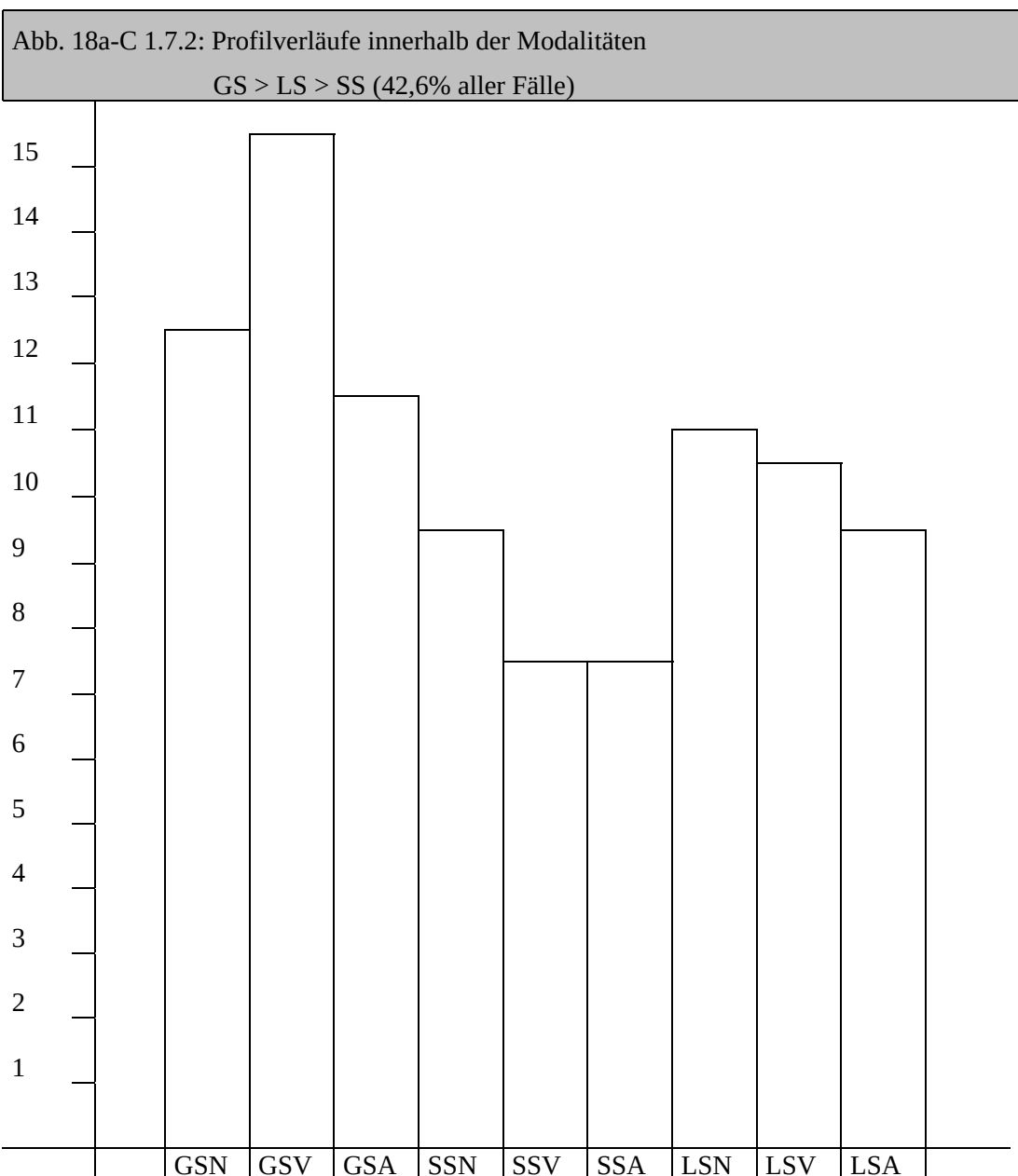
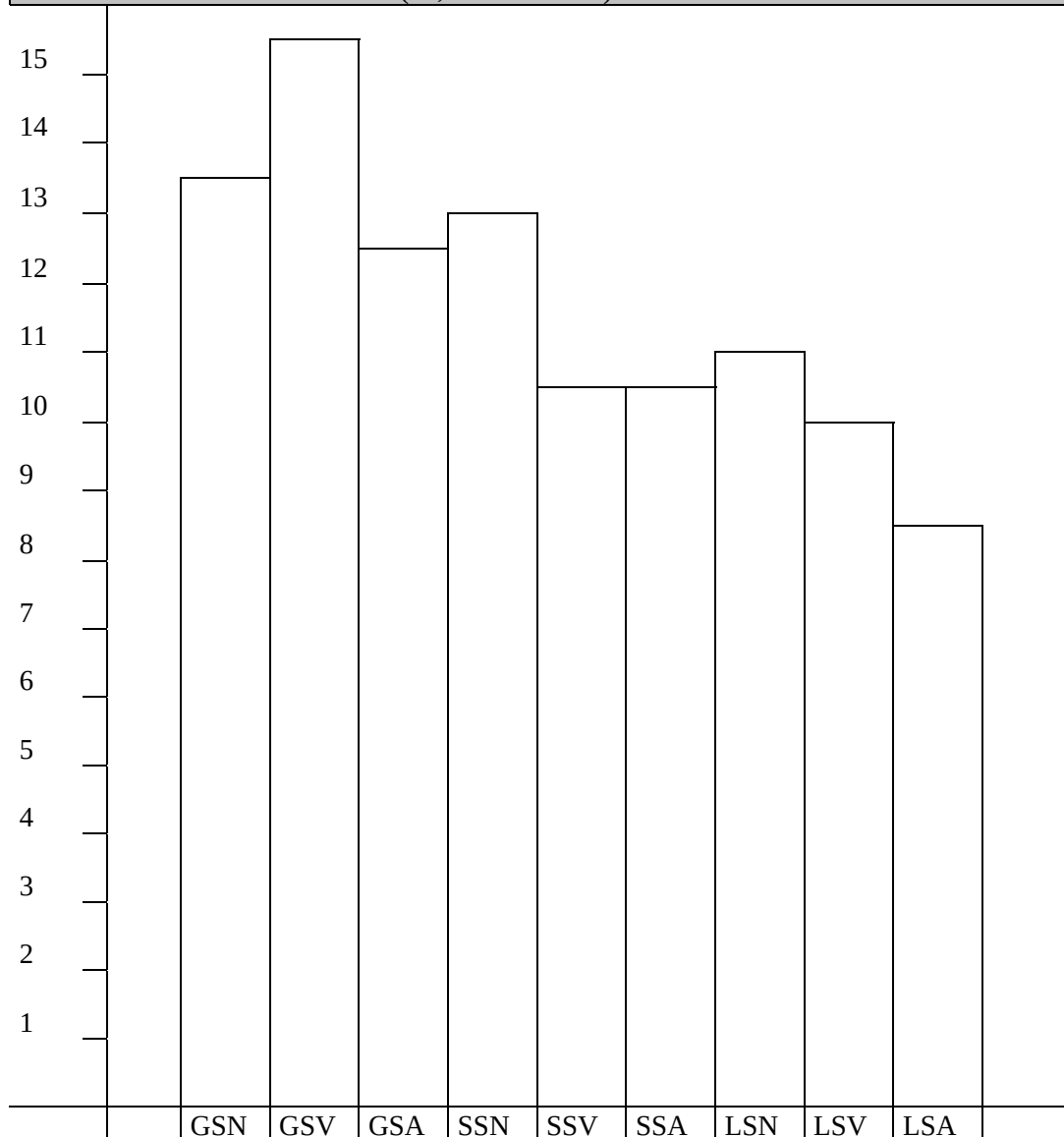


Abb. 18b-C 1.7.2: Profilverläufe innerhalb der Modalitäten

GS > SS > LS (24,8% aller Fälle)



Daraufhin wurde untersucht, welche Profilverläufe innerhalb der Modalität Gebärdensprache (GSN, GSV, GSA) häufig vorkommen. Von den 110 Testpersonen, die die Modalität Gebärdensprache bearbeitet haben, erzielten zwölf die beste von drei Leistungen in der Wortart Nomen. Davon haben zehn Testpersonen ihre zweitbeste Leistung in der Wortart Verben, zwei ihre zweitbeste Leistung in der Wortart Adjektive erreicht. Von den 70 Testpersonen, die in der Wortart Verben die stärkste von drei Leistungen aufweisen, haben 38 in der Wortart Nomen, 25 in der Wortart Adjektive die zweitbeste Leistung erzielt. Bei den restlichen sieben Testpersonen waren die Ergebnisse bei den Adjektiven und Verben gleich gut. Insgesamt haben sechs Testpersonen von den drei Leistungen ihre beste in der Wortart Adjektive erreicht, von denen alle in der Wortart Verben am zweitbesten abschnitten. Es ergaben sich bei 29 Testpersonen Testprofile, die in mindestens zwei Wortarten gleiche Werte aufweisen.

In der folgenden Tabelle sind die zwölf Teilleistungsprofile durch in Höhe variierende Kreuze gekennzeichnet. Die Profile sind ihrer Häufigkeit im Vorkommen nach geordnet:

Tab. 86-C 1.7.2: Teilleistungsprofile innerhalb der Modalität Gebärdensprache			
Nomen	Verben	Adjektive	Anzahl
	X		
X			38
		X	
	X		
		X	25
X			
X			
	X		10
		X	
X	X		
		X	10
	X		
X		X	7
		X	
	X		6
X			
	X	X	
X			5
X			
	X	X	3
X			
		X	2
	X		
X		X	
	X		2
X	X	X	
			1
		X	
X	X		1

In 13,6% aller Fälle wird die eindeutig beste Leistung in der Wortart Nomen erzielt. Dabei entstehen drei Profilverläufe. Davon besteht in 66,6% der Fälle eine erwartungsgemäße Leistungsabstufung von den Nomen bis zu den Adjektiven. In 63,6% aller Fälle und in drei Profilen wird in der Wortart Verben die eindeutig beste Leistung erbracht. Das häufigste (54,3%) unter diesen Profilen zeigt die zweitbeste Leistung bei den Nomen. In 6,4% aller

Fälle und in zwei Profilen wird in der Wortart Adjektive die eindeutig beste Leistung erbracht. Dabei zeigte das häufigste (85,7%) dieser Profile die zweitbeste Leistung in der Wortart Verben.

Im Folgenden werden diejenigen Profilverläufe innerhalb der Modalität Gebärdensprache im Koordinatensystem differenziert dargestellt, die in mindestens 10% aller Fälle vorkommen:

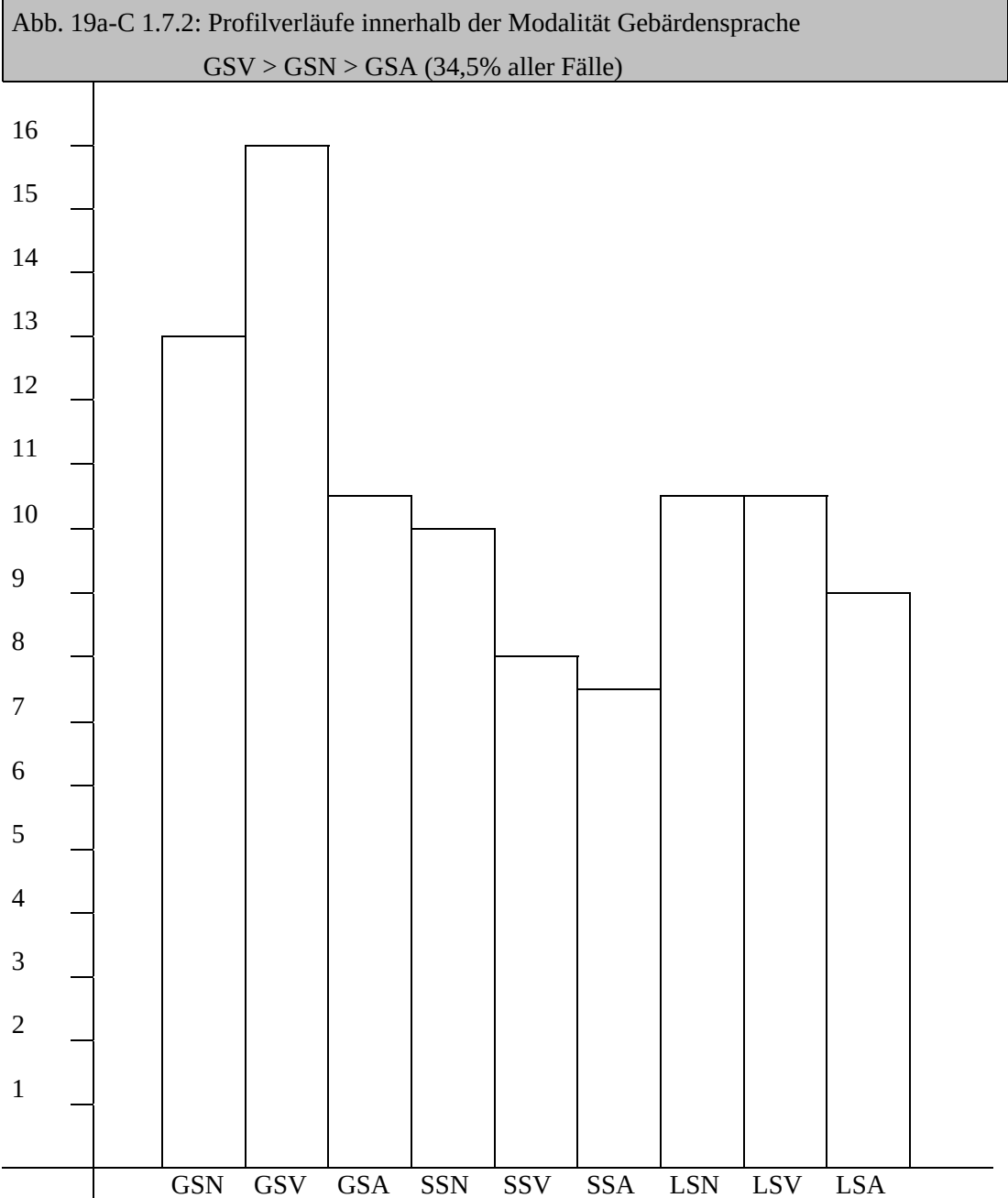
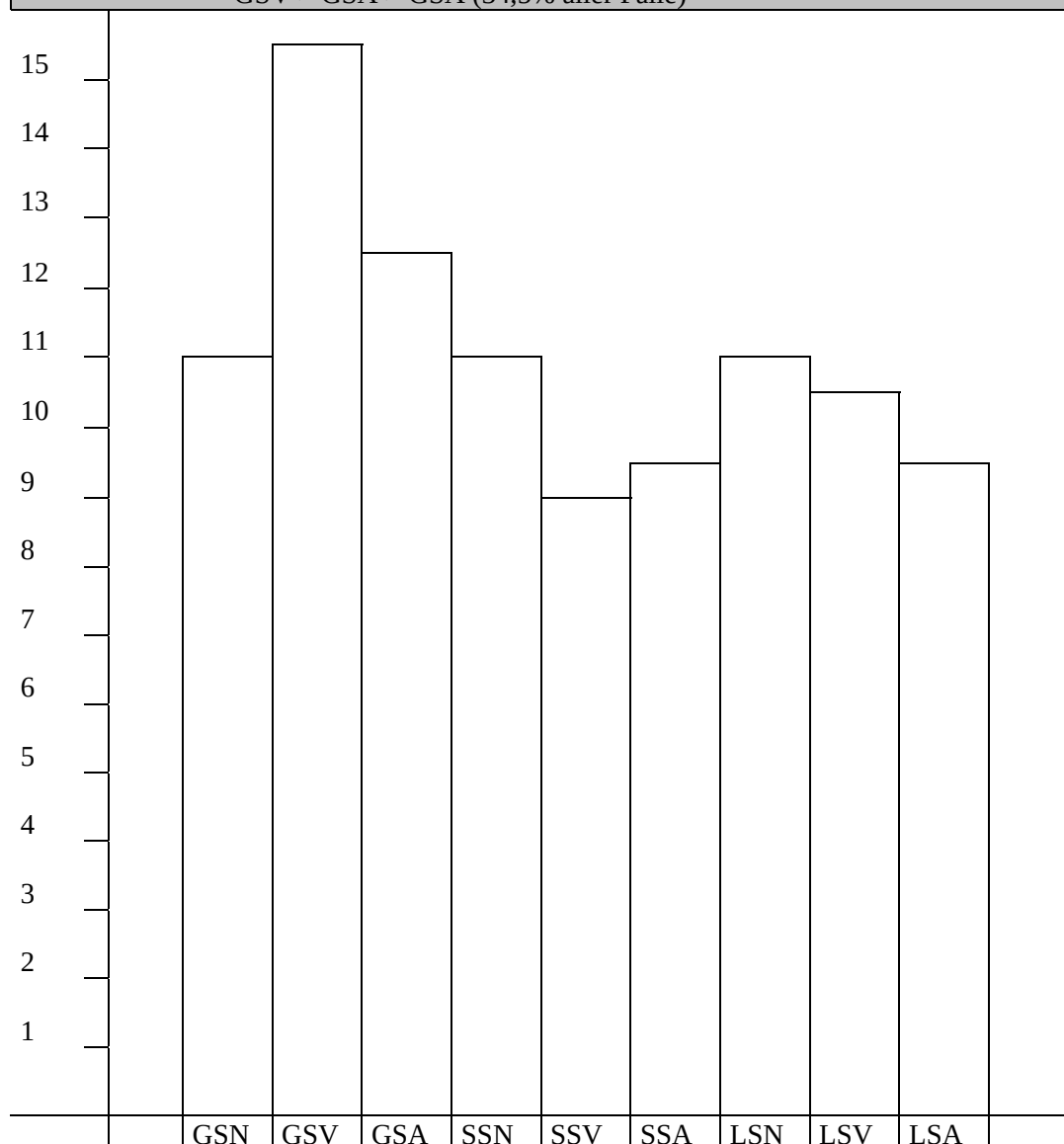


Abb. 19b-C 1.7.2: Profilverläufe innerhalb der Modalität Gebärdensprache

GSV > GSA > GSA (34,5% aller Fälle)



Entsprechend wurden die Profilverläufe innerhalb der Modalität Schriftsprache (SSN, SSV, SSA) auf ihre Häufigkeit hin untersucht. Von 108 Testpersonen, die die Modalität Schriftsprache bearbeitet haben, erzielten 61 die beste von drei Leistungen in der Wortart Nomen. Davon haben 19 Testpersonen ihre zweitbeste Leistung in der Wortart Verben, 28 ihre zweitbeste Leistung in der Wortart Adjektive erreicht. Von den sieben Testpersonen, die in der Wortart Verben die stärkste von drei Leistungen aufweisen, haben sechs in der Wortart Nomen, eine in der Wortart Adjektive die zweitbeste Leistung erzielt. Insgesamt haben acht Testpersonen von den drei Leistungen ihre beste in der Wortart Adjektive erreicht, von denen alle in der Wortart Nomen am zweitbesten abschnitten. Es ergaben sich bei 46 Testpersonen Testprofile, die in mindestens zwei Wortarten gleiche Werte aufweisen. In der folgenden Tabelle sind die zwölf Teilleistungsprofile durch in Höhe variierende Kreuze gekennzeichnet. Die Profile sind ihrer Häufigkeit im Vorkommen nach geordnet:

Tab. 87-C 1.7.2: Teilleistungsprofile innerhalb der Modalität Schriftsprache			
Nomen	Verben	Adjektive	Anzahl
X			
		X	28
	X		
X			
	X		19
		X	
X			
	X	X	14
X		X	
	X		12
		X	
X			8
	X		
	X		
X			6
		X	
	X		
X		X	5
		X	
X	X		5
X	X		
		X	4
X	X	X	
			4
	X	X	
X			2
	X		
		X	1
X			

In 56,5% aller Fälle wird die eindeutig beste Leistung in der Wortart Nomen erzielt. Dabei entstehen drei Profilverläufe. Davon besteht in 31,1% der Fälle eine erwartungsgemäße Leistungsabstufung von den Nomen bis zu den Adjektiven, in 45,9% der Fälle wird die zweitbeste Leistung in der Wortart Adjektive erzielt; ähnlich fällt das Ergebnis für die Fälle aus, in denen Verben und Adjektive gleich stark ausfallen. Diese Werte ähneln einander zu sehr, als dass weiterführende Schlüsse daraus abgeleitet werden können. In 11,1% aller Fälle und in drei Profilen wird in der Wortart Verben die eindeutig beste Leistung erbracht. Das häufigste (50,0%) unter diesen Profilen zeigt die zweitbeste Leistung in der Gebärdensprache. In

12,0% aller Fälle (zwei Profile) wird in der Wortart Adjektive die eindeutig beste Leistung erbracht. Dabei zeigte das häufigste (61,5%) dieser Profile die zweitbeste Leistung in der Wortart Verben.

Im Folgenden werden diejenigen Profilverläufe innerhalb der Modalität Schriftsprache im Koordinatensystem differenziert dargestellt, die in mindestens 10% aller Fälle vorkommen:

Abb. 20a-C 1.7.2: Profilverläufe innerhalb der Modalität Schriftsprache

SSN > SSA > SSV (25,9% aller Fälle)

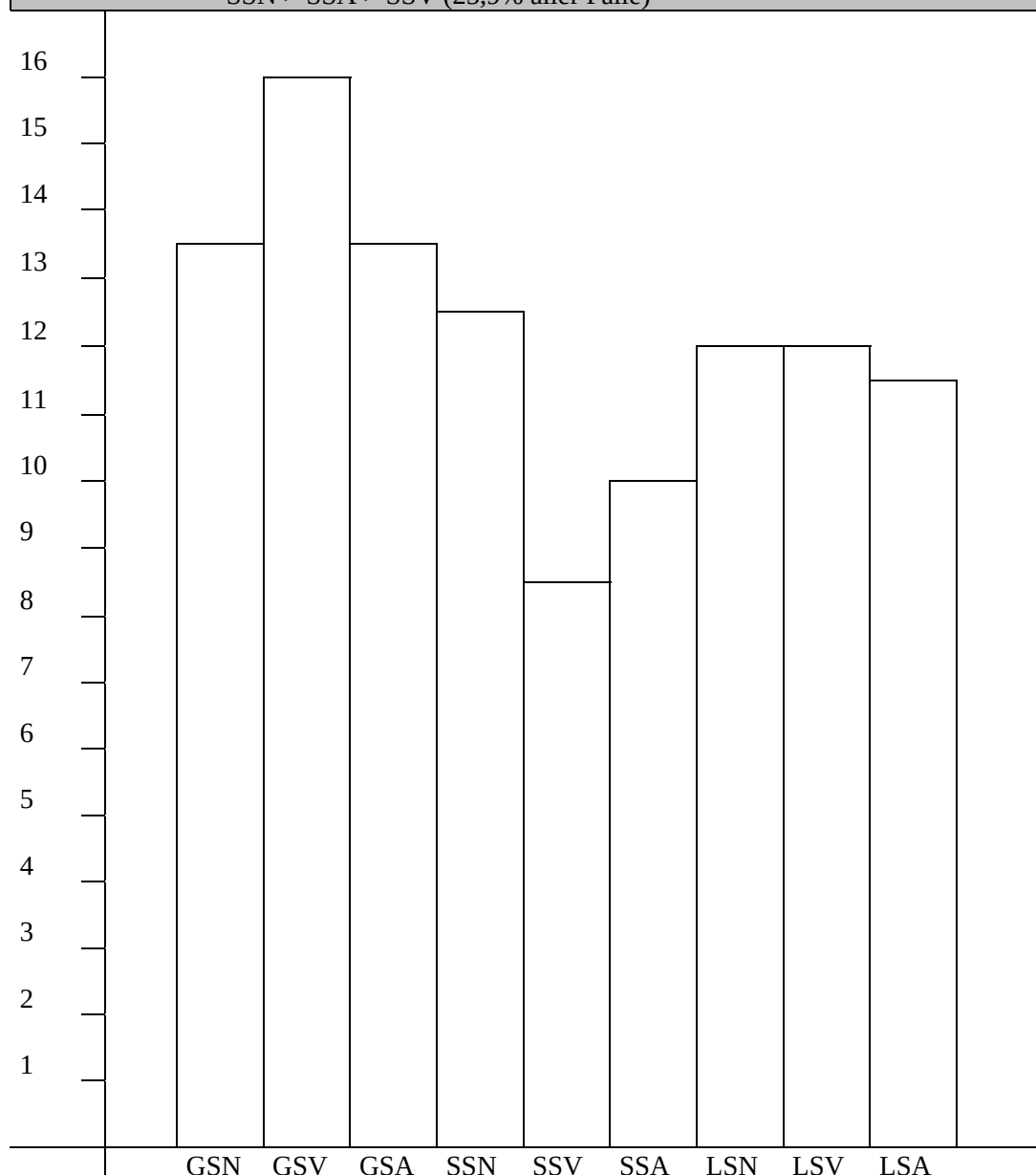
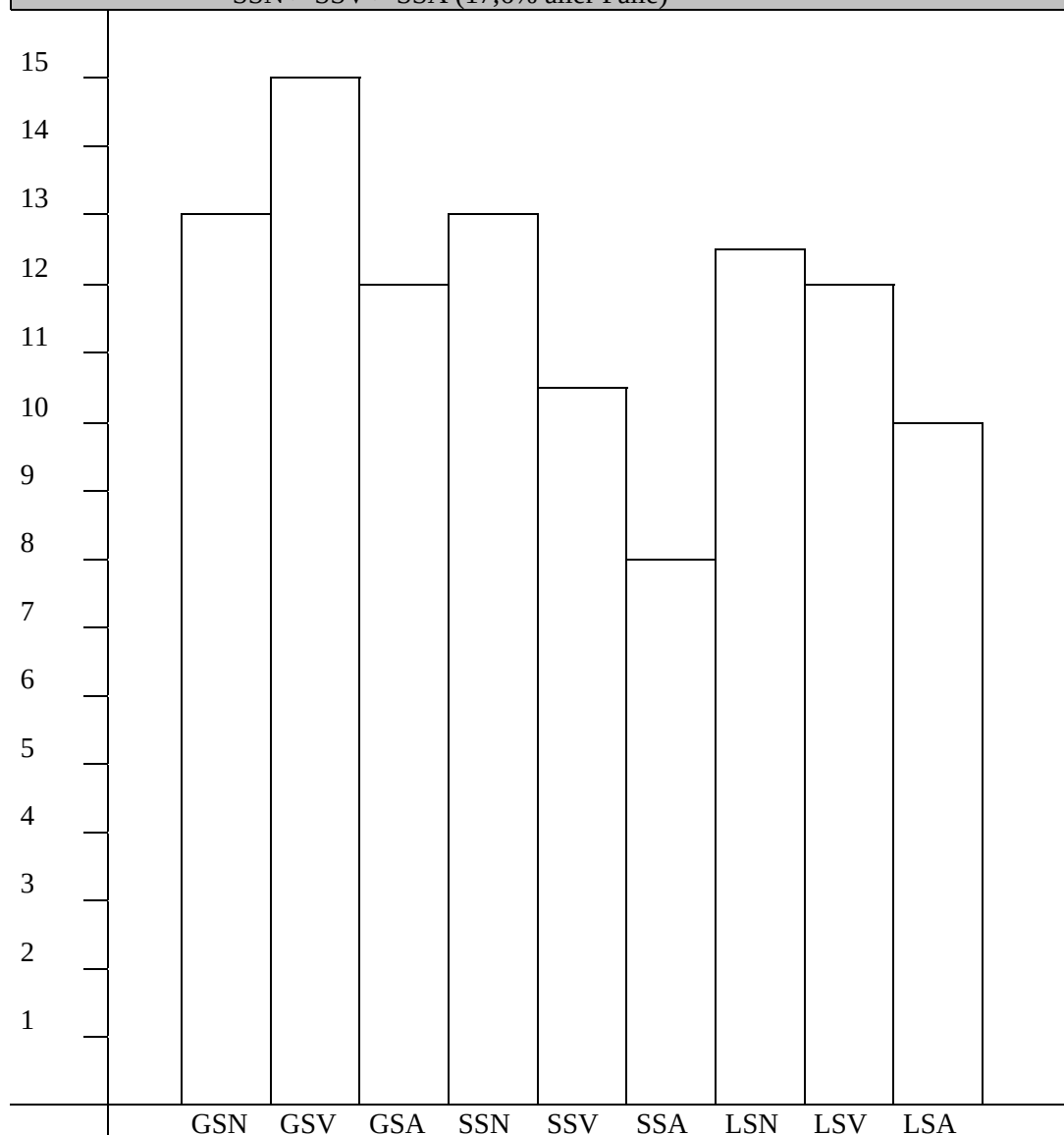


Abb. 20b-C 1.7.2: Profilverläufe innerhalb der Modalität Schriftsprache

SSN > SSV > SSA (17,6% aller Fälle)



Ebenso wurden die Profilverläufe innerhalb der Modalität Lautsprache (LSN, LSV, LSA) auf ihre Häufigkeit hin untersucht. Von den 106 Testpersonen, die die Modalität Schriftsprache bearbeitet haben, erzielten 42 die beste von drei Leistungen in der Wortart Nomen. Davon haben 27 Testpersonen ihre zweitbeste Leistung in der Wortart Verben, 8 ihre zweitbeste Leistung in der Wortart Adjektive erreicht. Von den 24 Testpersonen, die in der Wortart Verben die stärkste von drei Leistungen aufweisen, haben 19 in der Wortart Nomen, fünf in der Wortart Adjektive die zweitbeste Leistung erzielt. Insgesamt haben 13 Testpersonen von den drei Leistungen ihre beste in der Wortart Adjektive erreicht, von denen fünf in der Wortart Nomen, acht in der Wortart Verben am zweitbesten abschnitten. Es ergaben sich bei 34 Testpersonen Testprofile, die in mindestens zwei Wortarten gleiche Werte aufweisen. In der folgenden Tabelle sind die 13 Teilleistungsprofile durch in Höhe variierende Kreuze gekennzeichnet. Die Profile sind ihrer Häufigkeit im Vorkommen nach geordnet:

Tab. 88-C 1.7.2: Teilleistungsprofile innerhalb der Modalität Lautsprache			
Nomen	Verben	Adjektive	Anzahl
x			
	x		27
		x	
	x		
x			19
		x	
x			
		x	8
	x		
		x	
	x		8
x			
x		x	
	x		8
x			
	x	x	7
x	x		
		x	7
	x		
		x	5
x			
		x	
x			5
	x		
		x	
x	x		4
x	x	x	
			3
	x	x	
x			3
	x		
x		x	2

In 39,6% aller Fälle wird die eindeutig beste Leistung in der Wortart Nomen erzielt. Dabei entstehen drei Profilverläufe. Davon besteht in 64,3% der Fälle ein erwartungsgemäßer Leistungsabfall von den Nomen bis zu den Adjektiven, in 19,0% der Fälle wird die zweitbeste Leistung in der Wortart Adjektive erzielt. In 24,5% aller Fälle und in drei Profilen wird in der Wortart Verben die eindeutig beste Leistung erbracht. Das häufigste (73,1%) unter die-

sen Profilen zeigt die zweitbeste Leistung in der Gebärdensprache. In 16,0% aller Fälle (drei Profile) wird in der Wortart Adjektive die eindeutig beste Leistung erbracht. Dabei kommen zwei Profile, das mit der zweitbesten Leistung in der Wortart Nomen und das mit der zweitbesten Leistung in der Wortart Verben, gleich häufig (47,1%) vor.

Im Folgenden werden diejenigen Profilverläufe innerhalb der Modalität Schriftsprache im Koordinatensystem differenziert dargestellt, die in mindestens 10% aller Fälle vorkommen:

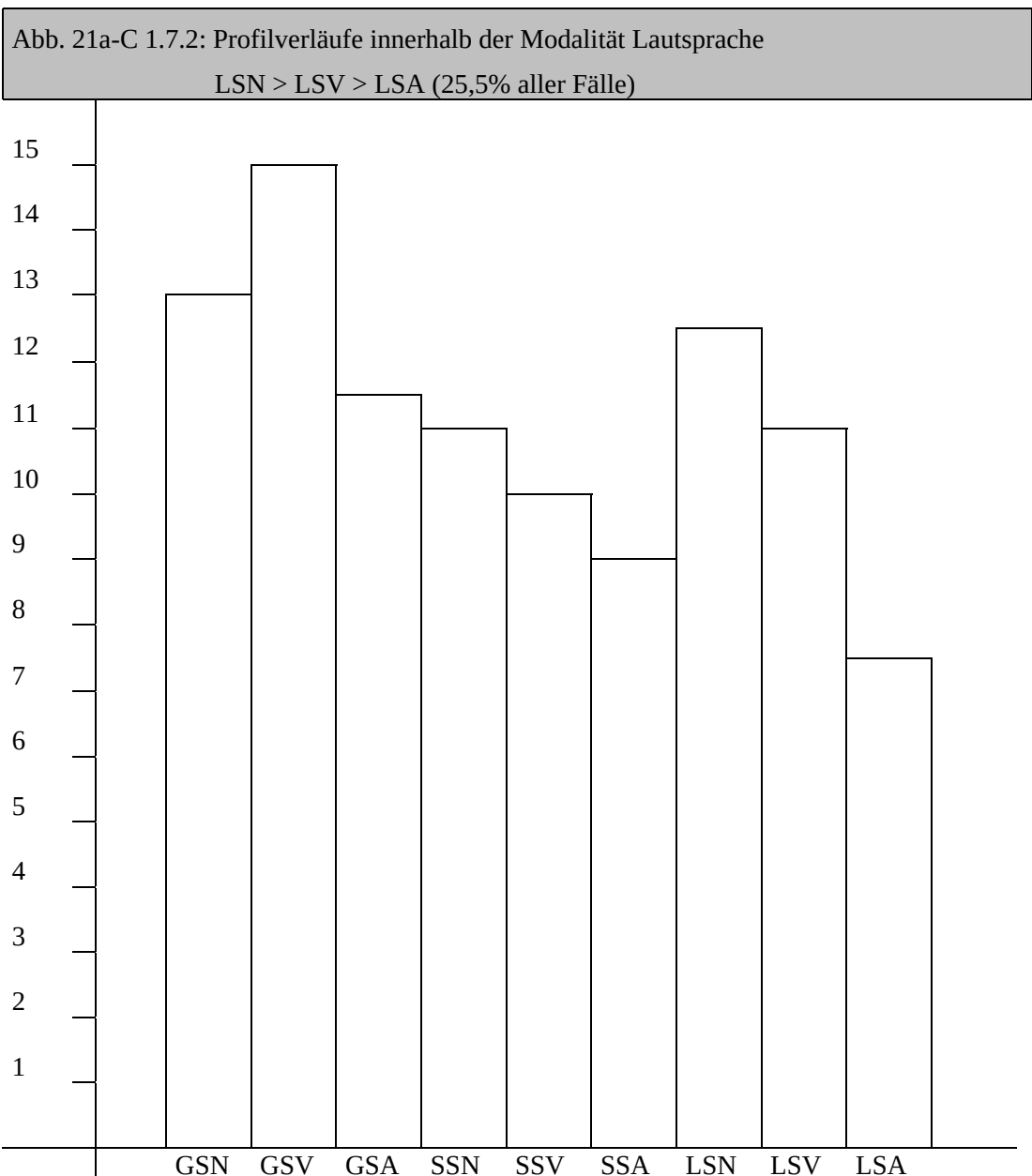
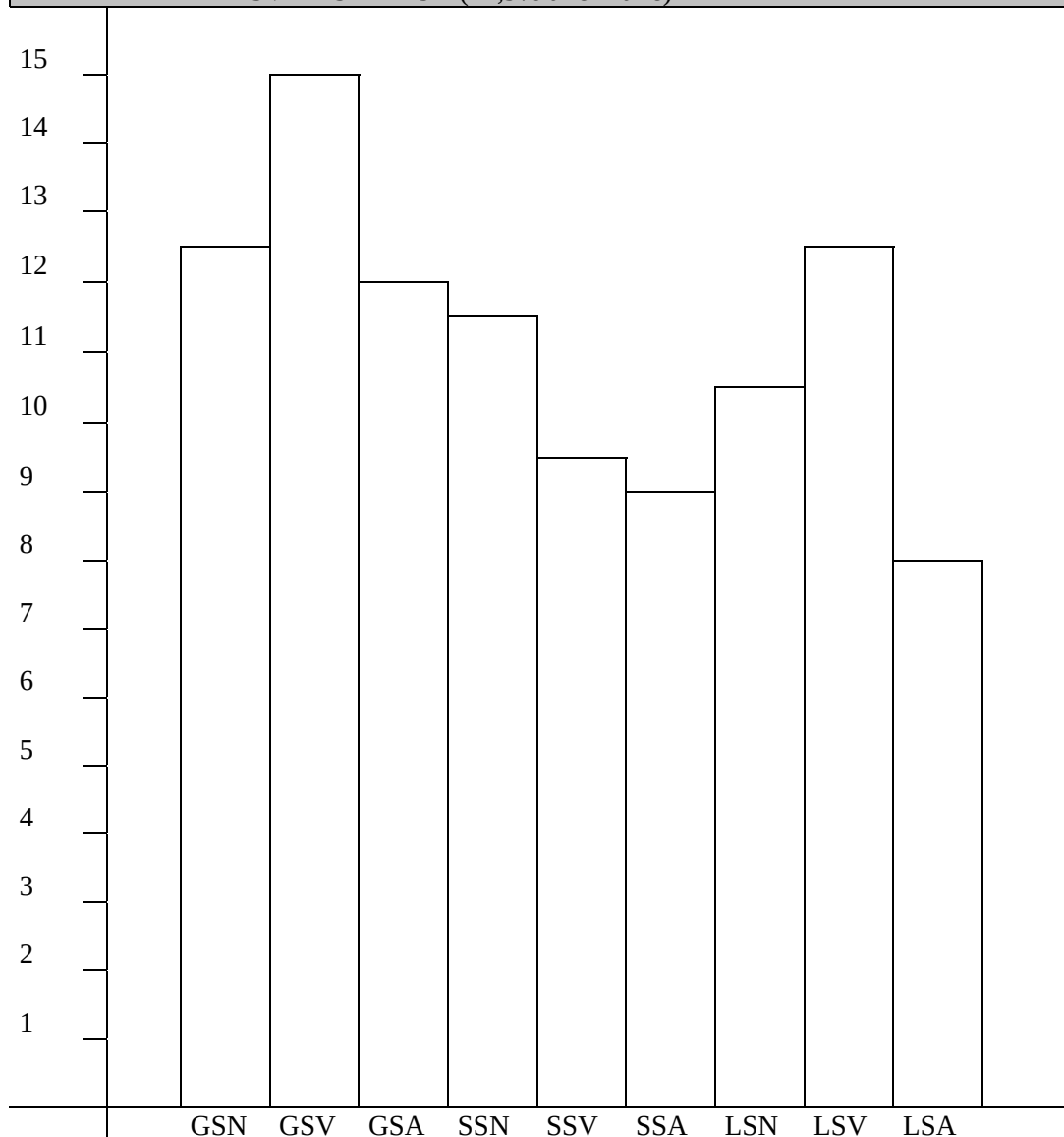


Abb. 21b-C 1.7.2: Profilverläufe innerhalb der Modalität Lautsprache

LSV > LSN > LSA (17,9% aller Fälle)



1.7.3 Überprüfung III

In einem weiteren Schritt wurde untersucht, welche Profilverläufe innerhalb der drei Wortarten (N, V, A) häufig vorkommen. Von 101 Testpersonen haben 51 die beste von drei Leistungen in der Wortart Nomen erzielt. Davon haben 35 Testpersonen ihre zweitbeste Leistung in der Wortart Verben, zehn ihre zweitbeste Leistung in der Wortart Adjektive erreicht. Von den 29 Testpersonen, die in der Wortart Verben die stärkste von drei Leistungen aufweisen, haben 21 die zweitbeste Leistung in der Wortart Nomen, 8 in der Wortart Adjektive erzielt. Insgesamt haben acht Testpersonen von den drei Leistungen ihre beste in der Wortart Adjektive erreicht, davon sind vier in der Wortart Nomen, vier in der Wortart Verben am zweitbesten. Auch in dieser Untersuchung ergaben sich bei 16 Testpersonen Testprofile, die in mindestens zwei Wortarten gleiche Werte aufweisen. In der folgenden Tabelle sind die

zwölf Teilleistungsprofile durch in Höhe variierende Kreuze gekennzeichnet. Die Profile sind ihrer Häufigkeit im Vorkommen nach geordnet:

Tab. 89-C 1.7.3: Teilleistungsprofile innerhalb der Wortarten			
Nomen	Verben	Adjektive	Anzahl
X			
	X		35
		X	
	X		
X			21
		X	
X			
		X	10
	X		
	X		
		X	8
X			
X			
	X	X	6
		X	
X			4
	X		
		X	
	X		4
X			
	X	X	
X			4
X	X		
		X	3
X		X	
	X		2
	X		
X		X	2
		X	
X	X		2

In 50,5% aller Fälle wird die eindeutig beste Leistung in der Wortart Nomen erzielt. Dabei ergeben sich drei Profilverläufe. Davon entsteht in 68,6% der Fälle eine erwartungsgemäße Leistungsabstufung von den Nomen bis zu den Adjektiven. In 30,7% aller Fälle und in drei Profilen wird in der Wortart Verben die eindeutig beste Leistung erbracht. Das häufigste (67,7%) unter diesen Profilen zeigt die zweitbeste Leistung in der Gebärdensprache. In 9,9%

aller Fälle (drei Profile) wird in der Wortart Adjektive die eindeutig beste Leistung erbracht. Derartige Profile kommen jedoch nur selten (maximal 4,0% aller Fälle) vor.

Im Folgenden werden diejenigen Profilverläufe innerhalb der Wortarten im Koordinatensystem differenziert dargestellt, die in mindestens 10% aller Fälle vorkommen:

Abb. 22a-C 1.7.3: Profilverläufe innerhalb der Wortarten

N > V > A (34,7% aller Fälle)

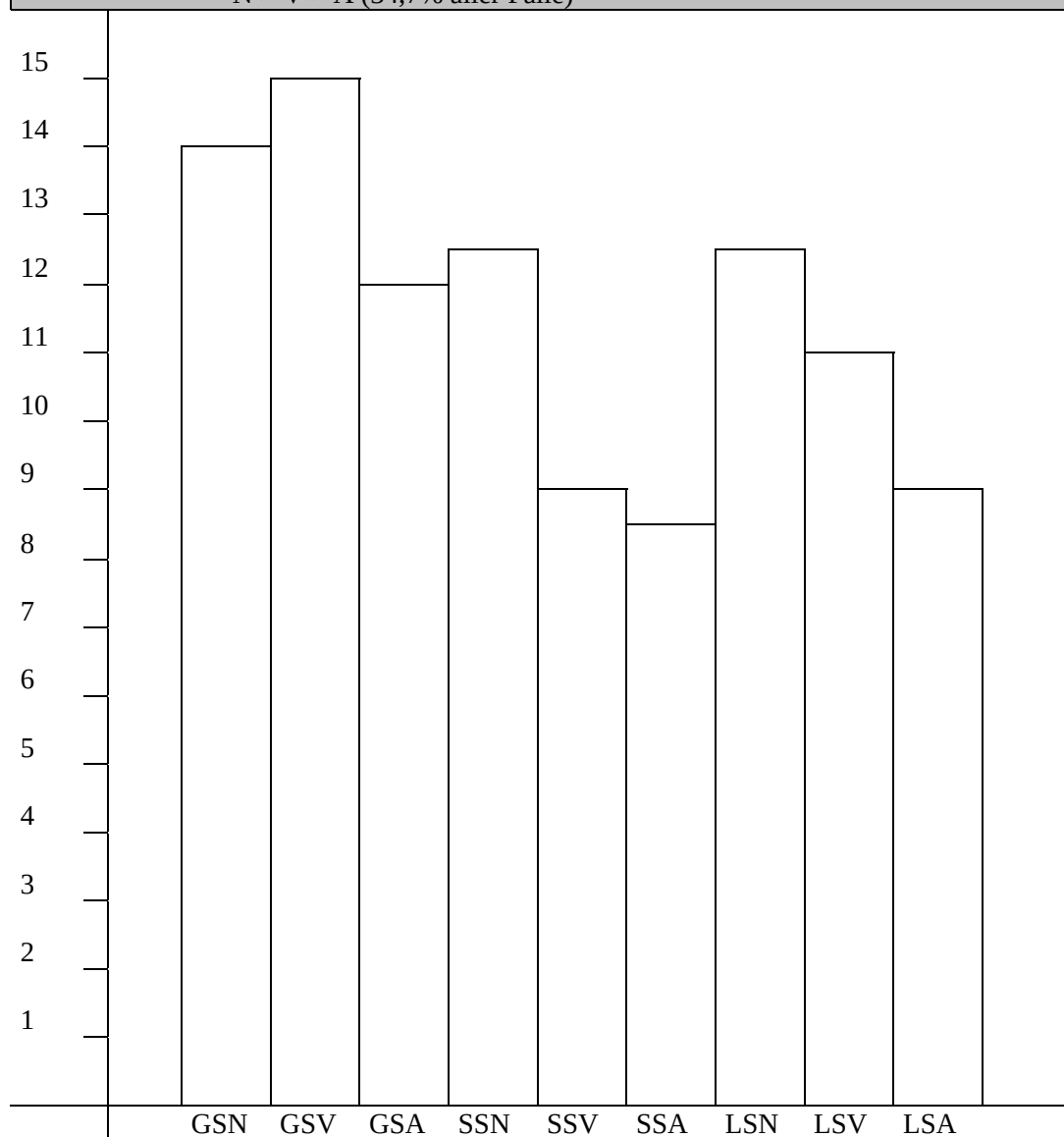
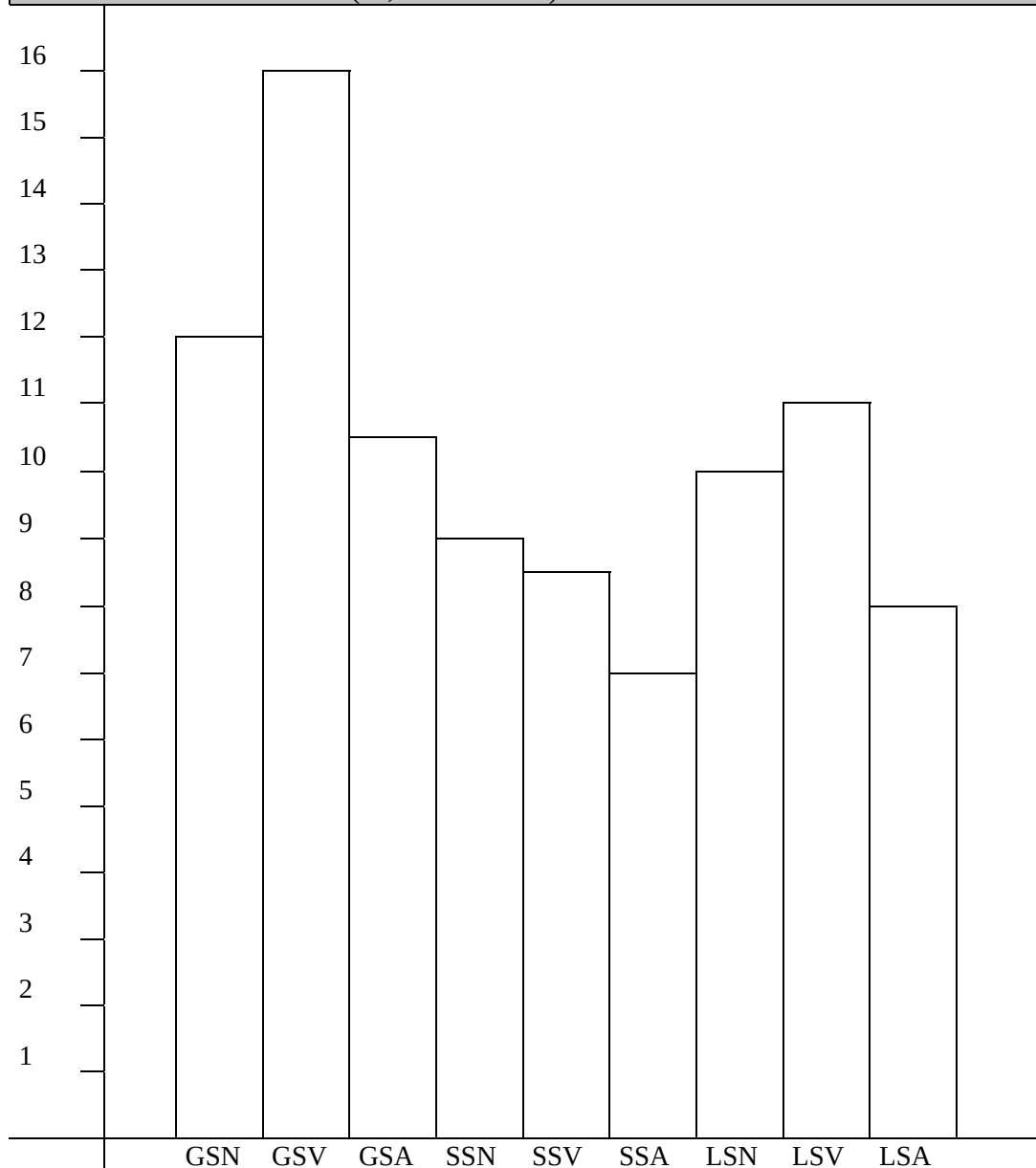


Abb. 22b-C 1.7.3: Profilverläufe innerhalb der Wortarten

V > N > A (20,8% aller Fälle)



1.7.4 Überprüfung IV

Über die Untersuchung der Profilverläufe innerhalb verschiedener Subtests oder Subtestteile hinaus, wurde überprüft, ob sich die Profilverläufe der Testpersonen unterschiedlicher Konzepte deutlich voneinander unterscheiden. Die dazugehörigen Werte sind im Anhang (Anh. 1: C 1.7.4) dargestellt (Abb. 23-C 1.7.4).

Bis auf eine Überschneidung der Konzepte 'oral' und 'mit Gebärden' im Subtestteil Gebärdensprache Verben zeigt sich eine deutliche – bereits in C 1.5 beschriebene – Unterscheidung der Leistungsergebnisse der Testpersonen unterschiedlicher Konzepte.

1.7.5 Überprüfung V

Schließlich wurden die Profilverläufe der Testpersonen unterschiedlicher Klassen und ihre Unterschiede untersucht. Die dazugehörigen Werte sind im Anhang (Anh. 1: C 1.7.5) dargestellt (Abb. 24-C 1.7.5).

Die Darstellung zeigt eine deutliche Differenz der Profilverläufe. Die Leistungen der Klasse 3 liegen durchgehend und fast gleichabständig in allen Subtestteilen unter denen der Klasse 4 (+5).

Zusammenfassend stellt sich heraus, dass sich im Laufe der Untersuchung verschiedene Profilverläufe ergeben haben, die sowohl aufgrund ähnlicher Geamttestverläufe als auch aufgrund ähnlicher Ergebnisse in einzelnen Subtests oder Subtestteilen zustande kamen. Die Ergebnisse weiterer Testdurchgänge können mit diesen Profilverläufen verglichen und somit zu der Grundgesamtheit oder speziellen Untergruppen der Grundgesamtheit in Beziehung gesetzt werden.

1.8 Häufig vorkommende typische Profilverläufe bzgl. der Distraktorenwahl

Es wird angenommen, dass es typische Profilverläufe bezüglich der Distraktorenwahl gibt, die besonders häufig vorkommen. Solche Häufungen können Hinweise geben auf Wahlstrategien der Testpersonen sowie die Qualität der Aufgabenkonzeption differenziert erhellen.

Selbstverständlich soll mit diesen Aussagen (wie auch in C 1.6.2 und C 1.7) über Leistungen in bestimmten Schulen keine Bewertung vorgenommen werden. Abweichungen im Leistungsniveau sind auf einen Selektionseffekt der Stichprobe und damit – zu großen Anteilen über das zu messende Merkmal hinausgehend – auf Zufälligkeiten zurückzuführen.

Vorwegnehmend ist zu sagen, dass sich im Rahmen der Untersuchung aufgrund von Clusteranalysen für jeden Subtestteil sowie für den Gesamttest jeweils sieben typische Profilverläufe ergaben. Die Zusammenstellung dieser Cluster wird im Folgenden im einzelnen aufgeführt und jeweils abschließend analysiert. Auf diese Weise wird die sich aus den Daten ergebende inhaltliche Relevanz deutlich. Zudem können so Ergebnisse zukünftiger Erhebungen eingeordnet und mit der Analysestichprobe in Beziehung gesetzt werden.

1.8.1 Überprüfung I

Zunächst wurde die durchschnittliche Wahlhäufigkeit (in %) der unterschiedlichen Distraktorenarten aller Testpersonen, die den jeweiligen Subtest bearbeitet haben, untersucht. Dabei

steht G für Ähnlichkeit zum gefragten Begriff in der Gebärde, M für Ähnlichkeit in Mundbild und I für Ähnlichkeit im Inhalt. Besonders starke inhaltliche Ähnlichkeit und damit besonders hohe Plausibilität des Ablenkens wird durch den Distraktor HRP angezeigt. Die Abkürzung MOD bedeutet die dem jeweiligen Subtest entsprechende Ähnlichkeit, das entspricht in den Subtests Gebärden- bzw. Lautsprache den Ablenkern G bzw. M, stellt für den Subtest Schriftsprache jedoch einen eigenständigen Ablenker S für Ähnlichkeit im Schriftbild dar.

Tab. 90-C 1.8.1: Durchschnittliche Wahlhäufigkeit der Distraktoren									
Distr.	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
HRP	2,35	1,11	1,87	1,79	1,95	0,43	1,71	1,50	1,58
G	2,82	1,90	4,08	2,82	4,12	3,15	2,53	2,44	3,48
M	2,10	1,01	1,54	2,13	3,82	3,58	3,17	3,21	3,55
I	3,16	1,82	4,16	3,89	5,03	5,06	2,89	3,39	4,97
MOD	2,82	1,90	4,08	1,35	3,58	2,45	3,17	3,21	3,55

Die Wahlhäufigkeiten sind im Koordinatensystem (Abb. 25-C 1.8.1) im Anhang (Anh. 1: C 1.8.1) graphisch dargestellt.

1.8.2 Überprüfung II

Für den Subtestteil GSN wurde eine Clusteranalyse bezüglich der Distraktorenwahl durchgeführt. Die Ergebnisse einer hierarchischen Clusteranalyse legten die Bildung von sieben Clustern für diesen sowie alle weiteren Subtestteile nahe. Die Ergebnisse der Clusterbildung für den ersten Subtestteil werden im Folgenden dargestellt und analysiert:

Tab. 91a-C 1.8.2: Clusteranalyse bzgl. der Distraktorenwahl GSN					
Cluster GSN	Anzahl	HRP	G=MOD	M	I
1	11	3,1	2,0	4,2	3,5
2	27	1,9	3,6	1,8	2,8
3	35	1,0	0,9	0,8	1,3
4	5	3,8	7,4	2,6	5,8
5	14	3,3	4,5	4,1	5,1
6	16	4,1	3,1	1,7	4,6
7	2	4,5	6,5	6,0	7,5

Cluster 1 beschreibt die Zusammenfassung von Ergebnissen, bei denen die plausibleren Ablenkerarten (G und HRP) dieses Subtestteils weniger häufig gewählt wurden als die übrigen.

Die elf Testpersonen dieses Clusters setzen sich wie folgt zusammen: Es handelt sich um sechs (13,6%) der Dritt-, einen (3,3%) der Viert- und vier (11,1%) der Fünftklässler, von denen vier (11,8%) nach oral orientiertem Konzept und sieben (10,0%) mit Gebärden unterrichtet werden. Es sind fünf Mädchen (9,4%) und sechs Jungen (10,5%), die alle hörende Eltern haben. Die betreffenden Testpersonen sind durchschnittlich 10;6 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 95. Es handelt sich um 14,3% aller Bremer, 11,8% aller Berliner, 16,7% aller Schleswiger, 8,3% aller Züricher, 15,4% aller Kölner und 18,2% aller Würzburger Testpersonen. Hamburger und Münsteraner Testpersonen befinden sich mit ihren Testleistungen nicht in diesem Cluster.

Cluster 2 beschreibt die Zusammenfassung der Profile, bei denen besonders häufig gebärdenähnliche Ablenker gewählt, insgesamt jedoch wenig Fehler gemacht wurden. Bei den 27 Testpersonen des Clusters 2 handelt es sich um acht (18,2%) der Dritt-, neun (30,0%) der Viert- und zehn (27,8%) der Fünftklässler, um neun (26,5%) der 34 oral orientiert und 18 (25,7%) der 70 mit Gebärden unterrichteten Testpersonen. Unter den 14 (26,4% aller) Mädchen und 13 (22,8% aller) Jungen ist eine der zehn Testpersonen (10%) mit gehörlosen Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;9 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 95. Es handelt sich um Testpersonen aus allen die Untersuchung betreffenden Schulen, drei (23%) der Hamburger, drei (21,4%) der Bremer, vier (23,5%) der Berliner, drei (25,0%) der Schleswiger, drei (16,7%) der Münsteraner, sechs (50,0%) der Züricher (davon überwiegend Viertklässler), eine (7,7%) der Kölner und vier (36,4%) der Würzburger Testpersonen.

Das Cluster 3 umfasst die Ergebnisse von 35 Testpersonen, in deren Profilen jede Distraktorenart besonders selten gewählt wurde. Unter diesen Testpersonen sind acht (18,2%) der Dritt-, zwölf (40,0%) der Viert- und 15 (41,7%) der Fünftklässler. Unter ihnen sind acht (23,5%) der oral orientiert, 21 (30,0%) der mit Gebärden und alle sechs (100%) bilingual erzogenen Testpersonen. Unter den 15 (28,3% aller) Mädchen und 20 (35,1% aller) Jungen sind neun der zehn Testpersonen (90%) mit gehörlosen Eltern. Die Testpersonen sind durchschnittlich 10;5 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 103. Es sind Testpersonen aus allen acht Städten; sieben (53,8%) der Hamburger, drei (21,4%) der Bremer, vier (23,5%) der Berliner, eine (8,3%) der Schleswiger, neun (50,0%) der Münsteraner, fünf (41,7%) der Züricher, drei (23,1%) der Kölner und drei (27,3%) der Würzburger Testpersonen.

Im Cluster 4 finden sich die Ergebnisse von fünf Testpersonen zusammen, die eine hohe Wahlhäufigkeit aller Distraktorenarten und besonders der gebärdenähnlichen Ablenker aufweisen. Die Testpersonen sind durchschnittlich bereits 11;2 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 82. Es sind drei (6,8%) der Dritt-, und je einer der Viert-

(3,3%) und Fünftklässler (2,8%). Unter ihnen sind drei (5,7% aller) Mädchen und zwei (3,5% aller) Jungen, eine (2,9%) der oral orientiert und vier (5,7%) der mit Gebärden unterrichteten Testpersonen, alle mit hörenden Eltern. Es sind eine (7,1%) der Bremer, eine (8,3%) der Schleswiger, eine (5,5%) der Münsteraner und zwei (15,4%) der Kölner Testpersonen. Testpersonen aus Hamburg, Berlin, Zürich und Würzburg passen mit ihren Ergebnissen nicht in dieses Cluster.

Cluster 5 stellt eine Zusammenfassung der Ergebnisse derjenigen Testpersonen dar, die besonders häufig die inhaltlich ähnlichen Distraktoren gewählt haben. Die Gruppe setzt sich zusammen aus 14 Testpersonen, darunter neun (20,5%) der Dritt-, drei (10,0%) der Viert- und zwei (5,6%) der Fünftklässler. Unter ihnen sind sieben (13,2% aller) Mädchen und sieben (12,3% aller) Jungen, sieben (20,6%) der oral orientiert und sieben (10,0%) der mit Gebärden unterrichteten Testpersonen, alle mit hörenden Eltern. Die Testpersonen dieses Clusters sind durchschnittlich 10;2 Jahre alt und erreichen einen IQ-Wert von durchschnittlich 99 Punkten. Es sind Testpersonen aus allen Städten bis auf Zürich; eine (7,7%) der Hamburger, zwei (14,3%) der Bremer, zwei (11,8%) der Berliner, vier (33,3%) der Schleswiger, zwei (11,1%) der Münsteraner, zwei (15,4%) der Kölner und eine (9,1%) der Würzburger Schüler.

Cluster 6 umfasst die Ergebnisse von 16 Testpersonen, deren Profile eine erhöhte Wahlhäufigkeit der inhaltlich ähnlichen Distraktoren sowie der Ablenker mit besonderer inhaltlicher Ähnlichkeit aufweisen. Die Testpersonen dieses Clusters sind durchschnittlich 10;2 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 97 Punkten. Es handelt sich um acht (18,2%) der Drittklässler und jeweils vier der Viert- (13,3%) und Fünftklässler (11,1%), um fünf (14,7%) der 34 oral orientiert und elf (15,7%) der 70 mit Gebärden unterrichteten Testpersonen. Darunter sind sieben (13,2% aller) Mädchen und neun (15,8% aller) Jungen, alle mit hörenden Eltern. Es sind Testpersonen aus allen Städten bis auf Zürich; zwei (15,4%) der Hamburger, zwei (14,3%) der Bremer, vier (23,5%) der Berliner, einer (8,3%) der Schleswiger, drei (16,7%) der Münsteraner, drei (23,1%) der Kölner und einer (9,1%) der Würzburger Schüler.

In Cluster 7 befinden sich die Ergebnisse von nur zwei Testpersonen, die jede Distraktorenart besonders häufig gewählt haben. Sie sind durchschnittlich 9;8 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 82. Es handelt sich um zwei Mädchen, die beide in die Klassenstufe 3 gehen und hörende Eltern haben. Sie werden beide mit Gebärden unterrichtet und gehen in Bremen bzw. Berlin zur Schule.

Ebenso wurde für den Subtestteil GSV eine Clusteranalyse bezüglich der Distraktorenwahl durchgeführt, die zu folgenden Ergebnissen führte:

Tab. 91b-C 1.8.2: Clusteranalyse bzgl. der Distraktorenwahl GSV					
Cluster GSV	Anzahl	HRP	G=MOD	M	I
1	2	2,5	6,0	2,5	5,5
2	22	2,0	3,6	1,4	2,8
3	3	4,3	5,0	4,7	6,7
4	24	1,4	1,6	0,9	2,3
5	32	0,8	1,6	1,0	1,0
6	25	0,1	0,4	0,1	0,4
7	2	0,5	2,0	3,0	4,5

Cluster 1 setzt sich aus nur zwei Testpersonen zusammen, die auffallend häufig die gebärdenähnlichen Distraktoren gewählt haben. Es handelt sich um zwei Jungen, von denen einer nach oral orientiertem Konzept und einer mit Gebärden unterrichtet wird, die beide die dritte Klasse besuchen und hörende Eltern haben. Sie sind durchschnittlich 10;9 Jahre alt und haben einen durchschnittlichen IQ-Wert von 97 erreicht. Sie gehen in Berlin und Köln zur Schule.

Cluster 2 beschreibt die Zusammenfassung der 22 Profile, bei denen besonders häufig gebärdenähnliche Ablenker gewählt, insgesamt jedoch weniger Fehler gemacht wurden als in Cluster 1. Bei den 22 Testpersonen des Clusters 2 handelt es sich um 14 (31,8%) der Dritt-, fünf (16,7%) der Viert- und drei (8,3%) der Fünftklässler, um neun (26,5%) der 34 oral orientiert, zwölf (17,1%) der 70 mit Gebärden unterrichteten und eine (16,7%) der sechs bilingual erzogenen Testpersonen. Unter den zehn (18,9% aller) Mädchen und zwölf (21,1% aller) Jungen ist eine der zehn Testpersonen (10%) mit gehörlosen Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;0 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 91. Es handelt sich um Testpersonen aus allen betreffenden Schulen außer Münster und Zürich; drei (23%) der Hamburger, vier (28,6%) der Bremer, fünf (29,4%) der Berliner, vier (33,3%) der Schleswiger, vier (30,8%) der Kölner und zwei (18,2%) der Würzburger Testpersonen.

Das Cluster 3 umfasst die Ergebnisse von drei Testpersonen, in deren Profilen jede Distraktorenart besonders häufig gewählt wurde. Unter diesen Testpersonen sind zwei (4,5%) der Dritt- und einer (2,8%) der Fünftklässler. Unter ihnen sind eine (2,9%) der oral orientiert und zwei (2,9%) der mit Gebärden erzogenen Testpersonen. Es sind zwei (3,8% aller) Mädchen und ein Junge (1,8% aller Jungen), alle mit hörenden Eltern. Die Testpersonen sind durchschnittlich bereits 11;0 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von nur 78. Es sind zwei Testpersonen aus Bremen und eine aus Schleswig.

Im Cluster 4 finden sich die Ergebnisse von 24 Testpersonen zusammen, die insgesamt wenig Distraktoren, dabei aber die inhaltlich ähnlichen etwas häufiger als die anderen wählen. Die Testpersonen sind durchschnittlich 10;3 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 99. Es sind elf (25,0%) der Dritt-, sechs (20,0%) der Viert- und sieben (19,4%) der Fünftklässler. Unter ihnen sind neun (17,0% aller) Mädchen und 15 (26,3% aller) Jungen, sieben (20,6%) der oral orientiert, 16 (22,9%) der mit Gebärden und eine (16,7%) der sechs bilingual unterrichteten Testpersonen. Bis auf eine Testperson haben alle hörende Eltern. Es sind Schüler aus allen betreffenden Städten; zwei (15,4%) der Hamburger, zwei (14,3%) der Bremer, fünf (29,4%) der Berliner, drei (25,0%) der Schleswiger, sieben (38,9%) der Münsteraner, einer (8,3%) der Züricher, zwei (15,4%) der Kölner und zwei (18,2%) der Würzburger Schüler.

Cluster 5 fasst die Ergebnisse von 32 Testpersonen zusammen, die insgesamt wenig Distraktoren, dabei aber die gebärdenähnlichen etwas häufiger als die anderen wählten. Die Gruppe setzt sich zusammen aus sieben (15,9%) der Dritt-, zwölf (40,0%) der Viert- und 13 (36,1%) der Fünftklässler. Unter ihnen sind 19 (35,8% aller) Mädchen und 13 (22,8% aller) Jungen, 14 (41,2%) der oral orientiert, 17 (24,3%) der mit Gebärden und eine (16,7%) der bilingual unterrichteten Testpersonen. Zwei von ihnen haben gehörlose Eltern. Die Testpersonen dieses Clusters sind durchschnittlich 10;8 Jahre alt und erreichen einen IQ-Wert von durchschnittlich 100 Punkten. Es sind Testpersonen aus allen Städten; drei (23,1%) der Hamburger, vier (28,6%) der Bremer, drei (17,6%) der Berliner, vier (33,3%) der Schleswiger, drei (16,7%) der Münsteraner, fünf (41,7%) der Züricher, drei (23,1%) der Kölner und sieben (63,6%) der Würzburger Schüler.

Cluster 6 umfasst die Ergebnisse von 25 Testpersonen, deren Profile die niedrigste Wahlhäufigkeit aller Distraktoren aufweisen. Die Testpersonen dieses Clusters sind durchschnittlich 10;9 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 100 Punkten. Es handelt sich um sieben (15,9%) der Drittklässler, sieben (23,3%) der Viert- und elf (30,6%) der Fünftklässler, um eine (2,9%) der 34 oral orientiert, 21 (30,0%) der 70 mit Gebärden und drei (50,0%) der sechs bilingual unterrichteten Testpersonen. Darunter sind 13 (24,5% aller) Mädchen und zwölf (21,1% aller) Jungen sowie sechs (60,0%) der zehn Testpersonen mit gehörlosen Eltern. Es sind Testpersonen aus allen Städten bis auf Schleswig und Würzburg; fünf (38,5%) der Hamburger, einer (7,1%) der Bremer, zwei (11,8%) der Berliner, acht (44,4%) der Münsteraner, sechs (50,0%) der Züricher und drei (23,1%) der Kölner Schüler.

In Cluster 7 befinden sich die Ergebnisse von nur zwei Testpersonen, deren Profil eine deutlich gesteigerte Wahlhäufigkeit bei Distraktoren mit inhaltlicher Ähnlichkeit zeigt. Sie sind durchschnittlich bereits 11;1 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 100. Es handelt sich um zwei Jungen, einen aus Klasse 3, einen aus Klasse 5, mit hörenden

Eltern. Einer von ihnen wird oral orientiert, der andere mit Gebärden unterrichtet, sie gehen in Bremen bzw. Berlin zur Schule.

Die entsprechende Clusteranalyse bezüglich der Distraktorenwahl für den Subtestteil GSA führt zu folgenden Ergebnissen:

Tab. 91c-C 1.8.2: Clusteranalyse bzgl. der Distraktorenwahl GSA					
Cluster GSA	Anzahl	HRP	G=MOD	M	I
1	7	3,4	7,4	2,1	6,3
2	28	1,5	3,0	3,4	3,4
3	21	2,1	5,3	1,4	4,3
4	16	2,4	5,6	3,2	5,8
5	5	2,2	3,4	2,8	7,2
6	8	2,9	7,6	2,4	8,4
7	25	0,8	1,4	0,4	1,3

Cluster 1 umfasst die Ergebnisse von 7 Testpersonen, deren Profile eine auffallend erhöhte Wahlhäufigkeit der gebärden- sowie der inhaltlich ähnlichen Distraktoren aufweisen. Die Testpersonen dieses Clusters sind durchschnittlich 10;6 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 95 Punkten. Es handelt sich um drei (6,8%) der Drittklässler, zwei (6,7%) der Viert- und zwei (5,6%) der Fünftklässler, um zwei (5,9%) der 34 oral orientiert und um fünf (7,1%) der 70 mit Gebärden unterrichteten Testpersonen. Darunter sind zwei (3,8% aller) Mädchen und fünf (8,8% aller) Jungen, von denen alle hörende Eltern haben. Es sind einer (7,7%) der Hamburger, zwei (14,3%) der Bremer, zwei (16,7%) der Schleswiger und zwei (16,7%) der Züricher Schüler. Testpersonen aus Berlin, Münster, Köln und Würzburg passen mit ihren Testergebnissen nicht in dieses Cluster.

Cluster 2 beschreibt die Zusammenfassung von 28 Profilen, die bei sonst geringer Distraktorenwahl im Vergleich zu den übrigen Clustern etwas erhöhte Wahlhäufigkeit bei den mundbildähnlichen Ablenkern aufweisen. Bei den Testpersonen des Clusters 2 handelt es sich um elf (25,0%) der Dritt-, sieben (23,3%) der Viert- und zehn (27,8%) der Fünftklässler, um sechs (17,6%) der 34 oral orientiert, 19 (27,1%) der 70 mit Gebärden unterrichteten und drei (50,0%) der sechs bilingual erzogenen Testpersonen. Unter den 15 (28,3% aller) Mädchen und 13 (22,8% aller) Jungen sind drei der zehn Testpersonen (30%) mit gehörlosen Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;3 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 100. Es handelt sich um Testpersonen aus allen betreffenden Schulen; sechs (46,2%) der Hamburger, zwei (14,3%) der Bremer, zwei (11,8%) der Berliner, eine (8,3%) der Schleswiger, zehn (55,6%) der Münsteraner,

eine (8,3%) der Züricher, drei (23,1%) der Kölner und drei (27,3%) der Würzburger Testpersonen.

Das Cluster 3 umfasst die Ergebnisse von 21 Testpersonen, deren Profile einen ähnlichen, aber durchgehend besseren Verlauf aufweisen als die aus Cluster 1. Unter diesen Testpersonen sind elf (25,0%) der Dritt-, sechs (20,0%) der Viert- und vier (11,1%) der Fünftklässler. Unter ihnen sind neun (26,5%) der oral orientiert und zwölf (17,1%) der mit Gebärden erzogenen Testpersonen. Es sind neun (17,0% aller) Mädchen und zwölf (21,1% aller) Jungen, alle mit hörenden Eltern. Die Testpersonen sind durchschnittlich 10;6 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 95. Es sind Testpersonen aus allen betreffenden Städten; eine (7,7%) der Hamburger, zwei (14,3%) der Bremer, sieben (41,2%) der Berliner, drei (25,0%) der Schleswiger, eine (5,6%) der Münsteraner, drei (25,0%) der Züricher, drei (23,1%) der Kölner und eine (9,1%) der Würzburger Testpersonen.

Im Cluster 4 finden sich die Ergebnisse von 16 Testpersonen zusammen, die auffallend häufig Distraktoren mit Ähnlichkeit in der Gebärde oder im Inhalt wählten. Die Testpersonen sind durchschnittlich 10;8 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 92. Es sind sechs (13,6%) der Dritt-, fünf (16,7%) der Viert- und fünf (13,9%) der Fünftklässler. Unter ihnen sind sieben (13,2% aller) Mädchen und neun (15,8% aller) Jungen, fünf (14,7%) der oral orientiert und elf (15,7%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Es sind Schüler aus allen betreffenden Städten bis auf Würzburg; einer (7,7%) der Hamburger, einer (7,1%) der Bremer, drei (17,6%) der Berliner, vier (33,3%) der Schleswiger, vier (22,2%) der Münsteraner, einer (8,3%) der Züricher und zwei (15,4%) der Kölner Schüler.

Cluster 5 fasst die Ergebnisse von fünf Testpersonen zusammen, die besonders häufig Distraktoren mit Ähnlichkeit im Inhalt wählen. Die Gruppe setzt sich zusammen aus drei (6,8%) der Dritt- und zwei (5,6%) der Fünftklässler. Unter ihnen sind drei (5,7% aller) Mädchen und zwei (3,5% aller) Jungen, drei (8,8%) der oral orientiert und zwei (2,9%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Die Testpersonen dieses Clusters sind durchschnittlich 10;1 Jahre alt und erreichen einen IQ-Wert von durchschnittlich 90 Punkten. Es sind zwei (14,3%) der Bremer, einer (5,9%) der Berliner und zwei (16,7%) der Schleswiger Schüler. Testpersonen aus Hamburg, Münster, Zürich, Köln und Würzburg sind mit ihren Testergebnissen kein Bestandteil dieses Clusters.

Cluster 6 umfasst die Ergebnisse von acht Testpersonen, deren Profile im Vergleich zu den übrigen Clustern die höchste Wahlhäufigkeit der Distraktoren mit Ähnlichkeit im Inhalt oder in der Gebärde aufweisen. Die Testpersonen dieses Clusters sind durchschnittlich 10;7 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 93 Punkten. Es handelt sich um sechs

(13,6%) der Drittklässler, einen (3,3%) der Viert- und einen (2,8%) der Fünftklässler, alle acht (11,4%) aus der Gruppe der 70 mit Gebärden unterrichteten Testpersonen. Darunter sind vier (7,5% aller) Mädchen und vier (7,0% aller) Jungen, alle mit hörenden Eltern. Es sind einer (7,7%) der Hamburger, zwei (14,3%) der Bremer, zwei (11,8%) der Berliner, einer (5,6%) der Münsteraner und zwei (15,4%) der Kölner Schüler.

In Cluster 7 befinden sich die Ergebnisse von 25 Testpersonen, deren Profil die deutlich geringste Wahlhäufigkeit aller Distraktoren zeigt. Sie sind durchschnittlich 10;7 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 105. Es sind vier (9,1%) der Dritt-, neun (30,0%) der Viert- und zwölf (33,3%) der Fünftklässler. Unter ihnen sind 13 (24,5% aller) Mädchen und zwölf (21,1% aller) Jungen, neun (26,5%) der oral orientiert, 13 (18,6%) der mit Gebärden und drei (50,0%) der bilingual unterrichteten Testpersonen. Sieben (70,0%) der zehn Testpersonen mit gehörlosen Eltern sind mit ihrer Leistung Bestandteil dieses Clusters. Insgesamt handelt es sich um Schüler aus allen betreffenden Städten bis auf Schleswig; drei (23,1%) der Hamburger, drei (21,4%) der Bremer, zwei (11,8%) der Berliner, zwei (11,1%) der Münsteraner, fünf (41,7%) der Züricher, drei (23,1%) der Kölner und sieben (63,6%) der Würzburger Schüler.

Die Überprüfung der Distraktorenwahl im Subtest Gebärdensprache verdeutlicht bestimmte Tendenzen. Diese zeichnen sich ab durch die Struktur der Zusammenstellung einzelner Cluster. So bestätigen mehrfach vorkommende Strukturen die ausgeprägten lexikalisch-semantischen Kompetenzen der Hamburger bilingualen Schüler. Dies belegt bspw. die Zusammenstellung des Cluster 1 für den Subtestteil GSN. Hier finden sich in der Gruppe derjenigen, die im Subtest Gebärdensprache die plausibleren Ablenkerarten seltener gewählt haben als die übrigen, keiner der Hamburger bilingualen Schüler. In Cluster 3 für den Subtestteil GSN befinden sich die Ergebnisse von Testpersonen, deren Ergebnisse insgesamt sehr gut sind, da sie jede Distraktorenart selten gewählt haben. Bemerkenswert ist hier, dass sämtliche bilingualen Kinder und neun der zehn Kinder mit gehörlosen Eltern, also Kinder mit muttersprachlichen Gebärdensprachkenntnissen, zu dieser Gruppe zählen und der durchschnittliche IQ 103 beträgt. In die gleiche Richtung weist die Zusammenstellung des Cluster 6 für den Subtestteil GSV, dessen Profile die niedrigste Wahlhäufigkeit aller Distraktoren haben. Dazu zählen 2,9% der oral orientiert, 30,0% der mit Gebärden und 50,0% der bilingual unterrichteten Testpersonen. Darunter sind sechs (60,0%) der zehn Testpersonen mit gehörlosen Eltern. Das Konzept der bilingualen Erziehung gehörloser Kinder übt demnach positiven Einfluss auf ihre passive Wortschatzkenntnis in allen drei Modalitäten aus.

Die Ergebnisse der Untersuchung von Profilverläufen in der Distraktorenwahl wurden insbesondere unter der Fragestellung ausgewertet, ob sich bestimmte Schlüsse bezüglich der Wahlstrategien ziehen lassen. Eine Zusammenschau der diesbezüglichen Cluster legt jedoch

den Schluss nahe, dass sich keine Strategien abzeichnen, die in linearer Beziehung zu Begleitumständen stehen. So befinden sich bspw. in Cluster 2 für den Subtestteil GSN Profile, bei denen besonders häufig gebärdenähnliche Ablenker gewählt, insgesamt jedoch wenig Fehler gemacht wurden. Diese Struktur trifft auf neun (26,5%) der 34 oral orientiert und 18 (25,7%) der 70 mit Gebärden unterrichteten Testpersonen zu, also etwa gleich häufig. Gebärdenorientierte Erziehung bringt demnach keine gebärdenorientierte Wahlstrategie mit sich, die sich auf das Erziehungskonzept zurückführen ließe, da die oral orientierten Testpersonen nicht übertroffen werden. Die Ergebnisse fallen teilweise sogar umgekehrt proportional aus. So zählen in der Untersuchung des Subtestteils GSV (Cluster 2) 26,5% der oral orientiert Erzogenen zu den Testpersonen, die besonders häufig gebärdenähnliche Fehler gemacht haben. Demgegenüber stehen 17,1% der mit Gebärden sowie 16,7% der bilingual erzogenen Schüler. Noch größer ist der Abstand bei den Testpersonen, die insgesamt wenig Distraktoren, die gebärdenähnlichen aber etwas häufiger als die anderen gewählt haben (Cluster 5). Dies trifft zu auf 41,2% der oral orientiert, 24,3% der mit Gebärden und 16,7% der bilingual unterrichteten Testpersonen.

1.8.3 Überprüfung III

Auch für die Modalität Schriftsprache wurden drei Clusteranalysen durchgeführt. Für den Subtestteil SSN führte diese zu folgenden Ergebnissen:

Tab. 92a-C 1.8.3: Clusteranalyse bzgl. der Distraktorenwahl SSN						
Cluster SSN	Anzahl	HRP	G	M	I	MOD (S)
1	22	1,6	1,9	2,5	3,6	1,4
2	9	2,4	5,8	5,0	6,0	3,0
3	17	2,7	3,3	1,1	5,1	0,6
4	7	3,9	2,3	5,0	7,0	3,9
5	5	0,6	2,6	6,4	4,0	4,6
6	11	3,5	3,9	3,2	7,5	1,9
7	37	0,5	0,5	0,2	1,1	0,2

Cluster 1 umfasst die Ergebnisse von 22 Testpersonen, deren Profil eine durchgehend geringe Wahlhäufigkeit aller Distraktoren aufweist. Die Testpersonen dieses Clusters sind durchschnittlich 10;7 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 98 Punkten. Es handelt sich um acht (18,6%) der Drittklässler, vier (13,3%) der Viert- und zehn (28,6%) der Fünftklässler, um neun (26,5%) der 34 oral orientiert und um 13 (19,1%) der 68 mit Gebärden unterrichteten Testpersonen. Darunter sind zehn (19,2% aller) Mädchen und zwölf (21,4% aller) Jungen sowie eine (10,0%) der zehn Testpersonen mit gehörlosen Eltern. Es

sind Testpersonen aus allen betreffenden Schulen außer aus Zürich; einer (7,7%) der Hamburger, drei (23,1%) der Bremer, zwei (12,5%) der Berliner, drei (27,3%) der Schleswiger, acht (47,1%) der Münsteraner, einer (7,7%) der Kölner und vier (30,8%) der Würzburger Schüler.

Cluster 2 beschreibt die Zusammenfassung von neun Profilen, die bei insgesamt erhöhter Wahlhäufigkeit vor allem die gebärdenähnlichen Ablenker häufiger wählten, als die Mitglieder der anderen Cluster. Bei den Testpersonen des Clusters 2 handelt es sich um sechs (14,0%) der Dritt-, zwei (6,7%) der Viert- und einen (2,9%) der Fünftklässler, um zwei (5,9%) der 34 oral orientiert und sieben (10,3%) der 68 mit Gebärden unterrichteten Testpersonen. Von den drei (5,8% aller) Mädchen und sechs (10,7% aller) Jungen haben alle hörende Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt nur 9;11 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 96. Unter den Testpersonen sind eine (7,7%) der Hamburger, eine (6,3%) der Berliner, zwei (18,2%) der Schleswiger, eine (5,9%) der Münsteraner und vier (30,8%) der Kölner Testpersonen. Testpersonen aus Bremen, Zürich und Würzburg gehören mit ihren Ergebnissen nicht in dieses Cluster.

Das Cluster 3 umfasst die Ergebnisse von 17 Testpersonen, deren Profile eine auffallend erhöhte Wahlhäufigkeit der inhaltlich ähnlichen Distraktoren zeigen. Unter diesen Testpersonen sind sieben (16,3%) der Dritt-, fünf (16,7%) der Viert- und fünf (14,3%) der Fünftklässler. Unter ihnen sind sechs (17,6%) der oral orientiert und elf (16,2%) der mit Gebärden erzogenen Testpersonen. Es sind neun (17,3% aller) Mädchen und acht (14,3% aller) Jungen sowie eine (10,0%) der Testpersonen mit gehörlosen Eltern. Die Testpersonen sind durchschnittlich 10;6 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 94. Es sind Testpersonen aus allen betreffenden Städten bis auf Hamburg und Würzburg; drei (23,1%) der Bremer, vier (25,0%) der Berliner, drei (27,3%) der Schleswiger, eine (5,9%) der Münsteraner, vier (33,3%) der Züricher und zwei (15,4%) der Kölner Testpersonen.

Im Cluster 4 finden sich die Ergebnisse von sieben Testpersonen zusammen, die eine erhöhte Wahlhäufigkeit vor allem bei mundbild- und inhaltlich ähnlichen, aber auch bei den modalitätenspezifischen schriftbildähnlichen Distraktoren aufweisen. Die Testpersonen sind durchschnittlich 10;4 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 94. Es sind vier (9,3%) der Dritt- und drei (8,6%) der Fünftklässler. Unter ihnen sind fünf (9,6% aller) Mädchen und zwei (3,6% aller) Jungen, drei (8,8%) der oral orientiert und vier (5,9%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Es handelt sich um einen (7,7%) der Hamburger, zwei (15,4%) der Bremer, zwei (18,2%) der Schleswiger, einen (7,7%) der Kölner und einen (7,7%) der Würzburger Schüler. Schüler aus Berlin, Münster und Zürich sind mit ihren Ergebnissen nicht Bestandteil dieses Clusters.

Cluster 5 fasst die Ergebnisse von fünf Testpersonen zusammen, die bei sonst eher geringer Wahlhäufigkeit besonders häufig Distraktoren mit Ähnlichkeit im Mund- oder Schriftbild wählen. Die Gruppe setzt sich zusammen aus vier (9,3%) der Dritt- und einem (3,3%) der Viertklässler. Unter ihnen sind ein (1,9% aller) Mädchen und vier (7,1% aller) Jungen, alle (7,4%) aus der Gruppe der Testpersonen, die mit Gebärden unterrichtet werden. Alle haben hörende Eltern. Die Testpersonen dieses Clusters sind durchschnittlich 10;7 Jahre alt und erreichen einen IQ-Wert von durchschnittlich 90 Punkten. Es sind zwei (15,4%) der Hamburger, einer (7,7%) der Bremer, einer (6,3%) der Berliner und einer (8,3%) der Züricher Schüler. Testpersonen aus Schleswig, Münster, Köln und Würzburg sind mit ihren Testergebnissen kein Bestandteil dieses Clusters.

Cluster 6 umfasst die Ergebnisse von elf Testpersonen, die besonders häufig Distraktoren mit einfacher oder besonderer Ähnlichkeit im Inhalt wählen. Die Testpersonen dieses Clusters sind durchschnittlich 10;3 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 94 Punkten. Es handelt sich um sechs (14,0%) der Drittklässler, drei (10,1%) der Viert- und zwei (5,7%) der Fünftklässler, eine (2,9%) der oral orientiert und zehn (14,7%) der mit Gebärden unterrichteten Testpersonen. Darunter sind fünf (9,6% aller) Mädchen und sechs (10,7% aller) Jungen, alle mit hörenden Eltern. Es sind einer (7,7%) der Hamburger, zwei (15,4%) der Bremer, zwei (12,5%) der Berliner, einer (9,1%) der Schleswiger, zwei (16,7%) der Züricher und einer (7,7%) der Kölner Schüler. Testpersonen aus Münster und Würzburg passen mit ihren Testergebnissen nicht in dieses Cluster.

In Cluster 7 befinden sich die Ergebnisse von 37 Testpersonen, deren Profil im Vergleich zu den übrigen Clustern die geringste Wahlhäufigkeit aller Distraktoren aufweisen. Sie sind durchschnittlich 10;5 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 104. Es sind acht (18,6%) der Dritt-, 15 (50,0%) der Viert- und 14 (40,0%) der Fünftklässler. Unter ihnen sind 19 (36,5% aller) Mädchen und 18 (32,1% aller) Jungen, 13 (38,2%) der oral orientiert, 18 (26,5%) der mit Gebärden und alle sechs (100,0%) der bilingual unterrichteten Testpersonen. Acht (80,0%) der zehn Testpersonen mit gehörlosen Eltern sind mit ihrer Leistung Bestandteil dieses Clusters. Insgesamt handelt es sich um Schüler aus allen betreffenden Städten bis auf Schleswig; sechs (46,2%) der Hamburger, zwei (15,4%) der Bremer, fünf (31,3%) der Berliner, sieben (41,2%) der Münsteraner, fünf (41,7%) der Züricher, vier (30,8%) der Kölner und acht (61,5%) der Würzburger Schüler.

Für den Subtestteil SSV führte die Clusteranalyse zu folgenden Ergebnissen:

Tab. 92b-C 1.8.3: Clusteranalyse bzgl. der Distraktorenwahl SSV						
Cluster SSV	Anzahl	HRP	G	M	I	MOD (S)
1	8	2,0	6,6	8,6	6,9	8,0
2	12	1,4	5,6	5,8	7,0	5,6
3	12	2,9	5,8	3,4	7,2	3,2
4	21	1,7	3,5	2,4	3,8	2,3
5	26	1,1	1,7	0,9	2,0	0,8
6	10	3,0	6,7	6,4	10,0	6,2
7	19	1,6	3,7	5,0	4,6	4,5

Cluster 1 umfasst die Ergebnisse von acht Testpersonen, deren Profil eine auffallend hohe Wahlhäufigkeit der Distraktoren mit Ähnlichkeit in der Gebärde, im Mund- sowie im Schriftbild aufweist. Die Testpersonen dieses Clusters sind durchschnittlich 10;5 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 91 Punkten. Es handelt sich um sieben (16,3%) der Drittklässler und einen (2,9%) der Fünftklässler, um zwei (5,9%) der 34 oral orientiert und um sechs (8,8%) der 68 mit Gebärden unterrichteten Testpersonen. Darunter sind drei (5,8% aller) Mädchen und fünf (8,9% aller) Jungen, alle mit hörenden Eltern. Es handelt sich um drei (23,1%) der Hamburger, einen (7,7%) der Bremer, zwei (12,5%) der Berliner, einen (7,7%) der Kölner und einen (7,7%) der Würzburger Schüler. Testpersonen aus Schleswig, Münster und Zürich passen mit ihren Ergebnissen nicht in dieses Cluster.

Cluster 2 beschreibt die Zusammenfassung von zwölf Profilen, die bei insgesamt erhöhter Wahlhäufigkeit vor allem die inhaltlich ähnlichen Ablenker häufig wählten. Bei den Testpersonen des Clusters 2 handelt es sich um fünf (11,6%) der Dritt-, vier (13,3%) der Viert- und drei (8,6%) der Fünftklässler, um zwei (5,9%) der 34 oral orientiert und zehn (14,7%) der 68 mit Gebärden unterrichteten Testpersonen. Von den sieben (13,5% aller) Mädchen und fünf (8,9% aller) Jungen haben alle hörende Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;7 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 97. Unter den Testpersonen sind zwei (15,4%) der Hamburger, eine (7,7%) der Bremer, zwei (18,2%) der Schleswiger, zwei (11,8%) der Münsteraner, drei (25,0%) der Züricher und zwei (15,4%) der Kölner Testpersonen. Berliner und Würzburger Testpersonen sind nicht Mitglieder dieses Clusters.

Das Cluster 3 umfasst die Ergebnisse von zwölf Testpersonen, deren Profile eine auffallend erhöhte Wahlhäufigkeit der inhaltlich ähnlichen Distraktoren zeigen. Unter diesen Testpersonen sind sieben (16,3%) der Dritt-, einer (3,3%) der Viert- und vier (11,4%) der Fünft-

klässler. Unter ihnen sind drei (8,8%) der oral orientiert und neun (13,2%) der mit Gebärden erzogenen Testpersonen. Es sind fünf (9,6% aller) Mädchen und sieben (12,5% aller) Jungen, alle mit hörenden Eltern. Die Testpersonen sind durchschnittlich 10;3 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 92. Es handelt sich um zwei (15,4%) der Hamburger, drei (23,1%) der Bremer, drei (18,8%) der Berliner, zwei (18,2%) der Schleswiger, eine (8,3%) der Züricher und eine (7,7%) der Kölner Testpersonen. Schüler aus Münster und Würzburg gehören mit ihren Ergebnissen nicht in dieses Cluster.

Im Cluster 4 finden sich die Ergebnisse von 21 Testpersonen zusammen, die eine geringe Wahlhäufigkeit bei allen Distraktorenarten aufweisen. Die Testpersonen sind durchschnittlich 10;7 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 100. Es sind vier (9,3%) der Dritt-, sieben (23,3%) der Viert- und zehn (14,7%) der Fünftklässler. Unter ihnen sind 17 (32,7% aller) Mädchen und vier (7,1% aller) Jungen, sieben (20,6%) der oral orientiert, zwölf (17,6%) der mit Gebärden und zwei (33,3%) der bilingual unterrichteten Testpersonen. Unter ihnen sind vier (40,0%) der Testpersonen mit gehörlosen Eltern. Insgesamt handelt es sich um Testpersonen aus allen betreffenden Städten; zwei (15,4%) der Hamburger, vier (30,8%) der Bremer, einen (6,3%) der Berliner, drei (27,3%) der Schleswiger, drei (17,6%) der Münsteraner, drei (25,0%) der Züricher, zwei (15,4%) der Kölner und drei (23,1%) der Würzburger Schüler.

Cluster 5 fasst die Ergebnisse von 26 Testpersonen zusammen, die im Vergleich zu allen anderen Clustern die durchgängig niedrigste Wahlhäufigkeit zeigen. Die Gruppe setzt sich zusammen aus vier (9,3%) der Dritt-, einem (3,3%) der Viert- und einem (2,9%) der Fünftklässler. Unter ihnen sind zwölf (23,1% aller) Mädchen und 14 (25,0% aller) Jungen, acht (23,5%) der oral orientiert, 14 (20,6%) der mit Gebärden und vier (66,7%) der bilingual unterrichteten Testpersonen. Sechs (60,0%) der zehn Testpersonen mit gehörlosen Eltern sind mit ihrer Leistung Bestandteil dieses Clusters. Die Testpersonen dieses Clusters sind durchschnittlich 10;5 Jahre alt und erreichen einen IQ-Wert von durchschnittlich 104 Punkten. Es sind Testpersonen aus allen Städten bis auf Schleswig; vier (30,8%) der Hamburger, einer (7,7%) der Bremer, drei (18,8%) der Berliner, sieben (41,2%) der Münsteraner, drei (25,0%) der Züricher, drei (23,1%) der Kölner und fünf (38,5%) der Würzburger Schüler.

Cluster 6 umfasst die Ergebnisse von zehn Testpersonen, die bei insgesamt erhöhter Wahlhäufigkeit vor allem die einfach und die besonders inhaltlich ähnlichen Distraktoren häufiger wählen, als die Mitglieder der anderen Cluster. Die Testpersonen dieses Clusters sind durchschnittlich 10;4 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 93 Punkten. Es handelt sich um fünf (11,6%) der Drittklässler, drei (10,1%) der Viert- und zwei (5,7%) der Fünftklässler, drei (8,8%) der oral orientiert und sieben (10,3%) der mit Gebärden unterrichteten Testpersonen. Darunter sind fünf (9,6% aller) Mädchen und fünf (8,9% aller) Jun-

gen, alle mit hörenden Eltern. Es sind drei (23,1%) der Bremer, drei (18,8%) der Berliner, einer (9,1%) der Schleswiger, einer (5,9%) der Münsteraner und zwei (15,4%) der Kölner Schüler. Testpersonen aus Hamburg, Zürich und Würzburg passen mit ihren Testergebnissen nicht in dieses Cluster.

In Cluster 7 befinden sich die Ergebnisse von 19 Testpersonen, deren Profil eine erhöhte Wahlhäufigkeit der mundbild- und der modalitätenspezifischen schriftbildähnlichen Distraktoren aufweist. Die Testpersonen sind durchschnittlich 10;2 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 98. Es sind elf (25,6%) der Dritt-, vier (13,3%) der Viert- und vier (11,4%) der Fünftklässler. Unter ihnen sind sieben (13,5% aller) Mädchen und zwölf (21,4% aller) Jungen, neun (26,5%) der oral orientiert und zehn (14,7%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Es handelt sich um Schüler aus allen betreffenden Städten bis auf Hamburg und Bremen; vier (25,0%) der Berliner, drei (27,3%) der Schleswiger, vier (23,5%) der Münsteraner, sechs (50,0%) der Züricher, zwei (15,4%) der Kölner und vier (30,8%) der Würzburger Schüler.

Für den Subtestteil SSA führte die Clusteranalyse zu folgenden Ergebnissen:

Tab. 92c-C 1.8.3: Clusteranalyse bzgl. der Distraktorenwahl SSA						
Cluster SSA	Anzahl	HRP	G	M	I	MOD (S)
1	14	0,9	3,6	7,8	7,6	5,5
2	17	0,4	2,6	1,7	4,4	1,2
3	11	0,5	6,8	6,6	6,6	4,5
4	22	0,4	3,6	4,6	5,5	2,5
5	1	2,0	7,0	9,0	12,0	8,0
6	30	0,1	0,5	0,4	1,2	0,3
7	13	0,8	5,5	5,5	9,4	3,5

Cluster 1 umfasst die Ergebnisse von 14 Testpersonen, deren Profil (mit Ausnahme der einen Testperson aus Cluster 5) für die Distraktoren mit Ähnlichkeit im Mund- und im Schriftbild die höchste Wahlhäufigkeit zeigt. Die Testpersonen dieses Clusters sind durchschnittlich 10;2 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 97 Punkten. Es handelt sich um neun (20,9%) der Drittklässler, zwei (6,7%) der Viert- und drei (8,6%) der Fünftklässler, um vier (11,8%) der 34 oral orientiert und um zehn (14,7%) der 68 mit Gebärden unterrichteten Testpersonen. Darunter sind fünf (9,6% aller) Mädchen und neun (16,1% aller) Jungen, alle mit hörenden Eltern. Es handelt sich um Testpersonen aus allen Städten; einen (7,7%) der Hamburger, zwei (15,4%) der Bremer, einen (6,3%) der Berliner, drei (27,3%) der

Schleswiger, drei (17,6%) der Münsteraner, einen (8,3%) der Züricher, zwei (15,4%) der Kölner und einen (7,7%) der Würzburger Schüler.

Cluster 2 beschreibt die Zusammenfassung von 17 Profilen, innerhalb derer alle Distraktorenarten selten gewählt wurden. Bei den Testpersonen des Clusters 2 handelt es sich um sechs (14,0%) der Dritt-, zwei (6,7%) der Viert- und neun (25,7%) der Fünftklässler, um sechs (17,6%) der 34 oral orientiert und elf (16,2%) der 68 mit Gebärden unterrichteten Testpersonen. Unter den zehn (19,2% aller) Mädchen und sieben (12,5% aller) Jungen befindet sich eine (10,0%) der Testpersonen mit gehörlosen Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;6 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 105. Es handelt sich um Testpersonen aus allen Städten bis auf Hamburg; einen (7,7%) der Bremer, einen (6,3%) der Berliner, zwei (18,2%) der Schleswiger, drei (17,6%) der Münsteraner, vier (33,3%) der Züricher, drei (23,1%) der Kölner und drei (23,1%) der Würzburger Schüler.

Das Cluster 3 umfasst die Ergebnisse von elf Testpersonen, deren Profile eine auffallend erhöhte Wahlhäufigkeit der gebärdenähnlichen Distraktoren zeigen. Unter diesen Testpersonen sind fünf (11,6%) der Dritt-, fünf (16,7%) der Viert- und eine (2,9%) der Fünftklässler. Unter ihnen sind eine (2,9%) der oral orientiert und zehn (14,7%) der mit Gebärden erzogenen Testpersonen. Es sind sechs (11,5% aller) Mädchen und fünf (8,9% aller) Jungen sowie eine (10,0%) der Testpersonen mit gehörlosen Eltern. Die Testpersonen sind durchschnittlich 10;4 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 93. Es handelt sich um Testpersonen aus allen Städten bis auf Würzburg; einen (7,7%) der Hamburger, einen (7,7%) der Bremer, zwei (12,5%) der Berliner, einen (9,1%) der Schleswiger, einen (5,9%) der Münsteraner, drei (25,0%) der Züricher und zwei (15,4%) der Kölner Schüler.

Im Cluster 4 finden sich die Ergebnisse von 22 Testpersonen zusammen, die eine erhöhte Wahlhäufigkeit bei den inhaltlich ähnlichen Distraktoren aufweisen. Die Testpersonen sind durchschnittlich 10;6 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 94. Es sind zehn (23,3%) der Dritt-, drei (10,0%) der Viert- und neun (25,7%) der Fünftklässler. Unter ihnen sind 13 (25,0% aller) Mädchen und neun (16,1% aller) Jungen, acht (23,5%) der oral orientiert, 13 (19,1%) der mit Gebärden und eine (16,7%) der bilingual unterrichteten Testpersonen. Alle haben hörende Eltern. Insgesamt handelt es sich um Testpersonen aus allen betreffenden Städten; drei (23,1%) der Hamburger, vier (30,8%) der Bremer, vier (25,0%) der Berliner, zwei (18,2%) der Schleswiger, vier (23,5%) der Münsteraner, einen (8,3%) der Züricher, einen (7,7%) der Kölner und drei (23,1%) der Würzburger Schüler.

Cluster 5 umfasst die Ergebnisse von lediglich einer Testperson. Das Profil dieser Testperson zeigt im Bezug zu allen anderen Profilen für alle Distraktorenarten die höchste Wahlhäufig-

keit. Damit unterscheidet sich ihr Distraktorenprofil so deutlich von denen der übrigen Testpersonen, dass es keine Möglichkeit der Zusammenfassung zu einem Cluster gab.

Cluster 6 umfasst die Ergebnisse von 30 Testpersonen, deren Profil im Vergleich zu den übrigen Clustern die geringste Wahlhäufigkeit aller Distraktoren aufweisen. Die Testpersonen dieses Clusters sind durchschnittlich 10;6 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 104 Punkten. Es handelt sich um drei (7,0%) der Drittklässler, 16 (53,3%) der Viert- und elf (31,4%) der Fünftklässler, elf (32,4%) der oral orientiert, 14 (20,6%) der mit Gebärden und fünf (83,3%) der bilingual unterrichteten Testpersonen. Darunter sind 13 (25,0% aller) Mädchen und 17 (30,4% aller) Jungen sowie neun (90,0%) der Testpersonen mit gehörlosen Eltern. Insgesamt handelt es sich um Testpersonen aus allen betreffenden Städten außer Schleswig; fünf (38,5%) der Hamburger, drei (23,1%) der Bremer, fünf (31,3%) der Berliner, sechs (35,3%) der Münsteraner, zwei (15,4%) der Züricher, drei (23,1%) der Kölner und sechs (46,2%) der Würzburger Schüler.

In Cluster 7 befinden sich die Ergebnisse von 13 Testpersonen, deren Profil ähnlich, jedoch mit durchgehend höherer Wahlhäufigkeit als das der Testpersonen in Cluster 4 verläuft. Die Testpersonen sind durchschnittlich 10;6 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 89. Es sind neun (20,9%) der Dritt-, zwei (6,7%) der Viert- und zwei (5,7%) der Fünftklässler. Unter ihnen sind fünf (9,6% aller) Mädchen und acht (14,3% aller) Jungen, vier (11,8%) der oral orientiert und neun (13,2%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Es sind zwei (15,4%) der Hamburger, zwei (15,4%) der Bremer, drei (18,8%) der Berliner, drei (27,3%) der Schleswiger, einer (8,3%) der Züricher und zwei (15,4%) der Kölner Schüler. Testpersonen aus Münster und Würzburg sind mit ihren Ergebnissen nicht Bestandteil dieses Clusters.

Eine Zusammenschau der Häufungen innerhalb der Cluster aus der Untersuchung des Subtests Schriftsprache ergibt zweierlei: Zum einen befinden sich die Hamburger bilingualen Schüler in der überwiegenden Anzahl in jedem der drei Subtestteile in dem Cluster, das die Ergebnisse der Testpersonen zusammenfasst, die die geringste Wahlhäufigkeit aller Distraktoren aufweist. So stehen in Cluster 7 des Subtestteils Schriftsprache/Nomen den sechs bilingualen Schülern, also 100%, 38,2% der oralen und 26,5% der gebärdenorientierten erzo-genen Schülern gegenüber. In Cluster 5 des Subtestteils Schriftsprache/Verben setzt sich die Gruppe zusammen aus 23,5% der oral orientiert, 20,6% der mit Gebärden und 66,7% der bilingual unterrichteten Testpersonen. In Cluster 6 des Subtestteils Schriftsprache/Adjektive schließlich befinden sich 32,4% der oral orientiert, 20,6% der mit Gebärden und 83,3% der bilingual unterrichteten Testpersonen. Diese Ergebnisse bestätigen die Tendenz im Subtest Gebärdensprache. In Bezug auf die Schriftsprache erhalten sie jedoch besondere Relevanz aufgrund ihrer schwereren Erlernbarkeit der Gebärdensprache gegenüber.

Die zweite Tendenz, die sich abzeichnet, betrifft die Cluster, in denen Ergebnisse von Personen zusammengefasst werden, die überwiegend die mundbild- sowie die modalitätenspezifischen schriftbildähnlichen Distraktoren gewählt haben. In Hinblick auf die qualitativ differenzierte Konstruktion der Testaufgaben wären Häufungen oder sonstige Besonderheiten innerhalb dieser Cluster interessant gewesen. Die Zusammenschau der drei betreffenden Cluster (4/Nomen, 7/Verben, 1/Adjektive) ergibt jedoch keine derartigen Häufungen oder Besonderheiten, sondern die Ergebnisse bleiben unspezifisch. Demnach lassen sich keine Hinweise darauf ausmachen, dass durch die Aufgabenkonstruktion gesetzmäßige Wahlstrategien nachvollziehbar werden, die ihrerseits in linearer Verbindung zu bestimmten Begleitmerkmalen stehen.

1.8.4 Überprüfung IV

Ebenso wurden für die Modalität Lautsprache drei Clusteranalysen durchgeführt. Für den Subtestteil LSN führte die Clusteranalyse zu folgenden Ergebnissen:

Tab. 93a-C 1.8.4: Clusteranalyse bzgl. der Distraktorenwahl LSN					
Cluster LSN	Anzahl	HRP	G	M=MOD	I
1	5	4,8	7,4	4,8	8,6
2	43	0,6	0,9	1,8	0,9
3	3	5,0	3,3	4,7	8,0
4	21	1,9	2,7	1,9	3,0
5	6	2,8	5,0	8,3	5,8
6	10	2,4	5,7	4,5	4,3
7	18	2,3	1,8	4,8	3,3

Cluster 1 umfasst die Ergebnisse von fünf Testpersonen, deren Profil bei durchgehend hoher Wahlhäufigkeit auffallend erhöhte Werte bei den Distraktoren mit Ähnlichkeit im Inhalt und in der Gebärde zeigt. Die Testpersonen dieses Clusters sind durchschnittlich 10;3 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 85 Punkten. Es handelt sich um zwei (5,1%) der Drittklässler, zwei (6,5%) der Viert- und einen (2,8%) der Fünftklässler, alle (7,6%) aus der Gruppe der 68 mit Gebärden unterrichteten Testpersonen. Darunter sind drei (5,9% aller) Mädchen und zwei (3,6% aller) Jungen, alle mit hörenden Eltern. Es handelt sich um zwei (15,4%) der Bremer, einen (7,1%) der Berliner, einen (8,3%) der Züricher und einen (8,3%) der Kölner Schüler. Testpersonen aus Hamburg, Schleswig, Münster und Würzburg gehören mit ihren Ergebnissen nicht in dieses Cluster.

Cluster 2 beschreibt die Zusammenfassung von 43 Profilen, deren Distraktorenwahl in allen Fällen am niedrigsten ausfällt. Bei den Testpersonen des Clusters 2 handelt es sich um elf

(28,2%) der Dritt-, 17 (54,8%) der Viert- und 15 (41,7%) der Fünftklässler, um 17 (50,0%) der 34 oral orientiert, 23 (34,8%) der 66 mit Gebärden und 3 (50,0%) der sechs bilingual unterrichteten Testpersonen. Unter den 16 (31,4% aller) Mädchen und 27 (49,1% aller) Jungen befinden sich acht (80,0%) der Testpersonen mit gehörlosen Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;6 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 102. Es handelt sich um Testpersonen aus allen Städten; drei (23,1%) der Hamburger, fünf (38,5%) der Bremer, sechs (42,9%) der Berliner, einen (8,3%) der Schleswiger, neun (50,0%) der Münsteraner, fünf (41,7%) der Züricher, vier (33,3%) der Kölner und zehn (83,3%) der Würzburger Schüler.

Das Cluster 3 umfasst die Ergebnisse von drei Testpersonen, deren Profile im Vergleich zu den anderen Clustern die höchste Wahlhäufigkeit der besonders inhaltlich ähnlichen Distraktoren zeigen. Unter diesen Testpersonen sind zwei (5,1%) der Dritt- und einer (2,8%) der Fünftklässler, alle (4,5%) aus der Gruppe der mit Gebärden erzogenen Testpersonen. Es sind ein (2,0% aller) Mädchen und zwei (3,6% aller) Jungen, alle mit hörenden Eltern. Die Testpersonen sind durchschnittlich erst 9;9 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 91. Es handelt sich um zwei (15,4%) der Hamburger und einen (8,3%) der Kölner Schüler.

Im Cluster 4 finden sich die Ergebnisse von 21 Testpersonen zusammen, die im Vergleich zu den Mitgliedern aller anderen Cluster durchgehend selten Distraktoren wählen. Die Testpersonen sind durchschnittlich 10;7 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 95. Es sind sieben (17,9%) der Dritt-, vier (12,9%) der Viert- und zehn (27,8%) der Fünftklässler. Unter ihnen sind elf (21,6% aller) Mädchen und zehn (18,2% aller) Jungen, sieben (20,6%) der oral orientiert, zwölf (18,2%) der mit Gebärden und zwei (33,3%) der bilingual unterrichteten Testpersonen. Unter ihnen sind zwei (20,0%) der Testpersonen mit gehörlosen Eltern. Insgesamt handelt es sich um Testpersonen aus allen betreffenden Städten; zwei (15,4%) der Hamburger, drei (23,1%) der Bremer, drei (21,4%) der Berliner, drei (25,0%) der Schleswiger, drei (16,7%) der Münsteraner, drei (25,0%) der Züricher, drei (25,0%) der Kölner und einen (8,3%) der Würzburger Schüler.

Cluster 5 fasst die Ergebnisse von sechs Testpersonen zusammen, die im Vergleich zu den Testpersonen der anderen Cluster die modalitätenspezifischen mundbildähnlichen Distraktoren am häufigsten wählen. Die Gruppe setzt sich zusammen aus vier (10,3%) der Dritt- und zwei (5,6%) der Fünftklässler. Unter ihnen sind zwei (3,9% aller) Mädchen und vier (7,3% aller) Jungen, eine (2,9%) der oral orientiert und fünf (7,6%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Die Testpersonen dieses Clusters sind durchschnittlich 10;7 Jahre alt und erreichen einen IQ-Wert von durchschnittlich 98 Punkten. Es sind Testpersonen aus allen Städten bis auf Zürich und Würzburg; einer (7,7%) der Hambur-

ger, einer (7,7%) der Bremer, drei (21,4%) der Berliner, einer (8,3%) der Schleswiger, einer (5,6%) der Münsteraner und einer (8,3%) der Kölner Schüler.

Cluster 6 umfasst die Ergebnisse von zehn Testpersonen, deren Profil erhöhte Wahlhäufigkeit der gebärdenähnlichen Distraktoren zeigt. Die Testpersonen dieses Clusters sind durchschnittlich 10;7 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 98 Punkten. Es handelt sich um sechs (15,4%) der Drittklässler, einen (3,2%) der Viert- und drei (8,3%) der Fünftklässler, vier (11,8%) der oral orientiert und sechs (9,1%) der mit Gebärden unterrichteten Testpersonen. Darunter sind sechs (11,8% aller) Mädchen und vier (7,3% aller) Jungen, alle mit hörenden Eltern. Insgesamt handelt es sich um zwei (15,4%) der Hamburger, einen (7,7%) der Bremer, zwei (14,3%) der Berliner, vier (33,3%) der Schleswiger und einen (8,3%) der Züricher Schüler. Testpersonen aus Münster, Köln und Würzburg passen nicht in dieses Cluster.

In Cluster 7 befinden sich die Ergebnisse von 18 Testpersonen, deren Profil insgesamt niedrige, jedoch leicht erhöhte Wahlhäufigkeit bei den mundbildähnlichen Distraktoren zeigt. Die Testpersonen sind durchschnittlich 10;3 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 97. Es sind sieben (17,9%) der Dritt-, sechs (19,4%) der Viert- und fünf (13,9%) der Fünftklässler. Unter ihnen sind zwölf (23,5% aller) Mädchen und sechs (10,9% aller) Jungen, fünf (14,7%) der oral orientiert, zwölf (18,2%) der mit Gebärden und eine (16,7%) der bilingual unterrichteten Testpersonen. Alle haben hörende Eltern. Es handelt sich um Testpersonen aus allen Städten; drei (23,1%) der Hamburger, einen (7,7%) der Bremer, einen (7,1%) der Berliner, drei (25,0%) der Schleswiger, fünf (27,8%) der Münsteraner, zwei (16,7%) der Züricher, zwei (16,7%) der Kölner und einen (8,3%) der Würzburger Schüler.

Die Clusteranalyse für den Subtestteil LSV führte zu folgenden Ergebnissen:

Tab. 93b-C 1.8.4: Clusteranalyse bzgl. der Distraktorenwahl LSV					
Cluster LSV	Anzahl	HRP	G	M=MOD	I
1	16	2,9	2,9	4,9	6,9
2	21	2,5	3,9	3,0	4,4
3	36	0,7	1,4	1,2	1,1
4	15	1,3	3,7	6,5	4,9
5	1	3,0	8,0	11,0	7,0
6	11	0,7	2,0	3,5	2,3
7	6	1,0	5,2	4,2	1,5

Cluster 1 umfasst die Ergebnisse von 16 Testpersonen, deren Profil nach dem der einen Testperson aus Cluster 5 die höchsten Werte vor allem in den Fällen der einfachen und be-

sonderen inhaltlichen und der mundbildähnlichen Distraktoren zeigt. Die Testpersonen dieses Clusters sind durchschnittlich 10;5 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 98 Punkten. Es handelt sich um acht (20,5%) der Drittklässler, sechs (19,4%) der Viert- und zwei (5,6%) der Fünftklässler, um fünf (14,7%) der 34 oral orientiert und elf (16,7%) der 66 mit Gebärden unterrichteten Testpersonen. Darunter sind sieben (13,7% aller) Mädchen und neun (16,4% aller) Jungen, alle mit hörenden Eltern. Es handelt sich um Testpersonen aus allen Städten bis auf Würzburg; drei (23,1%) der Hamburger, einen (7,7%) der Bremer, zwei (14,3%) der Berliner, vier (33,3%) der Schleswiger, einen (5,6%) der Münsteraner, zwei (16,7%) der Züricher und drei (25,0%) der Kölner Schüler.

Cluster 2 beschreibt die Zusammenfassung von 21 Profilen, denen ein erhöhte Wahlhäufigkeit der inhaltlich ähnlichen Distraktoren gemein ist. Bei den Testpersonen des Clusters 2 handelt es sich um sieben (17,9%) der Dritt-, fünf (16,1%) der Viert- und neun (25,0%) der Fünftklässler, um sechs (17,6%) der 34 oral orientiert, 14 (21,2%) der 66 mit Gebärden und eine (16,7%) der sechs bilingual unterrichteten Testpersonen. Unter den elf (21,6% aller) Mädchen und zehn (18,2% aller) Jungen befindet sich eine (10,0%) der Testpersonen mit gehörlosen Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;6 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 95. Es handelt sich um Testpersonen aus allen Städten; drei (23,1%) der Hamburger, drei (23,1%) der Bremer, einen (7,1%) der Berliner, vier (33,3%) der Schleswiger, drei (16,7%) der Münsteraner, drei (25,0%) der Züricher, drei (25,0%) der Kölner und einen (8,3%) der Würzburger Schüler.

Das Cluster 3 umfasst die Ergebnisse von 36 Testpersonen, deren Distraktorenwahl in allen Fällen am niedrigsten ausfällt. Unter diesen Testpersonen sind sieben (17,9%) der Dritt-, 15 (48,4%) der Viert- und 14 (38,9%) der Fünftklässler, 15 (44,1%) der 34 oral orientiert, 15 (22,7%) der 66 mit Gebärden und fünf (83,3%) der sechs bilingual unterrichteten Testpersonen. Es sind 18 (35,3% aller) Mädchen und 18 (32,7% aller) Jungen, darunter sieben (70,0%) der Testpersonen mit gehörlosen Eltern. Die Testpersonen sind durchschnittlich 10;6 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 102. Es handelt sich um Testpersonen aus allen Städten bis auf Schleswig; fünf (38,5%) der Hamburger, zwei (15,4%) der Bremer, sechs (42,9%) der Berliner, sechs (33,3%) der Münsteraner, fünf (41,7%) der Züricher, drei (25,0%) der Kölner und neun (75,0%) der Würzburger Schüler.

Im Cluster 4 finden sich die Ergebnisse von 15 Testpersonen zusammen, deren Profil erhöhte Wahlhäufigkeit der modalitätenspezifischen, mundbildähnlichen Distraktoren zeigt. Die Testpersonen sind durchschnittlich 10;4 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 97. Es sind sechs (15,4%) der Dritt-, vier (12,9%) der Viert- und fünf (13,9%) der Fünftklässler. Unter ihnen sind sieben (13,7% aller) Mädchen und acht (14,5% aller) Jungen, vier (11,8%) der oral orientiert und elf (16,7%) der mit Gebärden unterrichteten

Testpersonen. Unter ihnen ist eine (10,0%) der Testpersonen mit gehörlosen Eltern. Insgesamt handelt es sich um Testpersonen aus allen betreffenden Städten bis auf Würzburg; einen (7,7%) der Hamburger, drei (23,1%) der Bremer, zwei (14,3%) der Berliner, drei (25,0%) der Schleswiger, drei (16,7%) der Münsteraner, zwei (25,0%) der Züricher und einen (8,3%) der Kölner Schüler.

Cluster 5 umfasst die Ergebnisse von lediglich einer Testperson. Das Profil dieser Testperson zeigt im Bezug zu allen anderen Profilen für alle Distraktorenarten (bis auf die mit besonderer inhaltlicher Ähnlichkeit) die höchste Wahlhäufigkeit. Damit unterscheidet sich ihr Distraktorenprofil so deutlich von denen der übrigen Testpersonen, dass es keine Möglichkeit der Zusammenfassung zu einem Cluster gab.

Cluster 6 umfasst die Ergebnisse von elf Testpersonen, deren Profil eine durchgehend niedrige Wahlhäufigkeit aller Distraktoren zeigt. Die Testpersonen dieses Clusters sind durchschnittlich bereits 10;10 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 98 Punkten. Es handelt sich um sieben (15,4%) der Dritt- und vier (11,1%) der Fünftklässler, drei (8,8%) der oral orientiert und acht (12,1%) der mit Gebärden unterrichteten Testpersonen. Darunter sind vier (7,8% aller) Mädchen und sieben (12,7% aller) Jungen, alle mit hörenden Eltern. Insgesamt handelt es sich um einen (7,7%) der Bremer, zwei (14,3%) der Berliner, vier (33,3%) der Münsteraner, zwei (16,7%) der Kölner und zwei (16,7%) der Würzburger Schüler. Testpersonen aus Hamburg, Schleswig und Zürich passen mit ihren Testergebnissen nicht in dieses Cluster.

In Cluster 7 befinden sich die Ergebnisse von sechs Testpersonen, deren Profil im Vergleich zu denen der Mitglieder anderer Cluster die höchste Wahlhäufigkeit von Distraktoren mit Gebärdenähnlichkeit zeigt. Die Testpersonen sind durchschnittlich 10;0 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 97. Es sind drei (7,7%) der Dritt-, einer (3,2%) der Viert- und zwei (5,6%) der Fünftklässler. Unter ihnen sind drei (5,9% aller) Mädchen und drei (5,5% aller) Jungen, eine (2,9%) der oral orientiert und fünf (7,6%) der mit Gebärden unterrichteten Testpersonen. Unter ihnen ist eine (10,0%) der Testpersonen mit gehörlosen Eltern. Es handelt sich um einen (7,7%) der Hamburger, drei (23,1%) der Bremer, einen (8,3%) der Schleswiger und einen (5,6%) der Münsteraner Schüler. Testpersonen aus Berlin, Zürich, Köln und Würzburg sind nicht Bestandteil dieses Clusters.

Die Ergebnisse der Clusteranalyse des Subtestteils LSA sind im Folgenden dargestellt:

Tab. 93c-C 1.8.4: Clusteranalyse bzgl. der Distraktorenwahl LSA					
Cluster LSA	Anzahl	HRP	G	M=MOD	I
1	20	1,4	3,0	2,3	4,0
2	11	3,0	6,5	5,2	9,0
3	29	0,4	0,8	0,9	1,3
4	15	2,8	5,5	3,7	6,2
5	18	1,8	4,3	6,7	7,7
6	3	3,7	7,7	7,7	11,0
7	10	1,0	3,3	4,8	3,5

Cluster 1 umfasst die Ergebnisse von 20 Testpersonen, die bei sonst geringer Wahlhäufigkeit verstärkt die inhaltlich ähnlichen Distraktoren wählten. Die Testpersonen dieses Clusters sind durchschnittlich 10;7 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 94 Punkten. Es handelt sich um sechs (15,4%) der Drittklässler, sechs (19,4%) der Viert- und acht (22,2%) der Fünftklässler, um sieben (20,6%) der 34 oral orientiert und 13 (19,7%) der 66 mit Gebärden unterrichteten Testpersonen. Darunter sind zehn (19,6% aller) Mädchen und zehn (18,2% aller) Jungen, sowie vier (40,0%) der Testpersonen mit gehörlosen Eltern. Es handelt sich um Testpersonen aus allen Städten bis auf Hamburg; drei (23,1%) der Bremer, zwei (14,3%) der Berliner, drei (25,0%) der Schleswiger, vier (22,2%) der Münsteraner, vier (33,3%) der Züricher, zwei (16,7%) der Kölner und zwei (16,7%) der Würzburger Schüler.

Cluster 2 beschreibt die Zusammenfassung von elf Profilen, denen eine auffallend erhöhte Wahlhäufigkeit der Distraktoren mit einfacher und mit besonderer inhaltlicher Ähnlichkeit sowie mit Gebärdenähnlichkeit gemein ist. Bei den Testpersonen des Clusters 2 handelt es sich um fünf (12,8%) der Dritt-, drei (9,7%) der Viert- und drei (8,3%) der Fünftklässler, um zwei (5,9%) der 34 oral orientiert und neun (13,6%) der 66 mit Gebärden unterrichteten Testpersonen. Von den sechs (11,8% aller) Mädchen und fünf (9,1% aller) Jungen haben alle hörende Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;1 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 101. Es handelt sich um Testpersonen aus allen Städten bis auf Hamburg und Würzburg; einen (7,7%) der Bremer, drei (21,4%) der Berliner, zwei (16,7%) der Schleswiger, zwei (11,1%) der Münsteraner, zwei (16,7%) der Züricher und einen (8,3%) der Kölner Schüler.

Das Cluster 3 umfasst die Ergebnisse von 29 Testpersonen, deren Distraktorenwahl in allen Fällen am niedrigsten ausfällt. Unter diesen Testpersonen sind vier (10,3%) der Dritt-, 14 (45,2%) der Viert- und elf (30,6%) der Fünftklässler, 14 (41,2%) der 34 oral orientiert, elf

(16,7%) der 66 mit Gebärden und vier (66,7%) der sechs bilingual unterrichteten Testpersonen. Es sind 13 (25,5% aller) Mädchen und 16 (29,1% aller) Jungen, darunter sechs (60,0%) der Testpersonen mit gehörlosen Eltern. Die Testpersonen sind durchschnittlich 10;6 Jahre alt und erreichten einen durchschnittlichen IQ-Wert von 107. Es handelt sich um Testpersonen aus allen Städten bis auf Schleswig; vier (30,8%) der Hamburger, einen (7,7%) der Bremer, fünf (35,7%) der Berliner, fünf (27,8%) der Münsteraner, drei (25,0%) der Züricher, zwei (16,7%) der Kölner und neun (75,0%) der Würzburger Schüler.

Im Cluster 4 finden sich die Ergebnisse von 15 Testpersonen zusammen, deren Profil erhöhte Wahlhäufigkeit der inhaltsähnlichen Distraktoren zeigt. Die Testpersonen sind durchschnittlich 10;7 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 92. Es sind acht (20,5%) der Dritt-, einer (3,2%) der Viert- und sechs (16,7%) der Fünftklässler. Unter ihnen sind vier (7,8% aller) Mädchen und elf (20,0% aller) Jungen, vier (11,8%) der oral orientiert, neun (13,6%) der mit Gebärden und zwei (33,3%) der bilingual unterrichteten Testpersonen. Alle haben hörende Eltern. Insgesamt handelt es sich um drei (23,1%) der Hamburger, zwei (15,4%) der Bremer, drei (21,4%) der Berliner, fünf (27,8%) der Münsteraner, einen (8,3%) der Kölner und einen (8,3%) der Würzburger Schüler. Testpersonen aus Schleswig und Zürich sind nicht Mitglieder dieses Clusters.

Cluster 5 umfasst die Ergebnisse von 18 Testpersonen, die besonders häufig die mundbild- sowie die inhaltlich ähnlichen Distraktoren wählten. Die Gruppe setzt sich zusammen aus zehn (25,6%) der Dritt-, zwei (6,5%) der Viert- und sechs (16,7%) der Fünftklässler. Unter ihnen sind zehn (19,6% aller) Mädchen und acht (14,5% aller) Jungen, fünf (14,7%) der oral orientiert und 13 (19,7%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Die Testpersonen dieses Clusters sind durchschnittlich 10;5 Jahre alt und erreichen einen IQ-Wert von durchschnittlich 95 Punkten. Es sind Testpersonen aus allen Städten bis auf Berlin und Würzburg; vier (30,8%) der Hamburger, vier (30,8%) der Bremer, fünf (41,7%) der Schleswiger, einer (5,6%) der Münsteraner, zwei (16,7%) der Züricher und zwei (16,7%) der Kölner Schüler.

Cluster 6 umfasst die Ergebnisse von drei Testpersonen, deren Profil im Vergleich zu denen der Mitglieder anderer Cluster die durchgehend höchste Wahlhäufigkeit aller Distraktoren zeigt. Die Testpersonen dieses Clusters sind durchschnittlich bereits 11;0 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 86 Punkten. Es handelt sich um zwei (5,1%) der Dritt- und einen (3,2%) der Viertklässler, alle (4,5%) aus der Gruppe der mit Gebärden unterrichteten Testpersonen. Darunter sind ein (2,0% aller) Mädchen und zwei (3,6% aller) Jungen, alle mit hörenden Eltern. Es sind einer (7,7%) der Bremer, einer (7,1%) der Berliner und einer (5,6%) der Münsteraner Schüler. Testpersonen aus Hamburg, Schleswig, Zürich, Köln und Würzburg passen mit ihren Testergebnissen nicht in dieses Cluster.

In Cluster 7 befinden sich die Ergebnisse von zehn Testpersonen, die bei sonst geringer Wahlhäufigkeit verstärkt die Distraktoren mit Ähnlichkeit im Mundbild wählten. Die Testpersonen sind durchschnittlich 10;4 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 96. Es sind vier (10,3%) der Dritt-, vier (12,9%) der Viert- und zwei (5,6%) der Fünftklässler. Unter ihnen sind sieben (13,7% aller) Mädchen und drei (5,5% aller) Jungen, zwei (5,9%) der oral orientiert und acht (12,1%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Es handelt sich um zwei (15,4%) der Hamburger, einen (7,7%) der Bremer, zwei (16,7%) der Schleswiger, einen (16,7%) der Züricher und vier (33,3%) der Kölner Schüler. Testpersonen aus Berlin, Münster und Würzburg sind nicht Bestandteil dieses Clusters.

Die Analyseergebnisse des Subtests Lautsprache decken sich mit denen der beiden anderen Subtests. So bestätigen die Werte zunächst die ausgeprägten lexikalisch-semantischen Kenntnisse der bilingualen Schüler. Der überwiegende Anteil der bilingualen Gruppe findet sich in den drei Clustern wieder, die die Profile derjenigen Testpersonen zusammenfassen, die jeweils die niedrigste Anzahl von Distraktoren gewählt haben. Dies sind Cluster 2 bei den Nomen (50% bilinguale), Cluster 3 bei den Verben (83,3% bilinguale) und Cluster 3 bei den Adjektiven (66,7% bilinguale).

Zudem weist die Analyse der Cluster, in denen die Profile der Testpersonen zusammengefasst sind, die überwiegend die subtestspezifischen mundbildähnlichen Distraktoren gewählt haben, ebenso unspezifische Ergebnisse auf, wie die entsprechenden in den anderen beiden Subtests. So setzt sich Cluster 7 bei den Nomen zusammen aus 14,7% der oral erzogenen, 18,2% der mit Gebärden und 16,7% der bilingual unterrichteten Testpersonen. Demnach zeigt sich keine überwiegende Mundbildorientierung bei den oral erzogenen Testpersonen, sondern sie bilden im Gegenteil sogar den geringsten Anteil. In Cluster 1 und 4 bei den Verben sowie in Cluster 5 und 7 bei den Adjektiven ergeben sich gleichermaßen unspezifische Werte. So kommen in keiner dieser Zusammenstellungen bilingual unterrichtete Testpersonen vor; die beiden anderen Gruppen sind in etwa gleichermaßen vertreten.

1.8.5 Überprüfung V

Schließlich wurde auch für den Gesamttest eine Clusteranalyse durchgeführt, bei der es zu folgenden Ergebnissen kam:

Tab. 94-C 1.8.5: Clusteranalyse bzgl. der Distraktorenwahl Gesamttest						
Cluster SSA	Anzahl	HRP	G	M	I	MOD (S)
1	17	15,2	26,5	21,5	33,6	22,9
2	22	5,9	9,3	6,5	11,0	8,5
3	2	22,5	57,5	47,5	62,5	57,0
4	22	16,9	34,3	31,9	42,5	33,4
5	1	29,0	45,0	48,0	77,0	46,0
6	18	19,9	40,2	41,6	53,9	42,4
7	19	11,4	19,3	15,9	24,5	18,3

Cluster 1 umfasst die Ergebnisse von 17 Testpersonen, deren Profil eine ausgeglichene Verteilung der Wahlhäufigkeiten aller Distraktoren darstellt. Die Testpersonen dieses Clusters sind durchschnittlich 10;5 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 101 Punkten. Es handelt sich um fünf (13,5%) der Drittklässler, fünf (17,2%) der Viert- und sieben (20,0%) der Fünftklässler, um sechs (19,4%) der 31 oral orientiert und um elf (17,2%) der 64 mit Gebärden unterrichteten Testpersonen. Darunter sind neun (18,8% aller) Mädchen und acht (15,1% aller) Jungen, alle mit hörenden Eltern. Es handelt sich um Testpersonen aus allen Städten bis auf Hamburg; drei (25,0%) der Bremer, zwei (15,4%) der Berliner, zwei (18,2%) der Schleswiger, drei (17,6%) der Münsteraner, drei (25,0%) der Züricher, zwei (18,2%) der Kölner und zwei (18,2%) der Würzburger Schüler.

Cluster 2 beschreibt die Zusammenfassung von 22 Profilen, innerhalb derer alle Distraktorenarten im Vergleich zu den übrigen Clustern am seltensten gewählt wurden. Bei den Testpersonen des Clusters 2 handelt es sich um einen (2,7%) der Dritt-, elf (37,9%) der Viert- und zehn (28,6%) der Fünftklässler, um acht (25,8%) der 31 oral orientiert und zehn (15,6%) der 64 mit Gebärden und vier (66,6%) der sechs bilingual unterrichteten Testpersonen. Unter den neun (18,8% aller) Mädchen und 13 (24,5% aller) Jungen befinden sich sechs (60,0%) der Testpersonen mit gehörlosen Eltern. Das durchschnittliche Alter der betreffenden Testpersonen beträgt 10;8 Jahre und sie erreichten im Raven einen durchschnittlichen IQ-Wert von 107. Es handelt sich um Testpersonen aus allen Städten bis auf Schleswig; vier (30,8%) der Hamburger, einen (8,3%) der Bremer, drei (23,1%) der Berliner, vier (23,5%) der Münsteraner, zwei (16,7%) der Züricher, drei (27,3%) der Kölner und fünf (45,5%) der Würzburger Schüler.

Das Cluster 3 umfasst die Ergebnisse von zwei Testpersonen, deren Profil nach dem der Testperson aus Cluster 5 für alle Distraktorenarten die höchste Wahlhäufigkeit anzeigt. Beide Testpersonen sind Drittklässler, mit Gebärden unterrichtet. Es handelt sich um einen Jungen und ein Mädchen, beide mit hörenden Eltern. Sie sind durchschnittlich erst 9;8 Jahre

alt und erreichten einen durchschnittlichen IQ-Wert von 97. Es handelt sich um Testpersonen aus Berlin und Köln.

Im Cluster 4 finden sich die Ergebnisse von 22 Testpersonen zusammen, deren Profil ähnlich zu dem der Testpersonen aus Cluster 1, aber mit etwas erhöhter Wahlhäufigkeit aller Distraktoren verläuft. Die Testpersonen sind durchschnittlich 10;7 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 94. Es sind neun (24,3%) der Dritt-, sieben (24,1%) der Viert- und sechs (17,1%) der Fünftklässler. Unter ihnen sind elf (22,9% aller) Mädchen und elf (20,8% aller) Jungen, sieben (22,6%) der oral orientiert und 15 (23,4%) der mit Gebärden unterrichteten Testpersonen. Alle haben hörende Eltern. Insgesamt handelt es sich um Testpersonen aus allen betreffenden Städten bis auf Würzburg; zwei (15,4%) der Hamburger, zwei (16,7%) der Bremer, drei (23,1%) der Berliner, fünf (45,5%) der Schleswiger, vier (23,5%) der Münsteraner, drei (25,0%) der Züricher und drei (27,3%) der Kölner Schüler.

Cluster 5 umfasst die Ergebnisse von lediglich einer Testperson. Das Profil dieser Testperson zeigt im Bezug zu allen anderen Profilen für alle Distraktorenarten die eindeutig höchste Wahlhäufigkeit. Damit unterscheidet sich ihr Distraktorenprofil so deutlich von denen der übrigen Testpersonen, dass es keine Möglichkeit der Zusammenfassung zu einem Cluster gab.

Cluster 6 umfasst die Ergebnisse von 18 Testpersonen, die eine durchgehend hohe Wahlhäufigkeit aller Distraktoren zeigen. Die Testpersonen dieses Clusters sind durchschnittlich 10;2 Jahre alt und erzielten einen IQ-Wert von durchschnittlich 95 Punkten. Es handelt sich um elf (29,7%) der Drittklässler, drei (10,3%) der Viert- und vier (11,4%) der Fünftklässler, vier (12,9%) der oral orientiert und 14 (21,9%) der mit Gebärden unterrichteten Testpersonen. Darunter sind sechs (12,5% aller) Mädchen und zwölf (22,6% aller) Jungen. Alle haben hörende Eltern. Insgesamt handelt es sich um Testpersonen aus allen betreffenden Städten außer Würzburg; fünf (38,5%) der Hamburger, drei (25,0%) der Bremer, zwei (15,4%) der Berliner, vier (36,4%) der Schleswiger, einen (5,9%) der Münsteraner, einen (8,3%) der Züricher und zwei (18,2%) der Kölner Schüler.

In Cluster 7 befinden sich die Ergebnisse von 19 Testpersonen, die durchgehend wenig Distraktoren wählten. Die Testpersonen sind durchschnittlich 10;4 Jahre alt und erreichen einen durchschnittlichen IQ-Wert von 97. Es sind acht (21,6%) der Dritt-, drei (10,3%) der Viert- und acht (22,9%) der Fünftklässler. Unter ihnen sind elf (22,9% aller) Mädchen und acht (15,1% aller) Jungen, sechs (19,4%) der oral orientiert, elf (17,2%) der mit Gebärden und zwei (33,3%) der bilingual unterrichteten Testpersonen. Es befinden sich zwei (20,0%) der Testpersonen mit gehörlosen Eltern in diesem Cluster. Insgesamt sind es Schüler aus allen die Untersuchung betreffenden Städten außer Schleswig; zwei (15,4%) der Hamburger, zwei

(16,7%) der Bremer, zwei (15,4%) der Berliner, fünf (29,4%) der Münsteraner, drei (25,0%) der Züricher, einen (9,1%) der Kölner und vier (36,4%) der Würzburger Schüler.

Die Untersuchung der Profilverläufe für den Gesamttest ebnet Subtest- bzw. Subtestteil-spezifische Ausschläge grundsätzlich ein. Dadurch sind deutlich veränderte Ergebnisse denkbar. Die Zusammensetzung der Cluster für den Gesamttest zeichnet jedoch ein Bild, das mit denen aus den Analysen der Subtests hervorgegangenen übereinstimmt. So befinden sich auch im Gesamttest im Cluster 2, das die Profile zusammenfasst, innerhalb derer alle Distraktoarten im Vergleich zu den übrigen Clustern am seltensten gewählt wurden, mit 66,6% der überwiegende Teil der bilingual erzogenen Schüler. Die übrigen Cluster beinhalten gleichermaßen unspezifische Ergebnisse wie oben beschrieben.

2. Pädagogische Folgerungen

Die Ergebnisse der Pilotstudie belegen, dass die bilingual erzogenen gehörlosen Testpersonen in der Gebärdensprache im Vergleich zu den anderen beiden Modalitäten sehr gut abschneiden. Die Ergebnisse der Hauptuntersuchung (auf die im Folgenden Bezug genommen wird) weisen in die gleiche Richtung. So haben hier fast drei Viertel der Testpersonen in der Gebärdensprache am stärksten abgeschnitten. Die Annahme, dass die DGS am vollständigsten erlernt wird, da sie gehörlosen Kindern am leichtesten zugänglich ist, bestätigt sich damit.

In der Gruppe derjenigen Testpersonen, die ihr bestes Ergebnis im Subtest Gebärdensprache erzielten, erhält dieses Ergebnis besondere Beweiskraft. Die leichte Zugänglichkeit der Gebärdensprache bestätigt sich, da offensichtlich selbst diejenigen Testpersonen, denen die DGS pädagogisch-konzeptionell nicht angeboten wird, einen Zugang zu ihr gefunden haben. So haben über 80% der bilingual und der gebärdensorientiert erzogenen Schüler sowie deutlich über 50% der primär hörgerichtet erzogenen Testpersonen die besten Ergebnisse in der DGS erzielt und damit bereits gute Kenntnisse in diesem einen sprachlich-modalen Feld unter Beweis stellen können. Neben den Testpersonen, die konzeptionell bereits mit der Gebärdensprache vertraut sind (bilingual, gebärdensorientiert) haben demnach auch mehr als die Hälfte der primär hörgerichtet erzogenen Testpersonen einen guten Zugang zur Gebärdensprache gezeigt.

In der Gruppe, die in der Lautsprache am stärksten abgeschnitten hat, fällt auf, dass kein bilingual erzogener Schüler darunter ist. Dies entspricht dem bilingualen Konzept, welches davon ausgeht, dass sich die Lautsprachkompetenzen auf dem Hintergrund der erworbenen Gebärdensprachkompetenzen positiv entwickeln.

In Teil A war gezeigt worden, dass es aus Gründen der psycho-sozialen Entwicklung einer Persönlichkeit notwendig ist, mindestens ein vollständiges Kommunikationsmittel zur Verfügung zu haben. Die Ergebnisse zeigen nun, dass Gehörlose die Gebärdensprache in hohem Grad beherrschen. Demnach lässt sich die somit fundierte Behauptung aufstellen, dass sich die Gebärdensprache als Lernangebot für gehörlose Grundschüler eignet, sofern man dieses Ziel verfolgt, den betreffenden Kindern ein vollständiges Kommunikationsmittel zur Verfügung zu stellen und damit eine gesunde psycho-soziale Entwicklung zu ermöglichen. Die Ergebnisse ermutigen zum Ausbau dieser Fähigkeiten.

Die Abstufung der Wortarten im Schwierigkeitsgrad ergibt sich annahmegemäß. Im Schriftsprachtest kamen insgesamt etwas schlechtere, zum Teil deutlich schlechtere Ergebnisse als im Subtest Gebärdensprache zustande. Die größten Schwierigkeiten bereitete erwartungsgemäß der lautsprachliche Subtest. Neben den ausgeprägten Fähigkeitsbereichen kann so für jede einzelne Testperson abgelesen werden, in welchen sprachlich-modalen Bereichen bzw. in welchen Wortarten Förderbedarf besteht. So können die entsprechenden Maßnahmen gezielt eingesetzt werden.

Der Vergleich mit hörenden Testpersonen in modifizierter Fehlerbewertung rückt den Wortschatz Gehörloser in ein neues Licht. Wo die Fehlerangaben insgesamt noch den Eindruck erwecken, dass die Gehörlosen deutlich schlechtere Leistungen erbrächten, erhellt die Bewertung der durchgängig gemachten Fehler ihre tatsächlichen Leistungsniveaus. Die üblicherweise vertretene Ansicht, gehörlose Kinder seien sprachlich nicht altersgemäß entwickelt und hielten einem Vergleich mit gleichaltrigen hörenden Kindern daher nicht stand, hält ihrerseits den Ergebnissen nicht stand. Damit wäre für die erläuterte Theorie belegt, dass das Lexikon der Kinder größer ist, als man unter Einbeziehung ausschließlich lautsprachlicher Abfrage feststellen könnte. Die Überlegung, dass modalitätenspezifische Gründe für insgesamt schlechte Testergebnisse Gehörloser verantwortlich sein könnten, bestätigt sich. Es wird deutlich, dass man den Testpersonen erst gerecht werden kann, wenn man – anders als in gängig verwendeten Testverfahren üblich – ein gebärdensprachliches Angebot macht. Dieses Ergebnis bestätigt gleichzeitig erneut die Annahme, dass es sich bei der Gebärdensprache um die den Gehörlosen am vollständigsten zur Verfügung stehende Sprache handelt und unterstreicht die Bedeutung ihrer Verwendung in der Erforschung des (passiven) Wortschatzes Gehörloser sowie in gehörlosenpädagogischen Sprachvermittlungsverfahren. Grundsätzlich bestätigen die Ergebnisse die vorausgegangene Annahme, dass weiterreichende Fähigkeiten vorhanden sind, als bisher diagnostizierbar war.

Die gesonderte Untersuchung der Gruppe bilingual erzogener Schüler belegt in beeindruckender Weise, welchen hohen Kenntnisstand diese Schüler nicht bilingual erzogenen Vergleichspersonen gegenüber erreicht haben. So sind die bilingualen Testpersonen in allen

Subtests durch sämtliche Wortarten uneingeschränkt am besten, und zwar selbst ohne angepasste Fehlerbewertung. Dies ist für sich genommen ein Ergebnis, das die Arbeit an der Entwicklung des bilingualen Konzepts in außerordentlicher Weise bestätigt. Zudem beugt es dem Vorwurf vor, die modifizierte Fehlerbewertung rücke die Ergebnisse in ein unangemessenes Licht.

Auch durch die signifikanten Werte der Korrelationsüberprüfung zwischen den Erziehungskonzepten und den Ergebnissen im Test wird bestätigt, dass sich der passive Wortschatz bei bilingual erzogenen Kindern in allen Modalitäten signifikant besser entwickelt als bei oral- oder gebärdenorientierter Erziehung. Dabei sind signifikante Leistungsunterschiede zwischen Testpersonen der beiden letztgenannten Konzepte eindeutig in der Modalität Lautsprache, teils in der Modalität Schriftsprache und nicht in der Modalität Gebärdensprache erkennbar.

Signifikante Zusammenhänge bestehen zudem zwischen der passiven Wortschatzentwicklung und dem Intelligenzniveau der Testpersonen, und dies in allen Fällen mit Ausnahme von Subtestteil LSV. So stellt sich heraus, dass sich der passive Wortschatz bei höherem Intelligenzniveau signifikant besser entwickelt. Die Werte fallen im Sinne der Annahme korrelativer Zusammenhänge sowie der Konstruktbeschreibung aus.

Es bestehen außerdem lineare Zusammenhänge zwischen der passiven Wortschatzentwicklung und dem Hörstatus der Eltern. Diese Zusammenhänge bzgl. des Hörstatus der Eltern fallen ebenfalls im Sinne der Annahme sowie der Konstruktbeschreibung aus. Die Ergebnisse belegen, dass sich der passive Wortschatz in der Grundschule bei gehörlosen Kindern gehörloser Eltern in allen Modalitäten signifikant besser entwickelt als bei gehörlosen Kindern hörender Eltern.

Leicht abweichend von den Ergebnissen der Analyse der Gesamtstichprobe stellen sich die Betrachtungen der (größten) Gruppe dar, die in der Gebärdensprache die besten Werte erreichte. So wird in 75 von 101 Fällen in der Gebärdensprache die eindeutig beste Leistung erzielt. Das häufigste der Profile zeigt die zweitbeste Leistung in der Lautsprache, das zweithäufigste in der Schriftsprache und nicht, wie in der Gesamtstichprobe, vorwiegend umgekehrt. Dieses Ergebnis erhält dadurch Gewicht, dass der Profilverlauf Gebärdensprache-Lautsprache-Schriftsprache 43-mal vorkommt, während sich die verbleibenden 58 auf 12 weitere Profile verteilen. Das Ergebnis deutet darauf hin, dass Schüler mit guten Gebärdensprachkenntnissen gleichzeitig bzw. auf diesen aufbauend relativ gute Lautsprachkenntnisse entwickeln konnten. Dies spricht zunächst dafür, dass Gebärdensprachkenntnisse eine gute Basis darstellen zum Erlernen schwierigerer sprachlich-modaler Bereiche. Zudem be-

stätigt es die in Teil A erläuterte Situation, dass die Schriftsprachkenntnisse stärkere curriculare Berücksichtigung erfahren sollten.

Innerhalb der Modalität Gebärdensprache zeigen die Profilverläufe für die Verteilung der Wortartenkenntnisse an, dass die meisten Testpersonen (63,6%) in der Wortart Verben am stärksten abschneiden. Wie in Teil A gezeigt wurde, zählen die Verben zu den Wortarten, deren Kenntnis auf elaborierte sprachliche Fähigkeiten schließen lässt. Insofern beweist dieses Ergebnis, dass die gehörlosen Testpersonen über qualitativ hochwertige und vielfältige Fähigkeiten verfügen. Die Annahme, ihre Fähigkeiten beschränkten sich auf den vergleichsweise leicht zu erlernenden nominalen Wortschatz, kann damit zurückgewiesen werden.

Bei der Überprüfung der Profilverläufe innerhalb der drei Wortarten stellte sich heraus, dass in 50,5% aller Fälle die eindeutig beste Leistung in der Wortart Nomen erzielt wird. Dabei ergeben sich 3 Profilverläufe. Davon entsteht in 68,6% der Fälle eine Leistungsabstufung von den Nomen bis zu den Adjektiven. Dieses Profil entspricht der im Vorwege formulierten Annahme.

Die Profilverläufe der Testpersonen, die mit unterschiedlichen Konzepten erzogen werden, wurden daraufhin überprüft, ob deutliche Unterschiede voneinander erkennbar sind. Die Ergebnisse zeigen (mit Ausnahme einer Überschneidung) deutliche Unterscheidungen der Leistungen. Dies deckt sich mit den bereits mehrfach angeführten Belegen dafür, dass das Leistungsniveau der Testpersonen sich gemäß der jeweiligen pädagogischen Konzepte, mit denen sie sich konfrontiert sehen, voneinander unterscheidet.

Weiterhin bestätigt die Untersuchung anhand deutlicher Differenzen zwischen den Profilverläufen der Drittklässler denen der Viert- (und Fünft-)klässler gegenüber das angenommene Leistungsgefälle.

Die Untersuchung von Profilverläufen in der Distraktorenwahl in den drei Subtests sowie im Gesamttest verdeutlicht bestimmte Tendenzen. Diese zeichnen sich ab durch die Struktur der Zusammenstellung einzelner Cluster. Wie in der jeweiligen Zusammenschau gezeigt worden war, bestätigen mehrfach vorkommende Strukturen die ausgeprägten lexikalisch-semantischen Kompetenzen der Hamburger bilingualen Schüler. Sie befinden sich durchgehend in der Gruppe derjenigen Testpersonen mit den besten Ergebnissen.

Die Ergebnisse der Überprüfung der Distraktorenwahl wurden zudem unter der Fragestellung ausgewertet, ob sich bestimmte Schlüsse bezüglich der Wahlstrategien ziehen lassen. Die überwiegende Anzahl der Cluster fasst Ergebnisse zusammen, die keine Häufungen in den einzelnen untersuchten Aspekten (Alter, IQ, Klasse, Konzept, Schule) aufweisen und die

daher keine spezifischen Schlüsse auf Wahlstrategien zulassen. Dies trifft vornehmlich auf Zusammenstellungen von Ergebnissen zu, die Wahlhäufungen einer der drei Distraktorenarten abbilden und die eine relativ große Anzahl von Profilen zusammenfassen, von daher also prinzipiell aussagefähig wären. Die Annahme, die qualitativ abgestufte Konzeption von Testaufgaben ließe anhand der Auswertung solcher korrelativer Zusammenhänge Rückschlüsse auf spezifische Zugänge der Testpersonen zum Wortschatz zu, bestätigt sich daher nicht. Der Bereich qualitativer Auswertung wird unter C.4 weiterführend behandelt.

3. *Resümee klassischer testtheoretischer Aspekte*

Im theoretisch grundlegenden Teil A dieser Arbeit war das Vorhaben und der Anspruch formuliert worden, ein diagnostisches Instrument zur Erfassung des Niveaus lexikalisch-semantischer Fähigkeiten im Bereich der referenzsemantischen Wortarten zu entwickeln, das in den drei sprachlich-modalen Bereichen Gebärden-, Schrift- und Lautsprache einsetzbar sein sollte. Dieses Testinstrument wurde mit dem Vorsatz entwickelt, neben quantitativen auch qualitative Bewertungsmaßstäbe konzeptionell einzuarbeiten. Zum Erreichen dieser beiden Zielsetzungen wurde der Test zunächst nach den Regeln der klassischen Testtheorie entwickelt. Gleichzeitig wurden qualitative Auswertungsansätze konzeptionell eingearbeitet und schließlich der Versuch unternommen, diese endgültig in das Konzept des Tests zu integrieren.

Resumierend ist festzustellen, dass die Entwicklung des Testinstruments gemäß der sog. klassischen Testtheorie sowie die entsprechenden Auswertungsmöglichkeiten der Testergebnisse nach quantitativen Maßstäben realisiert wurden. Das Erreichen dieses Ziels kann als erwartungsgemäß bezeichnet werden, da hier auf bestehende theoretische Grundlagen zurückgegriffen werden konnte. So wurde, wie in Teil B 3 umfassend dargestellt, ein bestimmter Anforderungskatalog (nach Lienert, Bortz u. a.) angewendet. Demnach sind für das vorliegende Testinstrument die drei zentralen Testgütekriterien Objektivität, Reliabilität sowie Validität gewährleistet. Der bereits erfolgten Darstellung und jeweils abschließenden Bewertung der Einzelheiten dieser Anwendung ist hier nichts hinzuzufügen.

Die konzeptionellen Einzelheiten des Instruments sind unter B.1 detailliert dargestellt worden. Diese wurden umgesetzt wie dargestellt und beibehalten. Insofern sind die differentialdiagnostischen Elemente bzw. Hintergründe für die Konzeption sowie die gehörlosenspezifische Ausrichtung des Instruments gewährleistet. In Hinblick auf die Überprüfungen pädagogischer Fragestellungen sollen im Folgenden einzelne spezifische Aspekte des vorliegenden Instruments unter testtheoretischem Blickwinkel nochmals aufgenommen und beleuchtet werden. So war die Abfolge der Modalitäten im Anwendungsverlauf (1. GS, 2. SS, 3. LS)

auf dem Hintergrund bestimmter Annahmen festgelegt worden. Die Gebärdensprache sollte als leichteste Modalität als motivierender Auftakt fungieren. Die Ergebnisse bestätigen nun, dass die Gebärdensprache den Testpersonen am leichtesten fiel, und rechtfertigt daher diese Position im Verlauf. Die Schriftsprache sollte aufgrund ihrer Andersartigkeit als zweites eingesetzt werden. Dies rechtfertigt sich von selbst, da ein Erinnerungseffekt durch die Übereinstimmung des Testheftes I für Subtest GS und LS vermieden werden sollte. Schließlich sollte der Lautsprachtest am Schluss eingesetzt werden, da er im Vorfeld als schwerste Modalität eingestuft wurde und daher motivationsmindernd wirken könnte. Auch diese Annahme bestätigt sich anhand der Ergebnisse. Die Abfolge der Modalitäten wird daher als stimmig und geeignet bewertet und beibehalten. In gleicher Weise stimmt die Abstufung der Wortarten im Schwierigkeitsgrad mit den darüber im Vorfeld getroffenen Annahmen überein. So soll auch diese aus Gründen der Motivationskurve beibehalten werden.

Die Bestätigung korrelativer Zusammenhänge zwischen Erziehungskonzepten und den Ergebnissen im Test lässt nun nicht allein die Aussage einer Abstufung unter den unterschiedlich unterrichteten Testpersonen zu, sondern bestätigt in Hinblick auf das Instrument, dass der Test unterschiedlich ausgeprägte Kenntnisstände (gemäß der jeweiligen Erziehungskonzepte) abbildet. Insofern erfüllt das Instrument den Anspruch, der an es gestellt wird. Gleiches gilt für den Umstand, dass Korrelationen zwischen Intelligenzquotient und Testergebnis bestehen. Auch damit bestätigt sich die Qualität bzw. Aussagekraft des Instruments.

4. Resümee der qualitativen Auswertung

Die qualitative Auswertung nimmt im Rahmen vorliegender Arbeit eine Sonderstellung ein. Dies beruht massgeblich auf dem Umstand, dass die wissenschaftlich-theoretischen Hintergründe zu diesem Themenbereich keine vorgefertigten Anwenderkataloge bereitstellen, sondern sich das vorliegende Projekt in eine Anzahl von Pilotprojekten der Entwicklung qualitativer Tests einreicht. Daher war nicht zu erwarten, dass im Verlauf des Projekts gesicherte und praktisch anwendbare Richtlinien zu entwickeln gewesen wären. So konnten die Ziele teilweise oder annähernd erreicht werden. Die Einzelheiten werden im Folgenden unter der Fragestellung beleuchtet, was im Einzelnen umgesetzt werden konnte, welche Versuche sich als ungeeignet im Sinne der Entwicklung eines qualitativen Verfahrens erwiesen haben und welchen Informationswert diese Ergebnisse haben.

Die Betrachtung qualitativer Aspekte umfasst vier Anteile. Dies sind zunächst die testkonstruktiven Elemente, die den Test als qualitatives Instrument ausweisen. Weiterhin ist die qualitative Abstufung der Distraktoren als Kernstück dieses Bereichs zu nennen. Der dritte

Anteil ist die unter C. 1.2 und 1.4 dargestellte modifizierte Fehlerbewertung. Schließlich ist die Untersuchung typischer Profilverläufe bezüglich der Distraktorenauswahl zu nennen. Hier lag die Fragestellung zugrunde, ob korrelative Zusammenhänge zwischen den Strategien der Distraktorenwahl und bestimmten Begleitmerkmalen der Testpersonen bestehen.

Grundlegendes testkonstruktives Element unter qualitativen Gesichtspunkten ist zunächst die Auswahl und gleichgewichtete Einbeziehung der drei Sprachmodalitäten Laut-, Schrift- und Gebärdensprache. Hiermit und mit der Berücksichtigung gehörlosenspezifischer Bedingungen der Sprachrezeption wird die spezifische Eignung des Instruments für Gehörlose erreicht. Dieser Umstand, bereits in der Testkonstruktion die spezifischen Bedingungsgefüge der Testpersonen zu berücksichtigen, für die das Instrument entwickelt wird, zeichnet den Test qualitativ aus.

Gleiches gilt für die Bezugnahme auf die dargestellten Wortlisten bei der Auswahl der lexikalischen Prüfitems. Diese Bezugnahme auf differenzierte Angaben enthaltende Korpora bedeutet einen entscheidenden qualitativen Fortschritt gegenüber der üblichen Praxis des überwiegenden Anteils auf dem Markt befindlicher Tests. Wie dargestellt wurde, rekrutieren die meisten Testkonstrukteure ihre Items ohne Zugrundelegung differenzieller Angaben über Methodik und theoretische Begründungszusammenhänge. Gleiches gilt für die Auswahl und gleichgewichtete Einbeziehung der Wortarten und ihrer anteiligen Aufnahme in den Itemkorpus. Beides wurde auf dem Hintergrund entsprechender inhaltlicher Betrachtungen vorgenommen. Insofern erfüllt das vorliegende Instrument in Hinblick auf bestimmte konstruktive Elemente die Anforderungen eines qualitativen Tests.

Wesentliches Element der Konzeption vorliegenden Tests ist die qualitative Abstufung der Distraktoren. Diese stellt den einzigen Fall einer geringfügigen Abweichung von den vorgegebenen Richtlinien der klassischen Konstruktionstheorie dar. Das Vorgehen bei der Distraktorenanalyse wurde unter B 3.4.5 dargestellt. Die Darstellung weist das Verfahren als in sich schlüssig und testtheoretisch korrekt aus. Daher wird hier nichts hinzugefügt. Unter B. 3.10.1 wurde die Annahme, dass die Wahl einiger plausibler Ablenker auf ein höheres Wortschatzniveau hindeutet als die Wahl anderer Ablenker, die in dieser Modalität keinen bedeutsamen Bezug zum Item aufweisen, dass also die Konzeption der qualitativen Abstufung der Distraktoren eine qualitative Zusatzbewertung der Fehler ermöglicht, mit einer gesonderten Untersuchung überprüft und im Anschluss daran bereits ausgewertet. Wie dabei gezeigt worden war, ergab sich eine deutlich bestätigende Tendenz für die höhere Plausibilität einiger Ablenker gegenüber anderen. Dieses Ergebnis hat jedoch nach dem durch diese Arbeit vorgelegten Forschungsstand eingeschränkte Gültigkeit: So bestätigten annähernd alle Untersuchungen die Annahme, es ist jedoch eine Ausnahme im Schriftsprachtest zu verzeichnen. Zudem mussten zur Untersuchung unbestätigte Hilfsannahmen getroffen werden. Die

Annahme gilt jedoch unter Berücksichtigung dieser Einschränkung (vergl. C. 5 Ausblick) als bewiesen und die Konzeption wird für das vorgelegte Instrument übernommen. Das Testinstrument ermöglicht demnach durch die Ermittlung eines individuellen Distraktorenprofils ergänzend zum quantitativen Testergebnis einer Testperson differenziertere Aussagen zum Niveau ihrer Sprachkompetenz. So werden durch die Anwendung des Tests mit dem quantitativen Wert der eine und mit der qualitativen Auswertung vier bzw. fünf weitere Werte für eine Testperson in jedem Subtestteil ermittelt, die jeweils als Prozentränge in entsprechenden Normierungstabellen festgestellt werden können.

Die dargestellte modifizierte Fehlerbewertung, die das tatsächliche Leistungsniveau gehörloser Testpersonen sichtbar machte, wurde in pädagogischer Hinsicht bereits ausgewertet und bewertet. Zentrale Motivation für die Durchführung dieser Fehlerbewertungsvariante war das Vorhaben, die Testergebnisse differenzialdiagnostisch, d. h. unter konsequenter Berücksichtigung testpersonenspezifischer Merkmale anzuschauen. So fällt diese Untersuchung zentral in den Bereich qualitativer Auswertung der Testergebnisse. Da die Auswertung bereits erfolgte, soll dem an dieser Stelle nichts hinzugefügt, wohl aber bemerkt werden, dass sich die modifizierte Fehlerbewertung in die Elemente qualitativer Auswertung einreicht.

Die Untersuchung typischer Profilverläufe bezüglich der Distraktorenauswahl schließlich ergab keine signifikanten korrelativen Zusammenhänge zwischen der Distraktorenwahlstrategie und bestimmten Merkmalen der Testperson, wie z. B. dem Erziehungskonzept. So war in den jeweiligen Zusammenfassungen der Ergebnisse unter C. 1.8 gezeigt worden, dass die Ergebnisse unspezifisch bleiben und sich keine Häufungen oder sonstige Besonderheiten innerhalb der Cluster abzeichnen, die auf schlüssige Wahlstrategien im Sinne der Distraktorenkonzeption hindeuten. Die Daten im Gebärdensprachtest weisen bspw. darauf hin, dass gebärdenorientierte Erziehung keine gebärdenorientierte Wahlstrategie mit sich bringt, die sich auf das Erziehungskonzept zurückführen ließe, die Daten im Schriftsprachtest, dass sich keine überwiegende Mundbildorientierung bei den oral erzogenen Testpersonen zeigt. Direkte Bezüge zwischen Wahlstrategien im Test und sprachlichen Sozialisationsbedingungen der Testpersonen können demnach im Sinne einer nachvollziehbaren Beweisführung nicht festgestellt werden. Dieses Ergebnis wertet die Konzeption qualitativer Abstufung der Distraktoren ab. Dieser Umstand sowie die dargestellten Einschränkungen bei der Untersuchung des Distraktorenprofils führen zu der Überlegung, welche Anschlussprojekte sich daraus ableiten.

5. Ausblick

Das zentrale Ziel dieser Arbeit, die Bereitstellung eines Testinstruments zur Erfassung lexikalisch-semanticischer Kompetenzen gehörloser Kinder, konnte erreicht werden. Mit seiner Anwendung und durch quantitative sowie qualitative Auswertungen seiner Ergebnisse erhält man ein differenziertes Bild der einzelnen Testperson unter dem Aspekt Wortschatz bzw. semantisch-lexikalischer Kompetenz auf den sprachlichen Sektoren Gebärden-, Schrift- und Lautsprache. Dadurch wird der gezielte Einsatz pädagogischer Fördermittel möglich. Die Testergebnisse lassen ein Urteil zu über die Wirksamkeit der bei den betreffenden Testpersonen angewandten Methode der Sprachvermittlung und bieten eine Grundlage für die Weiterentwicklung der einzelnen Sprachvermittlungsverfahren. Aufgrund seiner Konzeption kann der Test zukünftigen Forschungsprojekten im Bereich der sprachlichen Entwicklung Gehörloser als Grundlage zur Erfassung differenzierter und zielgruppenspezifischer Daten dienen. Dies stellt eine ermutigende Basis für die Weiterarbeit im Bereich der Erforschung sprachlicher Entwicklungsbedingungen Gehörloser dar.

Die vorliegende Arbeit konzentriert sich auf die spezifischen Bedingungen Gehörloser. Das Instrument eignet sich jedoch grundsätzlich auch für Kinder mit anderen spezifischen Bedingungen, bspw. im Bereich der Sprachheilpädagogik. Als Anschlussprojekte bieten sich Modifikationen des Tests unter anderen spezifischen Blickwinkeln als dem vorliegenden an. Dabei können zentrale konzeptionelle Elemente bis hin zur Item- und Distraktorenwahl übernommen oder nur teilweise abgewandelt bzw. ersetzt werden.

Wie dargestellt wurde, zeigt die Distraktorenanalyse keine bruchlose Bestätigung der angenommenen Theorie bzgl. der qualitativen Abstufung sowie der korrelativen Zusammenhänge zwischen Fehlerwahlprofil und bestimmten Merkmalen der Testpersonen. Gleichwohl gibt es deutliche bestätigende Tendenzen. In einem nächsten Schritt sollten zukünftig mit dem Test zu erhebende Daten gesammelt und im Rahmen einer sich an die Ergebnisse vorliegender Arbeit anschließenden Folgeuntersuchung unter dieser Fragestellung ausgewertet werden. Auf einer solchen erweiterten Basis ist es denkbar, dass repräsentative Aussagen zu dieser Fragestellung getroffen werden können, die dann in ausgeprägterem Maße als gesichert eingestuft werden könnten als derzeit möglich. In Verbindung mit dieser Überlegung ergeht die (auch im Anwenderhandbuch formulierte) Bitte an zukünftige Nutzer, ihre Daten zu Zwecken der Optimierung des Instruments an die Autorinnen weiterzuleiten.

Eine weitere Folgeuntersuchung sollte den Geltungsbereich des Tests in der Weise erweitern, dass neben den Dritt- und Viertklässlern auch die Erst- und Zweitklässler einbezogen werden. In diesen Aufgabenbereich würden Datenerhebungen mit Erst- und Zweitklässlern sowie die Erstellung entsprechender Normierungstabellen fallen. Somit würde Gültigkeit für

den Grundschulbereich realisiert werden. Mit einer so erweiterten Stichprobe ergäbe sich gleichzeitig die Möglichkeit und auch die Notwendigkeit, die Normierung getrennt nach Klassenstufen vorzunehmen.

D. Anwenderhandbuch

1. Beschreibung des Tests

1.1 Vorgeschichte und Aufgabe des Tests

Der Perlesko²¹ ermöglicht objektive und somit vergleichbare Aussagen über die lexikalisch-semantiche Kompetenz gehörloser Schüler im Grundschulalter. Er entstand als ergänzendes und fortsetzendes Element der wissenschaftlichen Begleitforschung des bilingualen Schulversuchs in Hamburg (vgl. Günther 1999). Den Hintergrund bildet die kontroverse Diskussion in Deutschland um die Qualität unterschiedlicher gehörlosenpädagogischer Methoden.

Übereinstimmender Ausgangspunkt jedes Sprachvermittlungsverfahrens ist die Frage, auf welche Weise ein gehörloser Mensch zur Sprachfähigkeit und damit zu einer sprachlich strukturierten Denk- und Kommunikationsfähigkeit als wesentlicher Grundlage seiner personalen und sozialen Existenz geführt werden kann. Einigkeit besteht zudem in der Einsicht, dass dem gehörlosen Menschen (wie jedem nichtbehinderten Menschen auch) zur Entfaltung seiner Identität und positiven Gestaltung seines Lebens mindestens ein funktionsfähiges Kommunikationssystem zur Verfügung stehen muss (vgl. Ahrbeck 1992). Obgleich demnach Konsens in wesentlichen grundlegenden Annahmen besteht, differieren die Aufstellungen methodischer Richtlinien zur Sprachvermittlung deutlich. Übereinstimmende Beschreibungen von Ausgangslage und Zielvorstellung führen nicht zu einer einheitlichen Wegbeschreibung.

Mit der Veröffentlichung dieses Tests liegt erstmals ein standardisiertes Instrument vor, das die passive Wortschatzkenntnis anhand einer repräsentativen Auswahl – entnommen aus und sortiert nach Wortlisten zum Grundwortschatz Hörender von Günther (1989) nach Augst (1984, 1989), Pregel/Rickheit (1987) und Ruoff (1981) – von Nomen, Verben und Adjektiven zu gleichen Anteilen und in Form von drei Sprachmodalitäten (Gebärden-, Schrift- und Lautsprache) testet. Der Test selbst ist damit grundsätzlich neutral und ermöglicht eine angemessene Bewertung der sich unterscheidenden methodischen Wege.

Die mit dem Test zu erhebenden Daten und die sich daraus ergebenden Ergebnisse sollen ihren Nutzen finden in einer gehörlosenpädagogischen Landschaft, die gerade heute durch die aktuellen Entwicklungen neuer Methoden der Sprachvermittlung sowie durch die medizinische Versorgung mit Implantaten u. a. m. Flexibilität erfordert. Oberstes Ziel der Testan-

²¹ Für die Veröffentlichung des Tests wurde diesem der Name ‚Perlesko‘ als Abkürzung für ‚Prüfverfahren zur Erfassung lexikalisch-semantiche Kompetenz‘ gegeben.

wendung ist es, Voraussetzungen und Ansatzpunkte für den Einsatz individueller Fördermaßnahmen im Bereich der kommunikativen Kompetenz (vgl. Leonhardt 1996) gehörloser Grundschüler zu schaffen.

1.2 Anwendungsbereich

Der Perlesko wurde nach gehörlosenspezifischen Gesichtspunkten für 3. und 4. Klassen an Gehörlosenschulen konzipiert. Dabei bieten seine konstruktiven Grundelemente das Potential zur Entwicklung modifizierter Schwesterinstrumente mit spezifischer Berücksichtigung von Bedingungen anderer Behinderungen, bspw. im Bereich der Sprachheilpädagogik.

Als Anschlussprojekte bieten sich demnach Modifikationen des Tests unter anderen spezifischen Blickwinkeln als dem vorliegenden an. Dabei können zentrale konzeptionelle Elemente bis hin zur Item- und Distraktorenwahl übernommen oder nur teilweise abgewandelt bzw. ersetzt werden. Eine weitere Folgeuntersuchung sollte den Geltungsbereich des Tests um Erst- und Zweitklässler erweitern, sodass sein Einsatz im gesamten Grundschulbereich möglich wird.

Um diese Anschlussprojekte realisieren zu können, ergeht die Bitte an zukünftige Nutzer, ihre Daten zu Zwecken der Erweiterung des Datenpools und damit der Optimierung des Instruments an die Universität Hamburg weiterzuleiten:

Universität Hamburg
Fachbereich Erziehungswissenschaft
Institut 5, Behindertenpädagogik
z. Hd. Herrn Prof. Dr. K.-B. Günther
Sedanstraße 19
20146 Hamburg

Vielen Dank!

1.3 Eigenschaften des Tests

Das Testinstrument ist konzeptionell darauf angelegt, ein differenziertes Bild der einzelnen Testperson unter dem Aspekt Wortschatz bzw. semantisch-lexikalischer Kompetenz zu erhalten. Dies erfolgt unter anderem, indem drei Sprachmodalitäten, drei Wortarten und qualitative Auswertungsmaßstäbe Berücksichtigung finden. Dadurch kann ein Bild von den Lei-

stungen der jeweiligen Testperson auf den verschiedenen sprachlichen Sektoren gewonnen werden.

Der Perlesko setzt sich aus drei Subtests zusammen, welche die drei Sprachmodalitäten Gebärdensprache (GS), Schriftsprache (SS) und Lautsprache (LS) testen. Wahlweise können auch nur einer oder zwei der Subtests zur Anwendung kommen. Jeder Subtest testet in seiner Sprachmodalität den gleichen Grundwortschatz, bestehend aus Blöcken von je 18 schwierigkeitsmäßig ansteigenden Nomen (N), Verben (V) und Adjektiven (A) in dieser Reihenfolge. Somit besteht jeder Subtest aus drei, der gesamte Test aus neun Subtestteilen (GSN, GSV, GSA, SSN, SSV, SSA, LSN, LSV, LSA). Hinzu kommen zwei leicht lösbare Übungsbeispiele zu Anfang des Tests. Diese gehen nicht in die Wertung ein, sondern dienen ausschließlich der Einführung in das Verfahren und der Absicherung, dass die Vorgehensweise von der Testperson verstanden wurde.

Den Itemkorpus des vorliegenden Tests bilden die drei referenzsemantischen Hauptwortarten (vgl. Günther 1991) Nomen, Verben und Adjektive (+Inhalts-Adverbien) zu gleichen Anteilen. Die fast ausschließliche Verwendung von Nomen in bisher gebräuchlichen Wortschatztests mag insoweit sinnvoll erscheinen, als dass Hauptwörter über 50% des deutschen Wortschatzes ausmachen, zudem am einfachsten und deutlichsten bildlich darzustellen sind und somit eine Eindeutigkeit des Items am ehesten gewährleistet ist. Ein solches Vorgehen vernachlässigt jedoch die Tatsache, dass sich Nomen als besonders ungeeignet zur Erfassung lexikalisch-semantischer Fähigkeiten und eventueller Störungen erweisen, weil die wenigsten Nomen unabhängig von Alter, sozial-kultureller Erfahrung und angeeignetem Wissen sind. Die Entwicklung des nominalen Wortschatzes weist also in hohem Maße erfahrungs- und entwicklungsbedingte Abhängigkeiten auf, was sich vor allem in der geringen Übereinstimmung des kindlichen und des erwachsenen Nomenwortschatzes, aber auch in Wortschätzen von Kindern unterschiedlicher Herkunft und unterschiedlichen Umfelds ausdrückt. Insofern lässt sich kein allgemeiner nominaler Grundwortschatz feststellen, der diese Abhängigkeiten nicht aufweist. Aus dem Blickwinkel der Linguistik muss zudem nach Günther (1991, 187) festgestellt werden, „daß sich aus Umfang und Niveau des nominalen Wortschatzes für sich genommen kaum etwas über die Fähigkeit schließen lässt, Sprache in semantisch und syntaktisch kohärenten Äußerungen zu gebrauchen. Verben als Organisationszentrum des Satzes sind diesbezüglich viel aussagekräftiger. Adjektive schließlich implizieren durch ihren relativ-dimensionalen Bedeutungscharakter und ihre primär semantisch-qualifizierende Funktion über den reinen Wortschatzaspekt hinaus Hinweise auf das Niveau der Sprachfähigkeit im semantischen Bereich.“

Zu jedem Testitem gehören gemäß dem Aufgaben-Aufbauschema drei Distraktoren. Diese wurden nach verschiedenen Ähnlichkeitskriterien qualitativ abgestuft (vgl. D 5.2) ausge-

wählt, wobei jeder Distraktor gleichzeitig mehrere Ähnlichkeiten aufweisen kann. Ein Distraktor soll auf unterschiedliche Weise, durch Ähnlichkeiten im Hinblick auf Gebärde, Schriftbild, Mundbild oder auf inhaltlicher Ebene, vom Testitem ablenken.

1.4 Testmaterial

Die folgenden Tabellen geben den Itemkorpus inklusive der Distraktoren – nach Wortarten getrennt – wieder:

Tab.95a-D 1.4: Itemkorpus Nomen				
Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	Ei	Huhn	Kartoffel	Eis
2	Bett	Decke	Brett	Sofa
3	Bart	Haare	Kinn	Post
4	Zahn	Mund	Hahn	Zange
5	Angst	Ärger	Trauer	Freude
6	Welt	Planet	Ball	Vase
7	Stein	Felsen	Kaffee	Stange
8	Rose	Sonnenblume	Tulpe	Rasen
9	Zahl	Regen	Buchstabe	Pfahl
10	Leute	Paar	Mensch	Herde
11	Schuh	Stiefel	Strumpf	Uhr
12	Arbeit	Eimer	Stein	Ampel
13	Kopf	Zopf	Junge	Topf
14	Geld	Gold	Salz	Feld
15	Holz	Baum	Brot	Gold
16	Mutter	Mädchen	Bauer	Motte
17	Straße	Weg	Strand	Strauß
18	Garten	Baum	Geschenk	Arbeit

Tab. 95b-D 1.4: Itemkorpus Verben				
Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	schreiben	malen	reiben	schlafen
2	fliegen	springen	winken	fragen
3	stinken	riechen	staunen	winken
4	springen	sitzen	klettern	schimpfen
5	weinen	reiben	schwitzen	wiegen
6	lesen	sitzen	halten	blasen
7	waschen	wischen	schwimmen	warten
8	werfen	fangen	strecken	wärmen
9	schieben	schließen	heben	ziehen
10	verstecken	schlafen	erschrecken	stecken
11	schlafen	schlagen	streicheln	liegen
12	brennen	löschen	rechnen	rennen
13	gehen	stehen	rennen	sehen
14	singen	sagen	warnen	flöten
15	warten	sitzen	wandern	braten
16	gucken	schminken	denken	ducken
17	hinsetzen	sitzen	hinlegen	schwitzen
18	denken	träumen	warnen	schenken

Tab. 95c-D 1.4: Itemkorpus Adjektive				
Nummer	Item	Distraktor I	Distraktor II	Distraktor III
1	glatt	sauber	lockig	platt
2	spitz	struppig	schief	spannend
3	hellblau	dunkelblau	hellgrau	hellgrün
4	schmutzig	sauber	zerknittert	zerrissen
5	orange	rosa	gelb	rot
6	viereckig	dreieckig	fünfeckig	sechseckig
7	warm	heiß	kalt	weich
8	dünn	eng	dick	dumm
9	klein	klug	dünn	fein
10	gestreift	kariert	geblümt	gepunktet
11	nass	trocken	blass	nackt
12	schief	steil	gerade	stark
13	verletzt	scharf	gesund	verliebt
14	blond	braun	rot	schwarz
15	voll	viel	groß	stolz
16	weich	nass	wach	warm
17	nah	geschlossen	fern	nass
18	schwer	stark	leer	schwarz

Hinzu kommen zwei Übungsbeispiele aus dem Bereich der Nomen:

Hund mit den Distraktoren Katze (I), Pferd (I/G) und Hahn (I/M/S)

Auge mit den Distraktoren Nase (I/G), Auto (M/S) und Stirn (HRP/I/G)

Der Test besteht aus zwei Testheften, wobei Testheft I für die Subtests Gebärden- und Lautsprache, Testheft II für den Subtest Schriftsprache verwendet wird. In Testheft I befinden sich vier Schwarzweiß-Zeichnungen (mit Ausnahme der in Farbe dargestellten Farbwörter) pro Seite (DinA5/Querformat), von denen eine das gefragte Item und die drei übrigen die Distraktoren abbilden. In Testheft II befinden sich auf der linken Seite eine Zeichnung, die das gefragte Testitem abbildet, und auf der rechten Seite vier in Druckschrift geschriebene Wörter, von denen eines das Testitem bezeichnet und die drei übrigen die Distraktoren. Für die Anordnung der Bilder bzw. Wörter auf der einzelnen Seite wurde das System der formal bewährten Vierfelder-Wahltafeln übernommen. Die vier Felder sind dabei durch ein Kreuz voneinander abgegrenzt.

Im Verlauf der Testentwicklung wurde nach Anwendung aller testtheoretischen Kriterien die Reihenfolge der Items überarbeitet. Sie entspricht nun nicht mehr ihrem ursprünglichen Vorkommen in der Grundwortschatzliste nach Günther (1989) sondern ist den Ergebnissen der Hauptuntersuchung zufolge (für beide Testhefte unterschiedlich) modifiziert. Die folgenden Tabellen verdeutlichen die Gliederung des Itemkorpus in beiden Testheften:

Tab. 96a-D 1.4: Gliederung des Itemkorpus in Testheft I (Gebärdensprache und Lautsprache)			
Nr.	Nomen	Verben	Adjektive
1	Ei	schreiben	glatt
2	Bett	fliegen	spitz
3	Bart	stinken	hellblau
4	Zahn	springen	schmutzig
5	Angst	weinen	orange
6	Welt	lesen	viereckig
7	Stein	waschen	warm
8	Rose	werfen	dünn
9	Zahl	schieben	klein
10	Leute	verstecken	gestreift
11	Schuh	schlafen	nass
12	Arbeit	brennen	schief
13	Kopf	gehen	verletzt
14	Geld	singen	blond
15	Holz	warten	voll
16	Mutter	gucken	weich
17	Straße	hinsetzen	nah
18	Garten	denken	schwer

Tab. 96b-D 1.4: Gliederung des Itemkorpus in Testheft II (Schriftsprache)			
Nr.	Nomen	Verben	Adjektive
1	Schuh	lesen	klein
2	Ei	weinen	blond
3	Bett	fliegen	schmutzig
4	Zahn	schlafen	spitz
5	Geld	schreiben	hellblau
6	Kopf	gehen	viereckig
7	Zahl	werfen	dünn
8	Holz	springen	orange
9	Arbeit	waschen	nass
10	Rose	warten	warm
11	Straße	verstecken	voll
12	Angst	brennen	schief
13	Stein	stinken	glatt
14	Welt	singen	schwer
15	Bart	schieben	weich
16	Mutter	gucken	verletzt
17	Garten	denken	nah
18	Leute	hinsetzen	gestreift

1.5 Zeichnerische Gestaltung

Die zeichnerische Darstellung der gegenständlichen Nomen ist in aller Regel sowohl einfach als auch eindeutig. Eine solche Einfachheit der Darstellung ist wünschenswert, da auf diese Weise nichts (Zeichnungsimmanentes) von der Wahrnehmung des gefragten Gegenstandes ablenken kann. Wurden in einigen Fällen dennoch ausschmückende Details eingefügt, so geschah dies zur Schaffung eines Verständniskontextes. Allzu puristische Darstellungen können das Verständnis erschweren, da die Gegenstände auch in der Natur bzw. im Alltag der Testpersonen nicht in dieser Weise vorkommen, sondern i. d. R. Verzierungen, modische Elemente o. ä. an sich haben. Zudem sollte das Testinstrument ansprechend gestaltet sein, was ebenfalls für eine gewisse Ausgestaltung spricht.

Die Darstellung der Verben, die z. T. einen prozessualen Ablauf, eine Handlung oder einen Zustand beschreiben, wird vielfach durch angedeutete Bewegungslinien verdeutlicht.

Die zeichnerische Darstellung der Adjektive unterscheidet zwischen absoluten und relativen Adjektiven. Erstere, in vorliegendem Test die Farbadjektive, sind situationsunabhängig und lassen sich daher eindeutig darstellen. Die relativen Adjektive sind meist als Gegensatzpaare

dargestellt, dabei ist der jeweils nicht gefragte Gegensatz durchgestrichen. Dieses Prinzip wurde jedoch nur in den Fällen angewandt, in denen die Darstellung als Gegensatzpaar unerlässlich war. Wäre jedes relative Adjektiv auf diese Weise dargestellt worden, hätte die Testperson pro Seite statt vier Bilder fast immer acht betrachten müssen, was unübersichtlich geworden wäre.

Insgesamt wurde versucht, die charakteristischen Eigenschaften der gefragten Begriffe möglichst deutlich bildlich darzustellen. Die angemessenen Trennschärfeindizes der Aufgaben werden als hinreichendes Kriterium dafür gewertet, dass die bildlichen Darstellungen tatsächlich die intendierten Begriffen wiedergeben.

2 Normierung

2.1 Stichprobe

Mit der Version I des Perlesko wurde in der Zeit von August 1997 bis Januar 1998 eine Stichprobe von 112 gehörlosen und hochgradig schwerhörigen Kindern untersucht. Bei der Auswahl der Schulen wurde versucht, im Sinne der Repräsentativität flächendeckend vorzugehen. Es fanden sich sieben Schulen in deutschen Städten sowie eine Schule in der deutschsprachigen Schweiz bereit, an der Untersuchung teilzunehmen. Nicht in jeder Schule standen Schüler aller drei gewünschten Klassenstufen zur Verfügung. Die daraus resultierende Aufteilung der Testpersonen in Klassenstufen ergibt sich wie folgt: Getestet wurden 40,2% der Testpersonen aus 3. Klassen, 27,7% aus 4. Klassen und 32,1% aus 5. Klassen. Für die Anwendung einiger statistischer Verfahren werden im weiteren Verlauf zur Erweiterung des Datenpools der gehörlosen Grundschüler die beginnenden 5. zu den 4. Klassen gezählt, so dass 40,2% Drittklässler einem Anteil von 59,8% Viertklässlern gegenüberstehen. Diese Zuordnung legitimiert sich aus dem Umstand, dass die betreffenden Testpersonen zu Beginn der fünften Klassenstufe weder altersmäßig noch in Bezug auf Leistungs- bzw. Entwicklungsniveau gravierend von den Viertklässlern abweichen.

Damit wurde etwa ein Viertel aller Testpersonen der entsprechenden Grundgesamtheit erfasst. Dieser Umstand rechtfertigt die Wahl der anfallenden und nicht zufälligen Zusammenstellung der Stichprobe, da bei einer so umfangreichen Erhebung nahe an der Grundgesamtheit von Repräsentativität der Stichprobe ausgegangen werden muss. Gemäß der üblichen Vorgehensweise wurde nach der Erhebung die Annahme der Normalverteilung der Stichprobe überprüft und bestätigt.

2.2 Aufgabenanalyse

Die in der Pilotstudie mit der Testrohversion (18 Nomen, Verben und Adjektive je Subtest) und in der Hauptuntersuchung mit Testversion I (ergänzt auf 24 Nomen, Verben und Adjektive je Subtest) gewonnenen Daten von 11 bzw. 112 Testpersonen wurden in zwei Schritten einer umfangreichen Aufgabenanalyse unterzogen. Hauptaugenmerk im Hinblick auf Aufgabenselektion und -revision lag dabei auf den Schwierigkeits- und Trennschärfewerten der Aufgaben. So wurde für die vorliegende Testversion (18 Nomen, Verben und Adjektive je Subtest) ein Itemkorpus zusammengestellt, der folgende Schwierigkeits- (p) und Trennschärfewerte (r_{it}) der Aufgaben aufweist:

Tab. 97a-D 2.2: Schwierigkeit und Trennschärfe Nomen						
Nr.	GSN		SSN		LSN	
	p	r_{it}	p	r_{it}	p	r_{it}
1. Ei	.99	.17	.96	.20	.89	.10
2. Bett	.89	.25	.86	.45	.88	.39
3. Bart	.96	.25	.50	.45	.72	.33
4. Zahn	.83	.10	.76	.33	.82	.28
5. Angst	.84	.23	.61	.51	.76	.64
6. Welt	.71	.29	.60	.55	.78	.39
7. Stein	.66	.40	.60	.59	.74	.47
8. Rose	.71	.18	.65	.42	.68	.30
9. Zahl	.95	.10	.71	.42	.42	.34
10. Leute	.81	.31	.40	.26	.56	.49
11. Schuh	.61	.32	.96	.27	.72	.44
12. Arbeit	.69	.30	.66	.50	.62	.46
13. Kopf	.75	.04	.72	.39	.48	.34
14. Geld	.56	.07	.76	.50	.68	.31
15. Holz	.56	.52	.66	.58	.60	.14
16. Mutter	.63	.35	.47	.23	.52	.33
17. Straße	.50	.16	.64	.56	.61	.52
18. Garten	.47	.41	.46	.57	.59	.46

Tab. 97b-D 2.2: Schwierigkeit und Trennschärfe Verben						
Nr.	GSV		SSV		LSV	
	p	r_{it}	p	r_{it}	p	r_{it}
1. schreiben	.96	.30	.71	.61	.85	.32
2. fliegen	.99	.28	.74	.57	.82	.35
3. stinken	.96	.19	.44	.40	.82	.33
4. springen	.96	.07	.65	.62	.78	.37
5. weinen	.96	.09	.75	.52	.77	.30

Nr.	GSV		SSV		LSV	
	p	r _{it}	p	r _{it}	p	r _{it}
6. lesen	.96	.23	.77	.61	.77	.50
7. waschen	.97	.10	.61	.63	.72	.54
8. werfen	.96	.37	.66	.40	.72	.29
9. schieben	.94	.26	.31	.25	.65	.38
10. verstecken	.87	.23	.57	.56	.69	.41
11. schlafen	.76	.21	.74	.52	.77	.31
12. brennen	.94	.23	.48	.47	.56	.41
13. gehen	.90	.53	.69	.53	.47	.21
14. singen	.85	.36	.42	.62	.50	.42
15. warten	.75	.45	.57	.43	.58	.32
16. gucken	.68	.40	.28	.15	.42	.27
17. hinsetzen	.36	.17	.21	.36	.40	.28
18. denken	.46	.37	.23	.15	.29	.24

Tab. 97c-D 2.2: Schwierigkeit und Trennschärfe Adjektive

Nr.	GSA		SSA		LSA	
	p	r _{it}	p	r _{it}	p	r _{it}
1. glatt	.96	.17	.54	.37	.79	.35
2. spitz	.91	.18	.66	.47	.71	.39
3. hellblau	.79	.34	.66	.55	.81	.34
4. schmutzig	.77	.32	.68	.43	.76	.49
5. orange	.91	.14	.60	.28	.60	.53
6. viereckig	.92	.10	.65	.42	.54	.44
7. warm	.79	.29	.57	.54	.66	.53
8. dünn	.93	.02	.63	.48	.51	.46
9. klein	.89	.18	.73	.39	.52	.30
10. gestreift	.72	.19	.22	.29	.50	.54
11. nass	.55	.38	.58	.54	.66	.31
12. schief	.68	.15	.54	.42	.43	.52
13. verletzt	.45	.40	.48	.47	.59	.37
14. blond	.46	.46	.73	.50	.58	.47
15. voll	.44	.43	.54	.59	.59	.42
16. weich	.59	.36	.52	.56	.41	.58
17. nah	.46	.21	.29	.36	.43	.31
18. schwer	.17	.25	.52	.63	.42	.44

Da mit Hilfe des Wortschatztests explizit der Grundwortschatz überprüft werden soll, indem häufig verwendete und demzufolge leichte Items geprüft werden, ist eine Häufung vornehmlich im hohen Schwierigkeits-, und damit im mittleren Trennschärfewertebereich, gerech-

fertigt. Die generell erwartete parabolische Abhängigkeit zwischen Schwierigkeit und Trennschärfe muss demnach für diesen Test einer Verteilungsform weichen, die im Koordinatensystem vor allem im rechten Bereich der Abzisse und im mittleren Bereich der Ordinate differenziert.

In der Regel wird ein neuer Itemkorpus an einer neu erhobenen Stichprobe überprüft. Da die Modifikation von Testversion I zu II des Wortschatztests jedoch anhand der Ergebnisse einer im Bezug auf die relativ kleine Grundgesamtheit besonders umfangreichen Stichprobe vorgenommen wurde, wurde im vorliegenden Fall auf die Erhebung einer Kontrollstichprobe verzichtet. Alternativ wurde für die ursprüngliche Analysestichprobe eine neue Punktbewertung der Version I ohne die eliminierten Aufgaben vorgenommen und eine Aufgaben-Re-Analyse (vgl. Lienert/Raatz 1994, 126) durchgeführt. Aufgabenschwierigkeit und -trennschärfe nach einer zufälligen Halbierung dieser Re-Analysen-Stichprobe wiesen die Testendform als hinreichend stabil aus. Die Überprüfung der Gütekriterien und die Normierung erfolgten demgemäß anhand der Daten der in D 2.1 beschriebenen Analysestichprobe, für deren Bewertung die eliminierten Aufgaben ausgenommen werden.

2.3 Prozentrangskalen und lineare Standardwerte

Um eine eindeutige Interpretation der Rohdaten einer Testperson aus dem Wortschatztest zu gewährleisten, ist eine Orientierung an vergleichbaren Messträgern nötig. Derartige Vergleichswerte werden durch die Normierung bereitgestellt. Ziel der Normierung des Wortschatztests war dabei, die erhaltenen Rohdaten anschaulicher, direkter vergleichbar und messtechnisch möglichst höherwertig zu machen.

Um eine fundierte Interpretation der gewonnenen Daten zu gewährleisten, wurden die Rohscores des Tests zunächst auf die Prozentrangskala transformiert, um sie anschaulicher und direkt vergleichbar zu machen. Eine Schwierigkeit in der Prozentrangberechnung vorliegender Daten lag in der ganzzahligen Abstufung der Variablen, obgleich diese Variable als psychologische Fähigkeit an sich stetig ist. Genaugenommen befindet sich die Leistung einer Testperson, die den Rohscore x erhält, also irgendwo zwischen $x-0,5$ und $x+0,5$ Punkten. Aus diesem Grund wurde die Hilfsannahme getroffen, dass sich alle Testpersonen mit dem Rohscore x den Platz zwischen $x-0,5$ und $x+0,5$ gleichabständig teilen. So kommt es, dass möglicherweise mehr als die objektiv schlechteren Testpersonen (Rohscore $x-1$ und schlechter), nämlich auch diejenigen mit dem Rohscore x , die zwischen $x-0,5$ und $x+0,5$ angesiedelt sind, den Prozentrang der Testperson bestimmen.

Die Skala der Prozentränge erfüllt die Ansprüche an Anschaulichkeit und Vergleichbarkeit, weist allerdings auch Nachteile auf. Zum einen täuscht sie ein nicht geltendes Skalenniveau vor, da 60% zwar doppelt so viel sind wie 30%, eine Testperson mit dem Prozentrang 60 jedoch nicht zwangsläufig doppelt so gut ist wie eine mit Prozentrang 30. Diese Skala hat demnach kein Intervallskalenniveau. Zum anderen gibt die Prozentrangskala eine unrealistische Messpräzision vor, wenn aufgrund der Angabe von Prozenträngen angenommen wird, das Messinstrument sei zu hundert Unterscheidungen fähig. Das Instrument kann nur der Anzahl der Scores entsprechend viele Prozentränge, also maximal 18 pro Subtestteil, angeben.

Um zur Anschaulichkeit und Vergleichbarkeit auch die Anhebung auf Intervallskalenniveau zu gewährleisten, wurde zusätzlich zu den Prozenträngen die Transformation der Rohwerte in lineare z-Werte vorgenommen. Da die Rohwertverteilungen der Stichprobe annähernd normal verteilt sind, erhielt man bei der linearen Transformation nicht nur z-Werte, sondern Standard-z-Werte einer Standardnormalverteilung, die automatisch Intervallskalenniveau besitzen. Diese Werte haben den Mittelwert 0 und die Standardabweichung 1 und erfüllen somit den Anspruch an direkte Vergleichbarkeit. Um negative z-Werte zu vermeiden, werden die z-Werte in anschaulichere Z-Werte transformiert, die einen Mittelwert von 100 und eine Standardabweichung von 15 aufweisen.

Für die Normierung des Wortschatztests wurden die Ergebnisse der Dritt-, Viert- (und Fünft-) klässler zusammengefasst. Dies geschah zum einen, um eine Reduzierung des Datenpools und damit Zufallseinflüsse zu vermeiden, zum anderen unter der Voraussetzung, dass mit den Ergebnissen des Tests keine expliziten Aussagen über Leistungsunterschiede in der lexikalisch-semantischen Kompetenz zwischen Dritt- und Viertklässlern getroffen werden sollen. Vielmehr soll der Test dem unterrichtenden Pädagogen eine Aussage darüber ermöglichen, ob ein hörgeschädigter Grundschüler (egal welcher Klassenstufe) einen vergleichbaren Grundwortschatz besitzt wie ein hörender.

Auf die Darstellung eines Gesamtwertes aller Subtestteile wurde bei der Normierung bewusst verzichtet. Wesentliche Informationen über die Wortschatzkenntnis einer Testperson können aus der Interpretation der Subtestteilergebnisse und nicht aus einem Gesamtwert gewonnen werden. Zwar lassen extreme Gesamtwerte Rückschlüsse auf ein homogenes Leistungsvermögen in allen Subtestteilen zu, Gesamtwerte im mittleren Bereich können jedoch unterschiedlich bedingt sein. Die Testperson kann einen solchen Wert bei durchschnittlicher Leistung in allen Skalen oder aber durch überdurchschnittliche Leistung in einigen, unterdurchschnittliche Leistung in anderen Subtestteilen erzielen. Gesamtwerte verlieren somit an Aussagekraft; interessanter sind Aussagen dazu, in welchen Subtestteilen eine Testperson über- oder unterdurchschnittliche Leistungen erbracht hat.

Die Darstellung der Normierungstabellen erfolgt unter D 5.1.

3 Gütekriterien

3.1 Objektivität

Unter dem ersten Gütekriterium der klassischen Testtheorie, der Objektivität, wird Unabhängigkeit der Ergebnisse von der Durchführung, der Auswertung und der Interpretation verlangt.

Die Durchführungsobjektivität ist für den Perlesko durch die in schriftlicher Form vorliegende Testanweisung eindeutig gewährleistet, sofern sich der Testleiter an diese Anweisung hält oder sie zumindest nicht inhaltlich verändert. Im Subtest Gebärdensprache werden dem Testleiter die zu verwendenden Gebärden als Gebärdenphotos mit Ausführungsanweisung vorgegeben. Die Durchführungsobjektivität wird demnach zusätzlich durch die Art und Weise bestimmt, in der der Testleiter im Stande ist, diese Vorgaben in für die Testperson verständliche Gebärden umzusetzen. Die Auswertungsobjektivität ist dadurch gewährleistet, dass eindeutig festgelegt ist, welche der Antwortmöglichkeiten richtig ist. Die Interpretationsobjektivität spielt auf dieser Stufe der Testkonstruktion, auf der für erzielte Rohwerte Prozentrangangaben in Hinblick auf unterschiedliche Bezugsgruppen gemacht werden, keine entscheidende Rolle. Die Umwandlung von Rohwerten in Prozentränge ist nicht durch Interpretationsspielräume seitens des Testleiters beeinflussbar und somit objektiv.

3.2 Reliabilität

Die Reliabilität eines Tests ist gewährleistet, wenn mit ihm ein reproduzierbares Testergebnis gewonnen werden kann, das die Leistung einer Testperson genau auf der Testskala einordnet.

3.2.1 Innere Konsistenz und Split-half

Sowohl bei der Überarbeitung der Pilotversion dieses Tests zu Testversion I als auch bei der Modifikation zu vorliegender Version wurde auf eine sorgfältige Aufgabenanalyse Wert gelegt. Für die Endform der Subtestteile wurden die Aufgaben aus einem umfangreicheren Aufgabenpool (24 statt 18 Aufgaben) selektiert, deren Eignung im Hinblick auf die Zuverlässigkeit des Tests nach der Methode der Inneren Konsistenz (wobei jede einzelne Aufgabe

als gesonderter Testteil mit allen anderen korreliert wurde) sowie der Split-half-Methode (Testhalbierung nach odd-even-Methode) geprüft wurde. Während diese Methoden der einmaligen Untersuchung lediglich durch die Auswertungsobjektivität, die Inhomogenität der Items, sowie punktuelle Schwankungen der Untersuchungsbedingungen (z. B. Störungen) und/oder des Merkmals beeinflusst werden, sind Paralleltest- und Retest-Verfahren zusätzlich von der Veränderlichkeit der Testdurchführungsbedingungen bestimmt. Demzufolge geben die Innere Konsistenz und die Split-half-Methode Aufschluss über die Leistungsfähigkeit des Tests und kommen dem eigentlichen Wesen der Reliabilität, nämlich ihrer instrumentellen Seite, am nächsten (vgl. Lienert/Raatz 1994, 201).

Die Reliabilitätsschätzung über Retest-Verfahren ist erst üblich und sinnvoll, wenn die Methoden Split-half und Innere Konsistenz durchgeführt und angemessene Ergebnisse erzielt wurden. Dies sowie die Anwendung der Paralleltest-Methode mit einem dem zu untersuchenden Test streng vergleichbaren Instrument stehen noch aus.

Der Anspruch, für den Reliabilitätskoeffizienten Werte von .90 zu erreichen, wurde in Anlehnung an Lienert/Raatz für den Wortschatztest als unrealistisch hoch angesehen, wenn man den Einfluss verschiedener Faktoren (vgl. 1994, 202ff) berücksichtigt: Für Tests, die einen umschriebenen Wissensstand wie den Wortschatz prüfen, gilt die Untersuchung der Reliabilität als erschwert. In einem solchen Fall kommen Paralleltests aufgrund des speziellen Aufgabeninhaltes nicht in Betracht und auch Testwiederholungen sind aufgrund des Erinnerungseffekts nicht möglich. Grundsätzlich gilt zudem, dass mit der Länge eines Tests und bei sonst vergleichbaren Bedingungen seine Reliabilität steigt. Eine Verlängerung des Wortschatztests würde demzufolge zwar eine mögliche Reliabilitätsverbesserung mit sich bringen, ist aber unter Berücksichtigung der zeitlichen und kognitiven Leistungsmöglichkeiten der Testpersonen zu verwerfen. Auch die Streuung der Aufgabenschwierigkeiten ist von großer Bedeutung für die Reliabilität, die bei mittelschweren Aufgaben und bei sonst vergleichbaren Bedingungen potentiell am höchsten liegt. Die spezielle Konzeption des Wortschatztests sieht jedoch vor allem Aufgaben von geringer Schwierigkeit (also hohen Schwierigkeitskoeffizienten) vor, sodass von vornherein keine optimalen Reliabilitätsbedingungen für diesen Test gegeben sind. Zusätzlich vermindert der Zufallseinfluss bei der Darbietung von Mehrfachwahlaufgaben die Reliabilität. In Anbetracht dieser verschiedenen Bedingungen wurden im vorliegenden Fall niedrigere Werte ab .70 als zufriedenstellend bewertet.

In der folgenden Tabelle sind die Werte der Reliabilitätsschätzung (r_{tt}) nach der Split-half-Methode (über Spearman-Brown und Kristof) sowie nach der Methode der Inneren Konsistenz (über Horst und Kuder/Richardson 8) aufgeführt. Zum Erhalt realistischer Koeffizienten für den Gesamttest wurde darauf verzichtet, die Formeln auf eine Gesamtskala mit 81 Aufgaben pro Testhälfte anzuwenden, da dies zu einer Überschätzung der Reliabilität führen

würde. Statt dessen wurde die Berechnung des Reliabilitätskoeffizienten für den Gesamttest über Umwandlung in z' -Werte nach Fisher (in Lienert/Raatz 1994, 417) vorgenommen.

Tab. 98-D 3.2.1: Reliabilitätskoeffizienten				
	Split-half		Innere Konsistenz	
	r_{tt} Spearman- Brown	r_{tt} Kristof	r_{tt} Horst	r_{tt} Kuder / Richard- son 8
GSN	.79	.79	.85	.67
GSV	.83	.83	.86	.68
GSA	.79	.79	.83	.69
SSN	.87	.87	.91	.85
SSV	.86	.86	.90	.87
SSA	.87	.87	.91	.86
LSN	.85	.85	.90	.80
LSV	.87	.87	.88	.77
LSA	.87	.88	.91	.84
gesamt über z'	.85	.85	.89	.80

Die Reliabilitätskoeffizienten, die sich nach den Berechnungen mit unterschiedlichen Methoden ergeben, weisen den Gesamttest und alle Subtestteile als gut bis sehr gut reliabel aus.

3.2.2 Standardmessfehler

Wenn die Ergebnisse einer Testperson in verschiedenen Subtestteilen des Wortschatztests stark variieren, kann dies eine Folge des Messfehlers, also zufallsbedingt sein. Der Standardmessfehler gibt die Genauigkeit an, mit der aufgrund des Testergebnisses Aussagen über die wahre Merkmalsausprägung einer Testperson getroffen werden können. Mit einer Wahrscheinlichkeit von 2:3 liegt dieser wahre Wert einer Testperson in einem Bereich von einem Standardmessfehler ober- bzw. und unterhalb des gemessenen Wertes, mit einer Wahrscheinlichkeit von 19:20 liegt er in dem Streubereich von zwei Standardmessfehlern ober- bzw. und unterhalb des gemessenen Wertes:

Tab. 99-D 3.2.2: Standardmessfehler der Subtestteile									
Subtestteil	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
Standardmessfehler	1,31	0,92	1,32	1,52	1,69	1,67	1,49	1,43	1,58

Je geringer der Standardmessfehler einer Skala ausfällt, desto höher ist ihre Messgenauigkeit zu bewerten. Die Standardmessfehler des Wortschatztests weisen bei Werten zwischen 0,92 und 1,69 Punkten eine gute Messgenauigkeit der einzelnen Subtestteile aus.

3.3 Validität

Die Validität eines Tests beschreibt den Grad, mit der er das Merkmal, welches er zu messen beansprucht, tatsächlich misst.

Die Erwartungen an die Höhe des Validitätskoeffizienten hängen wesentlich von dem jeweiligen Verwendungszweck des Tests ab. Da der vorliegende Test zwar ein eindeutiges Urteil über den Ausprägungsgrad des zu untersuchenden Merkmals einer Testperson erlaubt, jedoch keine Beurteilung der Testperson überwiegend oder ausschließlich von seinem Testergebnis abgeleitet werden soll, muss der Test einen Validitätsanspruch von $r_{tc} \geq 0,5$, und nicht etwa von $r_{tc} > 0,7$ (vgl. Lienert/Raatz 1994, 270) erfüllen.

Grundsätzlich ist der Grad der externen bzw. der internen Validität von der Auswahl und der Kontrolle von Störvariablen abhängig. Bei der Erhebung des Wortschatzes können die Uhrzeit, der Termin, die Lichtverhältnisse u. ä. potentiell bedeutsam sein, obwohl sie weder als abhängige noch als unabhängige Variablen auf die Untersuchung Einfluss nehmen. Während einige dieser Variablen wie Lichtverhältnisse oder Gebärdenvorgabe von den Testleitern kontrolliert werden können, müssen andere Variablen wie Termin und Uhrzeit zufällig gewählt werden. Durch das so festgelegte Maß der Kontrolle dieser untersuchungsbedingten Störvariablen ist die Untersuchung im mittleren Bereich des Kontinuums von Untersuchungsvarianten angesiedelt. Dieser Kompromiss aus Labor- und Felduntersuchung gewährleistet sowohl interne als auch externe Validität. Damit sind die Ergebnisse der Untersuchung sowohl eindeutig interpretierbar, ohne dass Alternativerklärungen für ihr Zustandekommen herangezogen werden müssen, als auch über die besonderen Bedingungen der Untersuchungssituation und über die untersuchten Personen hinaus allgemein gültig, indem sie die Natürlichkeit der Untersuchungsbedingungen und die Repräsentativität der untersuchten Stichprobe berücksichtigt (vgl. Bortz 1993, 8).

3.3.1 Inhaltliche Validität

Die inhaltliche Validität eines Instrumentes ist gewährleistet, wenn der Test oder seine Teile das zu messende Merkmal so unzweifelhaft repräsentieren, dass sie selber das Kriterium für dieses Merkmal darstellen. Dies ist bei Kenntnistests wie dem Wortschatztest grundsätzlich

der Fall (vgl. Lienert/Raatz 1994, 225). Die Items des Wortschatztests entstammen einer den Grundwortschatz hörender Grundschüler abbildenden Wortliste und wurden aus dieser systematisch in regelmäßigen Abständen über den Gesamtkorpus verteilt ausgewählt. Sie bilden somit eine repräsentative Stichprobe aus dem Grundwortschatz. Demgemäß wurde auf eine Einschätzung durch Experten verzichtet, da die inhaltliche Validität des Wortschatztests trivial ist (vgl. Lienert/Raatz 1994, 225).

3.3.2 Konstruktvalidität

Für die Überprüfung der Konstruktvalidität des Wortschatztests wurden aus der theoretischen Analyse des Konstruktes Hypothesen deduziert, die daraufhin empirisch überprüft wurden. Da es sich bei der 'Kenntnis des passiven Wortschatzes' um eine vielschichtige, nicht direkt operational fassbare Einheit (vgl. Lienert/Raatz 1994, 226) handelt, ist eine differenzierte Auseinandersetzung mit diesem Konstrukt erforderlich. Theoretisch analysiert stellt sich das Konstrukt als die Wiedererkennung und richtige Zuordnung von konventionellen Bezeichnungen dar. Die Verfügbarkeit dieser Bezeichnungen wird als Anzeiger für die entwickelte Versprachlichung kognitiver Prozesse angesehen. Der Ausprägungsgrad dieses Konstruktes wird von verschiedenen Faktoren bestimmt. Generell wächst die Kenntnis des Wortschatzes mit zunehmendem Alter und mit steigendem Intelligenzniveau und ist vom Geschlecht relativ unabhängig. Einen weit deutlicheren Einfluss als das Alter übt bei Gehörlosen jedoch die Einordnung der Testpersonen in Klassenstufen aus. Diese erfolgt nach dem jeweiligen Förderbedarf, sodass Schüler unterschiedlichen Alters, aber gleichen Leistungsniveaus in einer Klasse zusammengefasst werden. Darüber hinaus nehmen sowohl schulische als auch familiäre Sozialisationsbedingungen Einfluss auf die Wortschatzentwicklung gehörloser Kinder: Sie wird auf schulischer Seite maßgeblich von dem angewandten sprachlichen Erziehungskonzept geprägt. Wie Untersuchungsergebnisse des Hamburger Bilingualen Schulversuchs u. a. belegen, muss bei bilingual erzogenen gehörlosen Kindern von einem differenzierter ausgebildeten Wortschatz ausgegangen werden als bei rein oral- oder gebärdenorientierter Erziehung. Als zentrale Begründung dafür gilt der Umstand, dass den bilingual erzogenen Kindern mit der Gebärdensprache ein Sprachsystem zur Verfügung steht, auf dessen Basis nicht muttersprachliche, lautsprachliche Fähigkeiten vollständiger erworben werden können (vgl. Günther u. a. 1999; Prillwitz / Wudtke 1988). Ein Vergleich der Fähigkeiten bzgl. der Sprachmodalitäten weist demnach die bilingual erzogenen Kinder als kenntnisreich sowohl in der Gebärden- als auch in der Lautsprache aus, wohingegen die oral erzogenen Schüler vor allem Stärken in der Laut-, die gebärdenorientiert erzogenen Schüler in der Gebärdensprache aufweisen. Dabei haben die Ergebnisse der Untersuchungen mit diesem Test belegt, dass aufgrund der generellen Neigung Gehörloser, mit der Gebärdensprache als der ihnen am leichtesten zugänglichen Modalität zu kommunizieren, alle Kinder in der Modalität Gebärdensprache am leistungsstärksten abschneiden, und dies unab-

hängig von dem ihrer jeweiligen Erziehung zugrundeliegenden Konzept. Zudem üben familiäre Rahmenbedingungen wie der Umstand, ob die Eltern des gehörlosen Kindes selber gehörlos oder hörend sind, ob also von der Gebärdensprache als Muttersprache ausgegangen werden kann oder nicht, einen Einfluss auf seine Wortschatzentwicklung aus. Gehörlose Kinder gehörloser Eltern erlernen aufgrund der Basis der Gebärdensprache, die sie sich von Geburt an entsprechend der lautsprachlichen Lernprozesse hörender Kinder aneignen, die Lautsprache als Zweitsprache leichter und vollständiger als gehörlose Kinder hörender Eltern. Daher kann davon ausgegangen werden, dass gehörlose Kinder gehörloser Eltern ein höheres Wortschatzniveau in allen Sprachmodalitäten aufweisen.

Zur Überprüfung der Konstruktvalidität wurde eine Analyse der interindividuellen Unterschiede in den Testresultaten vorgenommen. Untersuchungsziel dabei war die Überprüfung der Annahme, dass der Test für die verschiedenen die Wortschatzentwicklung beeinflussenden Faktoren sensibel ist. Es zeigten sich deutliche Zusammenhänge in oben beschriebener Weise, was als Beleg dafür angesehen wurde, dass das Testinstrument tatsächlich das beabsichtigte Konstrukt 'Kenntnis des passiven Wortschatzes' misst. Herangezogen wurden folgende Faktoren, korreliert nach Pearson und minderungskorrigiert:

Zeile 1: Klasse, in den gegebenen Abstufungen 3 und 4 (+5)

Zeile 2: Intelligenzquotient nach Raven (s. u.), in den Abstufungen der Testpunktwerte 67 bis 132

Zeile 3: Konzept, in den Abstufungen 1 (oral), 2 (gebärdenorientiert) und 3 (bilingual)

Zeile 4: Hörstatus der Eltern, in den Abstufungen 1 (hörende Eltern) und 2 (gehörlose Eltern)

Zeile 5: Geschlecht, in den Abstufungen 1 (weiblich) und 2 (männlich)

Zeile 6: Alter, in den gegebenen Abstufungen 7;11 bis 13;3 Jahre

Tab. 100-D 3.3.2: Korrelationen Wortschatztest – Faktoren									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
1. Korr. Klasse	.35	.36	.33	.44	.41	.40	.31	.29	.34
Mind.-korrektur	.39	.40	.37	.47	.43	.44	.33	.32	.36
p	.000	.000	.000	.000	.000	.000	.000	.000	.000
2. Korr. IQ	.33	.32	.40	.32	.36	.42	.30	.22	.38
Mind.-korrektur	.39	.37	.46	.36	.39	.47	.35	.25	.43
p	.000	.001	.000	.001	.000	.000	.002	.026	.000
3. Korr. Konzept	.31	.10	.29	.29	.27	.29	.20	.22	.18
Mind.-korrektur	.35	.12	.32	.31	.28	.31	.22	.24	.19
p	.001	.287	.002	.002	.005	.003	.042	.024	.073

	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
4. Korr. Eltern	.38	.20	.37	.31	.31	.37	.26	.24	.30
Mind.-korrektur	.42	.22	.41	.33	.32	.41	.29	.26	.32
p	.000	.039	.000	.001	.001	.000	.006	.013	.002
5. Korr. Sex	.12	-.10	-.07	-.07	-.03	-.01	.07	-.01	.02
Mind.-korrektur	.13	-.11	-.07	-.07	-.03	-.01	.08	-.02	.02
p	.213	.305	.501	.490	.794	.894	.472	.889	.879
6. Korr. Alter	-.00	.08	.01	.09	.06	.04	.05	.07	.03
Mind.-korrektur	-.01	.09	.01	.09	.06	.04	.06	.08	.04
p	.969	.395	.934	.359	.561	.686	.587	.475	.738

Gemäß der Hypothese bestehen sehr signifikante Zusammenhänge zwischen der Wortschatzentwicklung und den Einflussgrößen ‘Klassenstufe’, ‘Intelligenzniveau’ sowie ‘Hörstatus der Eltern’. In sieben von neun Subtestteilen hängt die Kenntnis des passiven Wortschatzes signifikant von dem ‘Erziehungskonzept’ ab. Im Falle von ‘Geschlecht’ und ‘Alter’ ist kein signifikanter Einfluss auf das Wortschatzniveau zu erkennen. Die Variation der Testergebnisse steht demnach in erwarteter Relation zu der Variation der schulischen und familiären Sozialisationsbedingungen. Diese Ergebnisse lassen darauf schließen, dass es sich bei dem vorliegenden Test tatsächlich um ein Messinstrument zur Erfassung des Konstruktes ‘Kenntnis des passiven Wortschatzes’ handelt, da sich die erwarteten Beziehungen zwischen dem Testergebnis und den untersuchten, den Wortschatz beeinflussenden Variablen bestätigt haben. Dadurch wird dem Wortschatztest eine Konstruktvalidität im Sinne der Abbildung der Wirkung unterschiedlicher Sozialisationsbedingungen zugeschrieben.

3.3.3 Übereinstimmungsvalidität (Ratingverfahren)

Zu Beginn des Validierungsprozesses eines Testinstruments steht in der Regel kein zufriedenstellendes Kriterium zur Verfügung (vgl. Lienert/Raatz 1994, 221). Erst im prozessualen Verlauf der Suche nach einem geeigneten Kriterium können vorläufige Kompromisse gegen bessere Kriterien ausgetauscht werden. So wurde auch für die Validierung des Wortschatztests zunächst ein vorläufiges, subjektives Außenkriterium in Form eines Ratingverfahrens (vgl. Lienert/Raatz 1994, 221ff) erhoben.

Für die kriterienbezogene Validitätsprüfung des Wortschatztests wurde parallel zu den Testpunktwerten im Sinne eines gesondert erhobenen Außenkriteriums ein Schätzurteil über die mutmaßliche Leistung einer Testperson durch die Klassenlehrer erhoben. Für diese Übereinstimmungsvalidität des Wortschatztests nahmen die Klassenlehrer der Testpersonen eine

Einschätzung der passiven Wortschatzkenntnis ihrer Schüler vor, indem sie sie auf einer qualitativ abgestuften Ratingskala einordneten. Es wurde eine siebenstufige Ratingskala mit Einteilungen der Leistung von sehr schlecht bis sehr gut vorgegeben. Für jede Testperson wurden drei Ratings erhoben, um ihre mutmaßliche Leistung in den drei Sprachmodalitäten getrennt zu erfassen. Nicht alle Klassen konnten an diesem Rating teilnehmen, sodass aufgrund zu geringer Testpersonenanzahl keine zufällige Stichprobe aus der Analysestichprobe gewählt werden konnte. Da eine Validitätsstichprobe von einer Größe ab 30 Testpersonen als ausreichend umfangreich angesehen wird (vgl. Lienert/Raatz 1994, 241), stellen die in etwa normalverteilten Rating-Ergebnisse von insgesamt 62 Testpersonen eine genügend große Stichprobe dar.

Die Skalenzuordnung wurde nach der Repräsentativgruppen-Methode (vgl. Lienert/Raatz 1994, 245) mit den Testergebnissen der dritten Version korreliert. Es wurden dabei jeweils die Testergebnisse der drei Subtestteile einer Modalität mit dem entsprechenden Rating-Ergebnis korreliert. Die Gesamtergebnisse einer Modalität oder einer Wortart sowie der Gesamtscore des Tests wurden über Transformation in z' -Werte nach Fisher berechnet:

Tab. 101-D 3.3.3: Korrelation Rating – Testergebnis				
Modalität	Nomen	Verben	Adjektive	gesamt
Gebärdensprache	.58	.44	.58	.54
Schriftsprache	.67	.79	.78	.76
Lautsprache	.62	.62	.70	.65
gesamt	.63	.64	.70	.65

Alle Ergebnisse sind sehr signifikant, sodass ein wahrer Zusammenhang zwischen Test und Kriterium angezeigt wird. Da bei der Erhebung des Kriteriums Skalenwerte für die Fähigkeiten in einzelnen Modalitäten verteilt wurden, die die Fähigkeiten in den einzelnen Wortarten der entsprechenden Modalität beinhalten und zusammenfassen, können die erhaltenen Koeffizienten möglicherweise als zu wenig differenzierend und damit zu gering angesehen werden. Unter Berücksichtigung dieses Sachverhalts wurden für die Gesamtwerte der Modalitäten zusätzlich direkte Korrelationen mit den erhobenen Kriterienwerten ohne Transformation in z' -Werte berechnet. Dafür wurde für jede Testperson ein Summenwert für jede Modalität aus den drei entsprechenden Subtestteilergebnissen des Wortschatztests errechnet und mit dem Ratingergebnis der Modalität korreliert. Durch derartige Korrelation ergeben sich sogar Werte von .64 (GS), .79 (SS) und .70 (LS). Über Transformation in z' -Werte lässt sich daraus eine Gesamtkorrelation für den Wortschatztest von .72 berechnen, die die Validität des Tests noch höher einstuft. Während die Korrelationswerte durch die z' -Transformation leicht unterschätzt werden, fallen die Zusammenhänge bei dieser Methode etwas zu hoch aus. Der wahre Korrelationswert wird vermutlich in der Mitte der Ergebnisse beider

Verfahren, also für Gebärdensprache um .59, für Schriftsprache um .78 und für Lautsprache um .68 anzusiedeln sein. Der wahre Wert für den Gesamttest liegt damit um .69.

3.3.4 Übereinstimmungsvalidität (Raven)

Zudem wurden die Ergebnisse des Wortschatztests mit denen eines Tests korreliert, der ein verwandtes Persönlichkeitsmerkmal erfasst. Da Wortschatz aus Sicht der Intelligenzforschung als ein Teilaspekt der Intelligenz angesehen wird (vgl. HAWIK-R Hamburg-Wechsler Intelligenztest für Kinder Revision 1983, Handbuch und Testanweisung 1985, 15/16), wurde der Analysenstichprobe neben dem Wortschatztest auch der RAVEN-Matrizen-Test, CPM (1980) vorgelegt. Der Raven wird in diesem Zusammenhang als ökonomisches Kontrollinstrument angesehen, mit dem in non-verbaler Form gravierende Einflüsse auf die Wortschatzentwicklung im Allgemeinen und die Überprüfung der passiven Wortschatzkenntnis mit dem Wortschatztest im Speziellen festgestellt und ausgeschlossen werden können.

Die Ergebnisse des Wortschatztests und des Raven wurden nach Pearson (erste Zeile) sowie einfach (zweite Zeile) und zweifach minderungskorrigiert (dritte Zeile) korreliert. Die Signifikanzwerte werden in der vierten Zeile aufgeführt. Bei der Minderungskorrektur wurde für den Raven ein Reliabilitätskoeffizient von .90 (RAVEN-Matrizen-Test, CPM, Manual 1980, 14), für den Wortschatztest der Koeffizient nach Kristof verwendet.

Tab. 102-D 3.3.4: Korrelation Raven – Testergebnis										
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA	ges.
Korrela-tion										
Raven	.33	.315	.40	.32	.36	.42	.30	.22	.38	.36
einfach										
mind.-korr.	.35	.332	.42	.34	.38	.44	.32	.23	.40	.37
zweifach										
mind.-korr.	.39	.371	.46	.36	.39	.47	.35	.25	.43	.39
p	.000	.001	.000	.001	.000	.000	.002	.026	.000	.000

Diese Ergebnisse sind unter zwei Gesichtspunkten zu interpretieren: Die theoretische Bedeutsamkeit einer Korrelation zeigt sich generell in der Unterschreitung eines Signifikanzniveaus. Sie liegt in diesem Fall hoch, da die Signifikanzergebnisse in acht von neun Fällen, wie auch für den Gesamttest, auf der 1%-Stufe, im neunten Fall immer noch auf der 5%-Stufe gesichert sind. Somit besteht bei geringer Irrtumswahrscheinlichkeit ein wahrer, interpretierbarer Zusammenhang zwischen Test und Kriterium. Die Wahl des Raven als Außenkriterium impliziert und rechtfertigt gleichzeitig die verhältnismäßig niedrigen Korrelations-

koeffizienten. Der Wortschatztest berücksichtigt in stärkerem Maße als ein Intelligenztest (und besonders als der non-verbale Raven) die verbale Entwicklung, zudem kann Wortschatz nur als ein Teilaspekt der gesamten Intelligenz angesehen werden, sodass nur ein bedingter Grad der Gemeinsamkeit zwischen Test und Kriterium besteht. Die entsprechenden Koeffizienten sind als Anzeiger der praktischen Relevanz demnach als gut zu bewerten, da ein stärkerer Zusammenhang zwischen Test und Kriterium nicht zu erwarten ist.

4 Durchführungsanleitung

4.1 Allgemeine Instruktionen

Der Test besteht aus maximal drei Testsequenzen, die im Abstand von jeweils ca. 14 Tagen durchgeführt werden. An den drei Terminen werden die drei Subtests in der Reihenfolge Gebärden-, Schrift- und Lautsprache getestet. Die Einhaltung dieses Zeitraums bei Anwendung mehrerer Subtests ist unter zwei Aspekten notwendig: Zum einen muss, da der Itemkorpus in den Subtests nicht variiert, sich also die Items bei mehrfacher Anwendung wiederholen, eine reine Wiedererkennung bei dem Kind vermieden werden. Dies fällt insbesondere bei der zweimaligen Verwendung des Testheftes I (Subtests Gebärden- und Lautsprache) ins Gewicht. Denn die Testperson soll ja ihre Wortschatzkenntnis der angebotenen Sprachmodalität unter Beweis stellen und nicht die Wiedererkennung der Zeichnungen. Zum zweiten sollte die Motivation der Testpersonen auch über die Anwendung mehrerer Subtests erhalten bleiben und nicht durch eine schnelle Folge von Testsituationen beeinträchtigt werden. Pro Subtest dauert der Test etwa acht bis zehn Minuten, wobei keine Zeitbegrenzung vorgegeben wird.

Das Testheft I mit vier Abbildungen auf jeder Seite kommt in den Subtests Gebärden- und Lautsprache zum Einsatz. Um einen Wiedererkennungseffekt der Zeichnungen auszuschließen, sollen diese beiden Subtests in der Anwendung nicht aufeinander folgen, sondern der Subtest Schriftsprache mit dem Testheft II dazwischen eingesetzt werden. Zu Anfang soll mit dem Gebärdensprachtest gearbeitet werden, da dieser spielerischen Charakter hat und somit positiv motiviert. Zudem wurde angenommen, dass hier für Gehörlose auf Grund ihres relativ leichten und positiv motivierten Zugangs zur Gebärdensprache die geringsten Schwierigkeiten auftreten würden, was ebenfalls die 'Eisbrecherfunktion' des Testeinstiegs unterstützt. Die Anwendung der Subtests in der Reihenfolge Gebärden-, Schrift- und Lautsprache entspricht der hypothetischen Annahme einer Steigerung des Schwierigkeitsgrades.

Testleiter und Testperson setzen sich wahlweise einander gegenüber oder über Eck an einen Tisch. Dem Kind liegt das Testheft vor, dem Testleiter die Ergebnisprotokolle und ein Stift. Die Protokolle sollen vom Testleiter so ausgefüllt werden, dass das Kind es nicht sieht und verfolgen kann. Das Kind kann sich dadurch ablenken oder verunsichern lassen. Es empfiehlt sich, eine Unterlage bereit zu legen, auf der geschrieben werden kann, sodass die Protokoll-Bögen nicht auf dem Tisch liegen und die Testperson sie nicht sieht.

Die Wahl des Raumes, in dem getestet wird, soll sich danach richten, Ablenkungen so weit wie möglich auszuschließen. Liegen die Fenster eines Raumes beispielsweise zu einem Schulhof oder einem Fußweg gewandt oder befinden sich in dem Raum attraktives Spielzeug, Bilder o. ä., kann die Aufmerksamkeit der Testperson leicht abgelenkt und damit beeinträchtigt werden. Gut geeignet ist ein möglichst neutral eingerichteter, wenn möglich etwas abgelegener Raum, in dem keine äußeren Bedingungen es der Testperson erschweren, seine Konzentration ausschließlich auf das Testgeschehen zu richten. (Sorgsamkeit bei der Auswahl des Raumes stellt sich als besonders bedeutsam heraus bei Kindern, die ohnehin Konzentrationsschwierigkeiten mitbringen und denen ablenkende Impulse im Raum es zusätzlich erschweren, mit der Aufmerksamkeit beim Testgeschehen zu bleiben. Auch für lustlose Kinder mit mangelnder Motivation bieten äußere Reize eine willkommene Ablenkung, die mit der richtigen Wahl der Umgebung, in der der Test stattfindet, von vornherein im Interesse eines objektiven Ergebnisses ausgeschlossen werden müssen.) Die Lichtverhältnisse schließlich sollen derart beschaffen sein, dass Testleiter und Testheft für das Kind einwandfrei sichtbar sind, sodass unter diesem Aspekt keine Beeinflussung des Ergebnisses auftreten kann. Auch für den Testleiter ist es wichtig, mögliche Gebärden oder Mundbilder der Testperson genau erkennen zu können.

Ein weiterer Aspekt, der reflektiert werden muss, ist die Tageszeit, zu der der Test durchgeführt wird. Hier wird in den meisten Fällen der Testpraxis kaum von der 'Wahl' der Zeit gesprochen werden können, da sowohl Testleiter als auch Kind, je nach institutionellem Rahmen, in dem der Test stattfindet, eingebunden sind in Schulalltag, Raumverfügbarkeit, Freizeit- und sonstige Termine, nach denen sich der Testtermin als außerordentlicher Termin richten muss. Es sollte dennoch darauf geachtet werden, die Testpersonen nicht aus Situationen heraus zu holen, denen ihr besonderes Interesse gilt, denn dies wird sich mit einiger Wahrscheinlichkeit negativ auf Aufmerksamkeit und Motivation auswirken.

Das Umblättern der Seiten des Testheftes kann – je nachdem, wie es die Situation ergibt – vom Testleiter oder von der Testperson selbst ausgeführt werden. Beide Varianten haben Vor- und Nachteile: Blättert das Kind selbst, so ist es in Form dieser Bewegung am Testgeschehen beteiligt, sozusagen aktiv. Das reine Schauen, Überlegen, Tippen und dabei still Sitzen kann – je nach Temperament des Kindes und Gesamtsituation – u. U. langweilig sein

und Ungeduld fördern. Ein weiterer Vorteil für den Testleiter liegt darin, die Hände frei zu haben zum Schreiben und nicht mit dem Stift in der Hand blättern zu müssen. Die Erfahrung der Testleiterinnen zeigt allerdings, dass es notwendig sein kann, selbst zu blättern, denn so wird Seite für Seite aufgeschlagen, während die Kinder u. U. zu weit vorblättern und abgelenkt werden, indem sie sich die Bilder ansehen, die erst später an der Reihe sind. Dies kann aus motorischer Ungeschicklichkeit geschehen oder weil es dem Kind als willkommene Ablenkung dient.

Schweift ein Kind hin und wieder mit seiner Aufmerksamkeit ab, stellt Fragen oder macht Bemerkungen über die Bilder, bleibt aber im Prinzip bei der Sache, sollte der Testleiter dies geschehen lassen, ohne näher darauf einzugehen, und dann fortfahren. Sollte ein Kind sehr lange bei einzelnen Aufgaben verweilen und sich schwer dabei tun, eine Entscheidung zu treffen, kann als indirekte Aufforderung zum Weitermachen das Item noch einmal wiederholt werden. Nimmt sich das Kind jedoch mehrmals zu viel Zeit, muss es zur Entscheidung gedrängt werden. Dies ist besonders in Hinblick darauf wichtig, dass die Dauer der Konzentration Grenzen hat, die durch Unterbrechungen nicht erweitert werden. Ein hinausgezögerter Testablauf wird insgesamt länger, was den letzten Abschnitt für das Kind zusätzlich erschwert.

4.2 Das Testprotokoll

Entsprechend der neun Subtestteile stellt der Perlesko neun Protokollbögen sowie ein Deckblatt (s. Anhang 1: B 3.5.7, Tab. 42a-j-B 3.5.7) bereit. Auf dem Deckblatt werden Name, Testdatum und –uhrzeit des Subtests, der Geburtstag bzw. das Alter (auf den Monat genau; gerundet) sowie die Klassenstufe eingetragen. Auf jedem folgenden Protokollbogen befinden sich die jeweils 18 Items eines Subtestteils mit den dazugehörigen Distraktoren, in der Wortart Nomen jeweils inklusive der beiden Übungsbeispiele.

Neben der Spalte mit den Items befinden sich in drei Spalten die Distraktoren zu jedem Item. Jeweils hinter den Items bzw. Distraktoren ist ein freies Feld, das bei entsprechender Antwort angekreuzt wird. Zum Zeitpunkt der Testdurchführung sind nur diese schwarz- und fettgedruckten Angaben von Bedeutung.

Die Summe der richtigen Antworten wird nach dem Testdurchgang in einem grau umrandeten Feld unterhalb der Itemspalte eingetragen. Anhand dieser Angabe lassen sich aus den Normierungstabellen (s. D 5.1.1, Tab. 103a/b/c-D 5.1.1 oder Tab. 104a/b/c-D 5.1.1 und/oder D 5.1.2, Tab. 105a/b/c-D 5.1.2) Prozentränge für die Leistung der Testperson ablesen.

Unter jedem Distraktor ist in grau dessen Art der Ähnlichkeit zum Item angegeben (s. D 5.2). Im Falle der Wahl eines Distraktors kann der Anwender im Anschluss an den Testdurchgang in vier (Subtest Gebärden- und Lautsprache) bzw. fünf (Subtest Schriftsprache) zusätzlichen grauen Spalten auf der rechten Seite des Testprotokolls die Art der Distraktoren ankreuzen. Das gibt ihm die Möglichkeit, die Anzahl der gewählten Distraktoren entsprechend ihrer Ähnlichkeiten zum Item getrennt auszuzählen und die Summe am Ende der Spalten einzutragen. Diese Information wird für das Ablesen von Werten aus den Tabellen zur Distraktoren-Normierung (s. D 5.2, Tab. 107a-i-D 5.2) benötigt.

4.3 Arbeitsanweisung

Sind die äußeren Umstände der Testdurchführung geklärt bzw. bereitgestellt, beginnt der Test mit einer Instruktion seitens des Testleiters an die Testperson. Ziel dieser Anweisung ist es, der Testperson vor Beginn der Testdurchführung Informationen über den Test bzw. über die Beantwortung der Aufgaben zu geben. In dieser Anweisung sind auch Übungsbeispiele enthalten, durch deren Lösung die Testperson zeigen kann, dass die Instruktion verstanden wurde und befolgt werden kann.

Eine kurze, eindeutig formulierte, methodische Erklärung wird jedem einzelnen Subtest des Wortschatztests vorangestellt. Sie wird beim ersten Testdurchgang etwas ausführlicher ausfallen als beim zweiten oder dritten, denn die Methode ändert sich prinzipiell nicht und das Kind erinnert sich i. d. R. an den vorangegangenen Testdurchlauf. Den Testaufgaben gehen zwei Übungsbeispiele voran, anhand derer das Kind die Vorgehensweise erproben und der Testleiter feststellen kann, ob das Prinzip verstanden wurde. Bevor jedoch die erste Seite aufgeschlagen wird und der Blick des Kindes auf die vier Zeichnungen bzw. vier Wörter und ein Bild eines der Übungsbeispiele gezogen werden kann, erfolgt die Erklärung wie folgt (hier beispielhaft für den Subtest Gebärdensprache):

"Hier hast du ein Heft, in dem wirst du auf jeder Seite vier Zeichnungen sehen. Ich werde eine Gebärde machen, du musst sie dir gut anschauen und überlegen, welches Bild dazu passt. Zuerst machen wir ein Beispiel zum Üben." (Testleiter schlägt die erste Seite des Testheftes auf. Testleiter gebärdet 'Hund'.)

Zeigt das Kind auf das richtige Bild, kann direkt und ohne weitere Anweisung das zweite Übungsbeispiel folgen. Sollte jedoch deutlich werden, dass der Testperson die Vorgehensweise noch nicht ganz klar ist, muss die Erklärung noch einmal wiederholt und am zweiten Übungsbeispiel umgesetzt werden.

Die Erklärung sollte um folgende Anmerkungen ergänzt werden: "Es geht nicht darum, besonders schnell darin zu sein, die Aufgaben zu lösen. Du kannst dir die Zeit lassen, die du brauchst. Hast du alles verstanden?"

Gibt das Kind an, alles verstanden zu haben, kann begonnen werden. Es sollte vorher allerdings noch ausdrücklich auf Folgendes hingewiesen werden: "Wenn du noch Fragen hast, musst du sie jetzt stellen, bevor wir angefangen haben. Während des Testens kann ich sie dir nicht beantworten. Hast du noch Fragen?"

Verneint das Kind, kann begonnen werden mit den Worten: "O.K. - Fangen wir an!"

Der Testleiter trägt die Anfangszeit in den ersten Protokollbogen ein und der Test beginnt: Das Kind wählt nun nach dem Prinzip der Antwort-Auswahl-Aufgabe eine von vier Zeichnungen aus und dies 54 mal. Der Testleiter kann bei Bedarf die Itemvorgabe zweimal wiederholen. Danach muss sich die Testperson für eine der vier Möglichkeiten entscheiden.

Sollte die Testperson zwischen zwei oder mehr Antwortmöglichkeiten schwanken, oder sich noch einmal umentscheiden so sollte der Testleiter auf Folgendes hinweisen: "Nur eine Antwort ist richtig, du musst dich entscheiden." Der Testperson soll die Möglichkeit gegeben werden, eine vorschnelle Wahl zu revidieren.

Der Übergang der Wortarten geschieht im Prinzip fließend und wird der Testperson nicht extra angezeigt. (Selbstverständlich kann das Kind den Wechsel theoretisch selbstständig feststellen. Einen Hinweis bekommt es möglicherweise dadurch, dass der Testleiter die Auswertungsprotokolle – ein Bogen pro Wortart – nach jeweils 18 Items umblättert.)

Kommt nach 14 Tagen der Subtest Schriftsprache zur Anwendung, so muss die Erklärung zum Verfahren erfahrungsgemäß lediglich teilweise wiederholt werden. Der Testleiter beginnt nun mit der Frage: „Erinnerst du dich noch an das letzte Mal?“

Wenn das Kind verneinen sollte, kann die Erklärung vom ersten Mal mit den spezifischen Modifikationen für den Subtest Schriftsprache wiederholt werden. Bejaht die Testperson, wird die Einführung fortgesetzt mit den Worten: „Dieses Mal geht es ein bisschen anders.“ Die erste Seite des Testhefts mit dem ersten Übungsbeispiel wird aufgeschlagen. „Hier siehst du ein Bild (Testleiter deutet auf die linke Seite) und hier vier Wörter. Du musst überlegen, welches Wort passt zu dem Bild.“

In der Regel deutet das Kind nun auf die von ihm angenommene Lösung. Andernfalls fordert der Testleiter dazu auf, das erste Übungsitem zu bearbeiten. Wird die richtige Lösung ge-

zeigt, kann davon ausgegangen werden, dass das Verfahren verstanden wurde und das zweite Übungsitem kann kommentarlos aufgeschlagen werden. Sollte jedoch deutlich werden, dass der Testperson die Vorgehensweise noch nicht klar ist, muss die Erklärung noch einmal wiederholt werden.

Die Erklärung sollte auch beim zweiten Testdurchlauf nochmals um folgende Anmerkungen ergänzt werden: "Es geht auch heute nicht darum, besonders schnell darin zu sein, die Aufgaben zu lösen. Du kannst dir die Zeit lassen, die du brauchst. Hast du alles verstanden?"

Bestehen keine weiteren Verständnisschwierigkeiten, wird begonnen mit den Worten: „O.K., fangen wir an.“

Bei Anwendung des dritten Subtests Lautsprache kann die Kenntnis des Verfahrens sicher vorausgesetzt werden. Es wird ausschließlich auf die subtestspezifische Vorgehensweise hingewiesen mit den Worten: „Ich sage ein Wort, du musst genau ablesen und auswählen, welches Bild dazu passt.“

Der Testleiter kreuzt nach jeder Antwort die richtige Antwort oder den gewählten Distraktor an. Das Auflegen der Folienschablone zum Ablesen der Distraktorenähnlichkeiten, die Auszählung der richtigen Antworten bzw. der Distraktoren sowie der Eintrag der Summenwerte in die dafür vorgesehen Kästchen am unteren Rand des Protokollbogens erfolgen nach Beendigung des Testdurchgangs anhand der eingetragenen Kreuze.

5. Auswertungsanleitung

5.1 Quantitative Auswertung

5.1.1 Reguläre Normierungstabellen

In den folgenden Normierungstabellen werden Häufigkeitswerte (f) der Eichstichprobe, lineare Standardwerte (Z) und Prozentrangwerte (PR) für jeden Subtestteil (Tab. 103a-c-D 5.1.1) gesondert sowie für die drei Subtests GS, SS und LS als Zusammenfassung der drei Wortarten innerhalb einer Modalität (Tab. 104a-c-D 5.1.1) aufgeführt. Entsprechend dem erreichten Rohwert ($RW = \text{Anzahl der richtig beantworteten Aufgaben je Subtestteil}$) kann damit abgelesen werden, wieviel Prozent der Eichstichprobe die Testperson mit ihrer Leistung übertrifft (PR).

Da der Mittelwert (a.M.) und die Standardabweichung (s) innerhalb der Skalen variieren, ändert sich auch die Höhe des Erwartungsrohwerths. Für den Z-Wert liegt er dagegen immer bei 100 und für Prozentränge bei 50. Jeweils eine Standardabweichung ober- und unterhalb des Mittelwerts liegen die Leistungen von etwa 66% aller Testpersonen, zwei Standardabweichungen ober- und unterhalb des Mittelwertes die Leistungen von 95% aller Testpersonen.

Tab. 103a-D 5.1.1: Normierung Subtestteile Gebärdensprache N/V/A

RW	GSN (a.M. = 13,1; s= 2,87)			GSV (a.M. = 15,2; s = 2,21)			GSA (a.M. = 12,4; s = 2,89)		
	f	Z	PR	f	Z	PR	f	Z	PR
0	0	-	0	0	-	0	0	-	0
1	0	-	0	0	-	0	0	-	0
2	0	-	0	0	-	0	0	-	0
3	0	-	0	0	-	0	0	-	0
4	0	-	0	0	-	0	0	-	0
5	0	-	0	0	-	0	0	76	1
6	3	76	1	0	-	0	0	79	2
7	1	79	2	1	-	0	6	83	4
8	0	83	5	1	-	0	6	86	8
9	8	86	8	1	-	0	7	90	16
10	8	89	14	1	77	1	12	93	24
11	15	93	24	5	82	4	13	96	34
12	11	96	35	4	86	8	14	100	50
13	11	100	50	5	91	18	10	104	66
14	16	104	66	12	95	31	12	107	76
15	12	107	76	21	100	50	10	110	84
16	8	111	87	23	105	70	13	114	92
17	13	114	92	26	109	82	5	117	96
18	4	117	96	10	114	92	2	121	98

Tab. 103b-D 5.1.1: Normierung Subtestteile Schriftsprache N/V/A

RW	SSN (a.M. = 11,9; s = 4,21)			SSV (a.M. = 9,8; s = 4,54)			SSA (a.M. = 10,1; s = 4,69)		
	f	Z	PR	f	Z	PR	f	Z	PR
0	0	-	0	1	78	1	0	79	2
1	0	-	0	1	80	2	0	81	3
2	0	76	1	5	82	3	2	83	4
3	0	79	2	3	85	7	4	85	7
4	4	81	3	6	87	10	6	87	10
5	3	83	4	7	89	13	8	89	13
6	6	86	8	9	91	18	11	91	18
7	6	88	12	6	93	24	8	94	27
8	9	90	16	4	95	31	9	96	34
9	9	93	24	6	98	42	5	98	42
10	5	95	31	5	100	50	6	100	50
11	11	98	42	9	102	58	8	102	58
12	7	100	50	9	104	66	8	104	66
13	5	102	58	10	107	76	2	106	73
14	7	105	69	7	109	82	3	109	82
15	8	107	76	7	111	86	6	111	86
16	8	110	84	10	113	90	5	113	90
17	11	112	88	2	115	93	13	115	93
18	11	114	92	1	118	96	4	117	96

Tab. 103c-D 5.1.1: Normierung Subtestteile Lautsprache N/V/A

RW	LSN (a.M. = 12,1; s = 3,86)			LSV (a.M. = 11,6; s = 3,66)			LSA (a.M. = 10,5; s = 4,46)		
	f	Z	PR	f	Z	PR	f	Z	PR
0	0	-	0	0	-	0	1	75	1
1	0	-	0	1	-	0	0	78	1
2	1	-	0	0	-	0	1	80	2
3	2	77	1	1	75	1	4	82	3
4	3	79	2	2	78	1	4	84	5
5	2	82	3	0	81	3	5	86	8
6	2	84	5	5	84	5	9	89	13
7	8	87	10	8	86	8	5	91	18
8	4	89	13	7	89	13	12	93	24
9	2	92	21	7	92	21	5	95	31
10	8	95	31	8	94	27	10	98	42
11	5	97	38	12	97	38	5	100	50
12	11	100	50	10	100	50	3	102	58
13	12	103	62	5	103	62	9	105	69
14	16	105	69	11	106	73	7	107	76
15	11	108	79	12	108	79	6	109	82
16	9	110	84	12	111	86	10	111	86
17	5	113	90	3	114	92	8	114	92
18	5	116	95	2	116	95	2	116	95

Die Ergebnisse jeder einzelnen Testperson in den neun Subtestteilen bilden ein individuelles Leistungsprofil. Starke Schwankungen in dem Profil einer Testperson sollten untersucht werden. Es muss dabei unterschieden werden zwischen Auffälligkeiten, die nur in einem einzelnen Subtestteil auftreten und solchen, die in mehreren Subtestteilen gleichzeitig deutlich werden. Besonderheiten in einzelnen Subtestteilen sollten nicht überbewertet werden, da sie durch zufällige Begebenheiten wie Konzentration oder Störungen während der Testsituation entstehen können und somit mehrdeutig sind. Gehäufte Auffälligkeiten, denen die gleiche Ursache zugrunde liegt, können dagegen möglicherweise auf besondere Probleme oder Fähigkeiten in bestimmten Modalitäten oder Wortarten hindeuten. Es sollte demnach im individuellen Profil untersucht werden, ob eine Testperson bei einer Teilgruppe von Subtestteilen, deren richtige Beantwortung die gleiche Fähigkeit voraussetzt, im Vergleich zu ihrer sonstigen Leistung besonders gut oder besonders schlecht abgeschnitten hat. Diese betreffenden Subtestteile sollten dann zusammenfassend unter Berücksichtigung ihrer Gemeinsamkeiten interpretiert werden.

Folgende Subtestteilkombinationen bilden für diagnostische Zwecke sinnvolle Zusammenfassungen: Die drei Subtestteile einer Modalität (GSN, GSV, GSA), (SSN, SSV, SSA), (LSN, LSV, LSA) sowie die drei Subtestteile einer Wortart (GSN, SSN, LSN), (GSV, SSV, LSV), (GSA, SSA, LSA).

Tab. 104a-D 5.1.1: Normierung Gesamtwert Subtest GS

RW	f	Z	PR		RW	f	Z	PR
1	-	-	0		28	1	80	2
2	-	-	0		29	-	82	3
3	-	-	0		30	1	84	5
4	-	-	0		31	2	85	7
5	-	-	0		32	2	87	10
6	-	-	0		33	4	88	12
7	-	-	0		34	4	90	16
8	-	-	0		35	4	91	18
9	-	-	0		36	2	93	24
10	-	-	0		37	6	94	27
11	-	-	0		38	5	96	34
12	-	-	0		39	11	97	38
13	-	-	0		40	8	99	46
14	-	-	0		41	6	99	46
15	-	-	0		42	2	102	58
16	-	-	0		43	10	104	66
17	-	-	0		44	2	105	69
18	-	-	0		45	9	107	76
19	-	-	0		46	4	108	79
20	-	-	0		47	4	110	84
21	-	-	0		48	3	111	86
22	-	-	0		49	4	113	90
23	1	73	0		50	4	114	92
24	1	74	0		51	4	116	95
25	-	76	1		52	2	117	96
26	1	77	1		53	1	119	97
27	2	79	2		54	-	120	98

Tab. 104b-D 5.1.1: Normierung Gesamtwert Subtest SS

RW	f	Z	PR		RW	f	Z	PR
1	-	76	1		28	-	97	38
2	-	76	1		29	3	98	42
3	-	77	1		30	3	98	42
4	-	78	1		31	-	99	46
5	-	79	2		32	2	100	50
6	-	80	2		33	2	101	54
7	-	80	2		34	3	102	58
8	-	81	3		35	1	102	58
9	1	82	3		36	1	103	62
10	-	83	4		37	1	104	66
11	-	83	4		38	3	105	69
12	2	84	5		39	4	106	73
13	3	85	7		40	6	106	73
14	6	86	8		41	1	107	76
15	1	87	10		42	2	108	79
16	2	87	10		43	1	109	82
17	1	88	12		44	1	109	82
18	3	89	13		45	4	110	84
19	1	90	16		46	1	111	86
20	4	91	18		47	2	112	88
21	3	91	18		48	3	113	90
22	2	92	21		49	6	113	90
23	8	93	24		50	1	114	92
24	4	94	27		51	5	115	93
25	2	94	27		52	4	116	95
26	-	95	31		53	-	117	96
27	5	96	34		54	-	117	96

Tab. 104c-D 5.1.1: Normierung Gesamtwert Subtest LS

RW	f	Z	PR		RW	f	Z	PR
1	-	70	0		28	4	94	27
2	-	71	0		29	3	95	31
3	-	72	0		30	1	96	34
4	-	73	0		31	4	97	38
5	-	74	0		32	2	98	42
6	-	75	1		33	2	99	46
7	1	75	1		34	6	100	50
8	-	76	1		35	4	101	54
9	-	77	1		36	3	102	58
10	1	78	1		37	3	103	62
11	-	79	2		38	2	104	66
12	1	80	2		39	2	104	66
13	-	81	3		40	5	105	69
14	-	82	3		41	4	106	73
15	1	83	4		42	4	107	76
16	1	84	5		43	3	108	79
17	3	85	7		44	4	109	82
18	4	85	7		45	4	110	84
19	4	86	8		46	1	111	86
20	1	87	10		47	4	112	88
21	1	88	12		48	2	113	90
22	1	89	13		49	4	114	92
23	3	90	16		50	1	114	92
24	2	91	18		51	2	115	93
25	3	92	21		52	1	116	95
26	2	93	24		53	2	117	96
27	-	94	27		54	-	118	96

5.1.2 Normierungstabellen aufgrund modifizierter Fehlerbewertung

Zusätzlich zur regulären Normierung bietet der Perlesko eine Normierung aufgrund einer modifizierten Fehlerbewertung an. Die Ergebnisse der Hauptuntersuchung haben gezeigt, dass ein Teil der Testpersonen zwar einen Begriff vom Testitem hat, diese Kenntnis jedoch durch Schwierigkeiten in der Modalität nicht unter Beweis stellen konnte. Demzufolge wurde in einer neuen Fehlerbewertung genau dann ein Punkt gegeben, wenn die Testperson das Item in Gebärden- oder in Schrift- oder in Lautsprache richtig beantworten konnte. Als falsch wurde die Aufgabe nur dann bewertet, wenn die Testperson sie durch alle drei Modalitäten hinweg falsch beantwortet hat. Daraus ergeben sich neue Rohwerte und demzufolge eine neue Normierung. Diese wurde über Flächentransformation vorgenommen, da keine Normalverteilung der Daten vorlag. Es ergeben sich nur drei Normierungstabellen, nämlich eine für Nomen, eine für Verben und eine für Adjektive, in denen sowohl Prozentränge als auch Z-Werte dargestellt sind. Auch hier kann entsprechend einem modifiziert errechneten Rohwert ein Prozentrang abgelesen werden, der den Anteil der Eichstichprobe beschreibt, den die Testperson mit ihrer Leistung übertroffen hat:

Tab. 105a-D 5.1.2: Normierung aufgrund modifizierter Fehlerbewertung				
Nomen				
RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	0	0	0	-
7	0	0	0	-
8	0	0	0	-
9	1	1	0,4	73
10	0	1	0,9	74
11	2	3	1,8	78
12	4	7	4,5	84
13	6	13	8,9	86
14	8	21	15,2	90
15	8	29	22,3	92
16	14	43	32,1	95
17	24	67	49,1	100
18	45	112	79,9	108

Tab. 105b-D 5.1.2: Normierung aufgrund modifizierter Fehlerbewertung				
Verben				
RW	f	f _{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	0	0	0	-
7	0	0	0	-
8	0	0	0	-
9	0	0	0	-
10	2	2	0,9	74
11	1	3	2,2	79
12	0	3	2,7	80
13	2	5	3,6	82
14	7	12	7,6	86
15	8	20	14,3	89
16	26	46	29,5	95
17	27	73	53,1	101
18	39	112	82,6	109

Tab. 105c-D 5.1.2: Normierung aufgrund modifizierter Fehlerbewertung				
Adjektive				
RW	f	f _{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	0	0	0	-
7	0	0	0	-
8	0	0	0	-
9	2	2	0,9	74
10	4	6	3,6	82
11	6	12	8,0	86
12	3	15	12,1	88
13	6	21	16,1	90
14	12	33	24,1	93
15	13	46	35,3	96
16	14	60	47,3	99
17	19	79	62,1	103
18	33	112	98,7	121

5.2 Qualitative Auswertung

Neben der normierten Rohwertskala stellt der Wortschatztest ergänzende qualitative Interpretationsmöglichkeiten der gemessenen Leistung bereit. Dadurch soll zusätzlich zu der ausschließlichen Betrachtung und Auszählung der richtigen Itemantworten auch der Wahl der Distraktoren bei falsch beantworteten Aufgaben Bedeutung zukommen. Da die Distraktoren der Items qualitativ abgestuft konzipiert sind, indem sie jeweils andere Ähnlichkeit zum Item aufweisen, soll ergänzend zur Auszählung des Gesamtpunktwertes die Art der Fehler untersucht werden. Es wird davon ausgegangen, dass damit Aufschluss über die Herangehensweise einer Testperson an sprachliche Anforderungen gegeben sowie Stärken oder Schwächen in einzelnen Modalitäten aufgezeigt werden können. Durch die differenzierte Betrachtung eines individuellen Distraktorenprofils einer Testperson soll dem Anspruch der Differentialdiagnose Genüge geleistet werden. Nur das Gesamtergebnis der quantitativen und der qualitativen Fehlerbewertung kann Grundlage für die differenzierte Bewertung einer Leistung sein. Angaben über die Homogenität oder Heterogenität des Distraktorenprofils können die Bedeutung des quantitativen Testergebnisses im Einzelfall relativieren.

Die Distraktoren wurden nach verschiedenen Prinzipien, qualitativ abgestuft, konstruiert. In allen Subtestteilen existieren vier Ablenkerarten: Für jede Aufgabe wurde mindestens ein Inhaltsablenker (I) konstruiert. In einigen Fällen weist die Wahl dieser Ablenker auf eine elaborierte Sprachkenntnis (HRP) hin, da dieser Distraktor eine beinahe ebenso plausible Antwort wie das Item selber darstellt. Nicht alle Ablenker inhaltlicher (I) Ähnlichkeit entsprechen dieser Kategorie (HRP), sondern nur einige von ihnen; alle Ablenker dieser Art (HRP) sind jedoch gleichzeitig Ablenker mit einfacher inhaltlicher (I) Ähnlichkeit zum Item. Zusätzlich stellt jede Aufgabe (mit drei konzeptionell begründeten Ausnahmen) mindestens einen Ablenker gebärdenähnlicher (G) und mundbildähnlicher (M) Art bereit. Darüber hinaus wurde eine Kategorie der modalitätenentsprechenden Ablenkerart entwickelt. Diese Distraktorenart stellt in den Subtests Gebärden- bzw. Lautsprache ein Pendant zu gebärden- (G) bzw. mundbildähnlichen (M) Ablenkern dar, bildet in der Modalität Schriftsprache jedoch eine eigenständige Gruppe, die sich aus schriftbildähnlichen (S) Ablenkern zusammensetzt. Mundbild- (M) und schriftbildähnliche (S) Ablenker überschneiden sich häufig, aber nicht immer.

Die Verteilung des Items und der Distraktoren innerhalb der vier möglichen Positionen auf einer Seite unterliegt keiner Regelmäßigkeit, nach der sich ein aufmerksames Kind womöglich richten könnte, sondern dem Zufallsprinzip. Da die Distraktorenarten konzeptionell bedingt im Vergleich zueinander und in den drei Subtests unterschiedlich attraktive und plausible Alternativen zur korrekten Antwort darstellen und sich in ihrer relativen Anzahl unterscheiden, weisen sie keine gleich große Wahlmöglichkeit auf. Die einzelnen Fehlerarten, die

sich in fünf Klassen zusammenfassen lassen, sowie ihre Verteilung auf die einzelnen Aufgaben lassen sich in den folgenden Tabellen ablesen:

Tab. 106a-D 5.2: Distraktorenverteilung Nomen							
Nr.	Item	Distraktor	HRP	G	M	I	S
1	Ei	Huhn	x (GS, LS)			x	
		Kartoffel		x		x	
		Eis			x		x
2	Bett	Decke	x	x		x	
		Brett			x		x
		Sofa				x	
3	Bart	Haare	x			x	
		Kinn	x	x		x	
		Post		x	x		
4	Zahn	Mund	x (GS, LS)	x		x	
		Hahn			x		x
		Zange			x	x	x
5	Angst	Ärger		x	x	x	
		Trauer		x		x	
		Freude				x	
6	Welt	Planet	x	x		x	
		Ball		x			
		Vase		x	x		
7	Stein	Felsen	x (GS, LS)			x	
		Kaffee		x			
		Stange			x		x
8	Rose	Sonnenblume		x		x	
		Tulpe		x		x	
		Rasen			x	x	x
9	Zahl	Regen		x			
		Buchstabe				x	
		Pfahl			x		x
10	Leute	Paar	x (GS, LS)			x	
		Mensch		x		x	
		Herde			x	x	
11	Schuh	Stiefel	x (GS, LS)	x		x	
		Strumpf		x		x	
		Uhr			x		
12	Arbeit	Eimer	x (GS, LS)		x	x	
		Stein		x			
		Ampel			x		
13	Kopf	Zopf	x (GS, LS)		x	x	x
		Junge		x			
		Topf			x		x
14	Geld	Gold	x		x	x	x
		Salz		x			
		Feld			x		x

Nr.	Item	Distraktor	HRP	G	M	I	S
15	Holz	Baum	x			x	
		Brot		x	x		
		Gold			x		
16	Mutter	Mädchen	x			x	
		Bauer		x			
		Motte			x		x
17	Straße	Weg	x	x		x	
		Strand			x		x
		Strauß			x		x
18	Garten	Baum	x			x	
		Geschenk		x			
		Arbeit			x		

Tab. 106b-D 5.2: Distraktorenverteilung Verben							
Nr.	Item	Distraktor	HRP	G	M	I	S
1	schreiben	malen		x		x	
		reiben		x	x		x
		schlafen			x		x
2	fliegen	springen	x			x	
		winken		x			
		fragen			x		x
3	stinken	riechen	x	x		x	
		staunen			x		x
		winken			x		x
4	springen	sitzen		x			
		klettern				x	
		schimpfen			x		
5	weinen	reiben	x (GS, LS)	x		x	x
		schwitzen				x	
		wiegen			x		x
6	lesen	sitzen	x (GS, LS)			x	
		halten	x (SS)	x		x (SS)	
		blasen			x		x
7	waschen	wischen	x	x	x	x	x
		schwimmen		x			
		warten			x		x
8	werfen	fangen		x		x	
		strecken				x	
		wärmen			x		x
9	schieben	schließen	x (GS, LS)	x	x	x	x
		heben				x	
		ziehen				x	
10	verstecken	schlafen		x			
		erschrecken			x	x	x
		stecken			x		x
11	schlafen	schlagen		x	x		x
		streicheln		x			
		liegen	x (SS)			x	

Nr.	Item	Distraktor	HRP	G	M	I	S
12	brennen	löschen	x			x	
		rechnen		x	x		x
		rennen			x		x
13	gehen	stehen		x	x	x	x
		rennen				x	
		sehen			x		x
14	singen	sagen		x	x	x	x
		warnen		x			
		flöten				x	
15	warten	sitzen	x (GS, LS)			x	
		wandern		x	x		x
		braten			x		x
16	gucken	schminken	x (GS, LS)	x		x	
		denken		x			
		ducken			x		x
17	hinsetzen	sitzen	x	x	x	x	x
		hinlegen			x	x	x
		schwitzen			x		
18	denken	träumen	x	x		x	
		warnen		x			
		schenken			x		x

Tab. 106c-D 5.2: Distraktorenverteilung Adjektive

Nr.	Item	Distraktor	HRP	G	M	I	S
1	glatt	sauber		x			
		lockig				x	
		platt			x		x
2	spitz	struppig		x			
		schief	x (SS)			x	x
		spannend			x		
3	hellblau	dunkelblau		x	x	x	x
		hellgrau		x	x	x	x
		hellgrün		x	x	x	x
4	schmutzig	sauber				x	
		zerknittert				x	
		zerrissen				x	
5	orange	rosa		x	x	x	
		gelb				x	
		rot				x	
6	viereckig	dreieckig		x	x	x	x
		fünfeckig		x	x	x	x
		sechseckig		x	x	x	x
7	warm	heiß	x	x		x	
		kalt				x	
		weich			x		x
8	dünn	eng	x (GS, LS)	x	x	x	
		dick				x	
		dumm			x		x

Nr.	Item	Distraktor	HRP	G	M	I	S
9	klein	klug		x	x		
		dünn				x	
		fein			x		x
10	gestreift	kariert		x		x	
		geblümt				x	
		gepunktet				x	
11	nass	trocken		x		x	
		blass			x		x
		nackt			x		x
12	schief	steil	x (GS, LS)	x		x	
		gerade				x	
		stark			x		
13	verletzt	scharf	x (GS, LS)	x		x	
		gesund				x	
		verliebt			x		x
14	blond	braun			x	x	
		rot				x	
		schwarz				x	
15	voll	viel	x	x	x	x	x
		groß		x	x	x	
		stolz		x	x		
16	weich	nass		x			
		wach			x	x	x
		warm			x		
17	nah	geschlossen		x			
		fern				x	
		nass			x		x
18	schwer	stark	x (GS, LS)	x		x	
		leer			x		
		schwarz			x		x

Für die Erstellung eines individuellen Distraktorenprofils werden Normierungen der Ablenker-Wahlhäufigkeiten benötigt, um die Fehlerwahl einer Testperson mit denen vergleichbarer Testpersonen in Beziehung setzen zu können. Die Berechnung der Prozentränge für einzelne Ablenkerarten erfolgte dabei über Flächentransformation, da Normalverteilung nicht nachgewiesen wurde. Die Prozentrangwerte müssen in diesem Zusammenhang besonders interpretiert werden. Der Prozentrang einer Testperson in einer Ablenkerart in einem Subtestteil gibt nach Definition den Prozentanteil aller vergleichbaren getesteten Testpersonen an, den diese Person mit ihrer 'Leistung' übertroffen hat. Da es sich bei der gemessenen Größe in diesem Fall aber nicht um eine 'Leistung', sondern um die Anzahl der gemachten Fehler handelt, bedeutet ein höherer Rohwert (viele Fehler) eine schlechtere Leistung als ein niedriger Rohwert (wenige Fehler). Dementsprechend repräsentiert der Prozentrang einer Testperson in einer Ablenkerart in einem Subtestteil den Prozentanteil aller vergleichbaren Testpersonen, der weniger Fehler gemacht hat, als diese Testperson. Ein hoher Prozentrang

bedeutet also, dass viele Testpersonen besser, ein niedriger Prozentrang, dass wenige Testpersonen besser waren als diese Testperson.

Mit Hilfe der Prozenträge dieser Normierungstabellen lässt sich feststellen, ob eine Testperson mit ihrer fehlerhaften Aufgabenbeantwortung im Durchschnitt der Normierungsstichprobe liegt, also durchschnittlich oft, oder auffallend selten oder häufig bestimmte Distraktoren wählt. Dieser Bezug zur Vergleichsstichprobe ist Grundlage für die Untersuchung, ob eine Testperson im Vergleich zu ihren sonstigen Leistungen in einzelnen Subtestteilen auffallend gut oder auffallend schlecht abschneidet. So werden für eine Testperson zusätzlich zu dem Wert aus der Normierungstabelle im Rahmen der klassischen quantitativen Bewertung (D 5.1) vier (Gebärden- und Lautsprache) bzw. fünf (Schriftsprache) Werte aus folgenden Tabellen abgelesen.

In Klammern unter der jeweiligen Ablenkerart ist die Gesamtanzahl der entsprechenden Fehler innerhalb aller Distraktoren des Subtestteils angegeben. Da mehrere Distraktoren eines einzelnen Items die gleiche Ablenkerart aufweisen können, also z. B. das Item ‘Angst’ zwei gebärdenähnliche und drei inhaltsähnliche Ablenker besitzt, kann die Angabe in Klammern die Aufgabenanzahl von 18 Items pro Subtestteil überschreiten. Trotzdem kann eine Testperson jede Ablenkerart nur maximal 18-mal wählen, da sie sich pro Aufgabe nur für eine Antwort entscheiden kann. Aus diesem Grund bestehen die folgenden Tabellen aus jeweils 18 und nicht der Anzahl der vorkommenden Ablenkerarten entsprechend vielen Zeilen.

Die Normierung der Distraktorenwahl wurde für jeden Subtestteil gesondert vorgenommen. Die Tabellen zu Gebärden- und Lautsprache enthalten jeweils vier Distraktorenspalten, da in diesen Fällen der Distraktorentyp des modalitätenähnlichen Ablenkers mit dem des gebärden- bzw. mundbildähnlichen Ablenkers übereinstimmt. In der Modalität Schriftsprache wird für die Ablenkerart mit Modalitätenähnlichkeit zum Item eine gesonderte Spalte mit schriftbildähnlichen Distraktoren aufgeführt, die Tabellen zur Schriftsprache weisen aus diesem Grund fünf Distraktorenspalten auf:

Tab. 107a-D 5.2: Distraktoren-Normierung Subtestteil Gebärdensprache

Nomen								
Fehler- anzahl	HRP (16)		G = MOD (24)		M (24)		I (29)	
	f	PR	f	PR	f	PR	f	PR
0	8	4	13	6	15	7	6	3
1	27	20	19	21	35	30	17	14
2	28	45	21	39	20	55	20	30
3	22	68	18	57	19	73	20	49
4	17	85	18	73	11	86	21	67
5	7	96	11	86	5	94	14	83
6	1	100	5	94	5	98	8	93
7	0	100	3	97	0	98	3	98
8	0	100	1	99	0	98	1	100
9	0	100	1	100	0	98	0	100
10	0	100	0	100	0	98	0	100
11	0	100	0	100	0	98	0	100
12	0	100	0	100	0	98	0	100
13	0	100	0	100	0	98	0	100
14	0	100	0	100	0	98	0	100
15	0	100	0	100	0	98	0	100
16	0	100	0	100	0	98	0	100
17	-		0	100	0	98	0	100
18	-		0	100	0	98	0	100

Tab. 107b-D 5.2: Distraktoren-Normierung Subtestteil Gebärdensprache

Verben								
Fehler- anzahl	HRP (11)		G = MOD (23)		M (27)		I (26)	
	f	PR	f	PR	f	PR	f	PR
0	31	15	16	8	32	15	16	8
1	48	50	36	31	58	56	40	33
2	23	83	26	60	12	88	28	64
3	5	95	18	80	5	95	12	82
4	2	99	5	90	2	99	8	91
5	1	100	7	95	0	99	2	96
6	0	100	2	100	1	100	3	98
7	0	100	0	100	0	100	0	98
8	0	100	0	100	0	100	1	100
9	0	100	0	100	0	100	0	100
10	0	100	0	100	0	100	0	100
11	0	100	0	100	0	100	0	100
12	-		0	100	0	100	0	100
13	-		0	100	0	100	0	100
14	-		0	100	0	100	0	100
15	-		0	100	0	100	0	100
16	-		0	100	0	100	0	100
17	-		0	100	0	100	0	100
18			0	100	0	100	0	100

Tab. 107c-D 5.2: Distraktoren-Normierung Subtestteil Gebärdensprache
Adjektive

Fehler- anzahl	HRP (6)		G = MOD (22)		M (26)		I (35)	
	f	PR	f	PR	f	PR	f	PR
0	6	3	3	2	27	13	2	1
1	37	23	11	8	28	38	13	8
2	39	57	18	21	31	65	11	19
3	22	85	17	37	18	87	21	34
4	5	97	9	49	5	97	14	50
5	1	100	21	63	1	100	21	65
6	0	100	16	80	0	100	9	79
7	-		8	90	0	100	10	88
8	-		5	96	0	100	6	95
9	-		2	100	0	100	3	99
10	-		0	100	0	100	0	99
11	-		0	100	0	100	0	99
12	-		0	100	0	100	0	99
13	-		0	100	0	100	0	99
14	-		0	100	0	100	0	99
15	-		0	100	0	100	0	99
16	-		0	100	0	100	0	99
17	-		0	100	0	100	0	99
18			0	100	0	100	0	99

Tab. 107d-D 5.2: Distraktoren-Normierung Subtestteil Schriftsprache Nomen

Fehler- anzahl	HRP (9)		S = MOD (14)		G (24)		M (24)		I (29)	
	f	PR	f	PR	f	PR	f	PR	f	PR
0	25	12	39	18	23	11	34	16	12	6
1	24	34	32	50	21	31	20	40	11	16
2	26	57	18	73	18	49	14	56	13	27
3	23	79	9	85	20	66	13	68	15	40
4	8	93	7	93	15	82	13	80	12	52
5	4	99	4	98	7	92	6	89	15	65
6	0	99	1	100	4	97	7	95	12	77
7	0	99	0	100	2	100	2	99	12	88
8	0	99	0	100	0	100	0	99	5	95
9	0	99	0	100	0	100	1	100	3	99
10	-		0	100	0	100	0	100	0	99
11	-		0	100	0	100	0	100	0	99
12	-		0	100	0	100	0	100	0	99
13	-		0	100	0	100	0	100	0	99
14	-		0	100	0	100	0	100	0	99
15	-		-		0	100	0	100	0	99
16	-		-		0	100	0	100	0	99
17	-		-		0	100	0	100	0	99
18	-		-		0	100	0	100	0	99

Tab. 107e-D 5.2: Distraktoren-Normierung Subtestteil Schriftsprache Verben										
Fehler- anzahl	HRP (8)		S = MOD (26)		G (23)		M (28)		I (28)	
	f	PR	f	PR	f	PR	f	PR	f	PR
0	9	5	9	5	3	2	9	5	1	1
1	38	26	17	17	6	6	15	16	6	4
2	35	60	12	30	21	19	13	29	18	15
3	21	86	18	44	15	35	16	42	11	29
4	5	98	13	58	17	50	8	53	13	40
5	4	99	15	71	15	65	19	66	14	52
6	0	99	12	84	15	79	12	80	11	64
7	0	99	6	92	10	90	8	89	14	75
8	0	99	4	97	5	97	6	96	8	86
9	-		1	99	1	100	0	96	5	92
10	-		1	100	0	100	1	99	5	96
11	-		0	100	0	100	1	100	1	99
12	-		0	100	0	100	0	100	1	100
13	-		0	100	0	100	0	100	0	100
14	-		0	100	0	100	0	100	0	100
15	-		0	100	0	100	0	100	0	100
16	-		0	100	0	100	0	100	0	100
17	-		0	100	0	100	0	100	0	100
18	-		0	100	0	100	0	100	0	100

Tab. 107f-D 5.2: Distraktoren-Normierung Subtestteil Schriftsprache Adjektive										
Fehler- anzahl	HRP (3)		S = MOD (18)		G (22)		M (27)		I (35)	
	f	PR	f	PR	f	PR	f	PR	f	PR
0	73	34	24	12	19	9	21	10	6	3
1	27	81	19	31	13	24	16	27	16	13
2	5	95	17	48	11	35	7	38	5	23
3	3	99	14	63	18	49	7	44	9	30
4	-		11	74	17	65	9	52	7	37
5	-		17	87	9	77	12	62	14	47
6	-		2	96	13	87	13	73	15	60
7	-		3	98	5	95	13	85	9	71
8	-		1	100	2	99	4	93	10	80
9	-		0	100	1	100	3	96	11	90
10	-		0	100	0	100	3	99	4	97
11	-		0	100	0	100	1	100	1	99
12	-		0	100	0	100	0	100	1	100
13	-		0	100	0	100	0	100	0	100
14	-		0	100	0	100	0	100	0	100
15	-		0	100	0	100	0	100	0	100
16	-		0	100	0	100	0	100	0	100
17	-		0	100	0	100	0	100	0	100
18	-		0	100	0	100	0	100	0	100

Tab. 107g-D 5.2: Distraktoren-Normierung Subtestteil Lautsprache Nomen								
Fehler- anzahl	HRP (16)		G (24)		M = MOD (24)		I (29)	
	f	PR	f	PR	f	PR	f	PR
0	20	10	18	9	10	5	11	6
1	34	35	25	29	11	15	26	23
2	22	62	22	51	26	33	18	44
3	18	81	15	69	20	54	16	60
4	7	92	8	80	13	70	14	74
5	3	97	5	86	11	81	7	84
6	2	100	6	91	7	90	4	89
7	0	100	4	96	3	94	3	92
8	0	100	3	99	2	97	4	96
9	0	100	1	100	3	99	2	99
10	0	100	0	100	0	99	1	100
11	0	100	0	100	0	99	0	100
12	0	100	0	100	0	99	0	100
13	0	100	0	100	0	99	0	100
14	0	100	0	100	0	99	0	100
15	0	100	0	100	0	99	0	100
16	0	100	0	100	0	99	0	100
17	-		0	100	0	99	0	100
18	-		0	100	0	99	0	100

Tab. 107h-D 5.2: Distraktoren-Normierung Subtestteil Lautsprache Verben								
Fehler- anzahl	HRP (10)		G (23)		M = MOD (27)		I (26)	
	f	PR	f	PR	f	PR	f	PR
0	21	10	18	9	6	3	6	3
1	42	40	25	29	21	16	27	19
2	19	69	22	51	15	33	13	38
3	18	86	15	69	18	49	9	48
4	5	97	8	80	14	64	17	60
5	1	100	5	86	13	76	12	74
6	0	100	6	91	10	87	7	83
7	0	100	4	96	4	94	12	92
8	0	100	3	99	4	98	2	99
9	0	100	0	99	0	98	1	100
10	0	100	0	99	0	98	0	100
11	-		0	99	0	98	0	100
12	-		0	99	0	98	0	100
13	-		0	99	0	98	0	100
14	-		0	99	0	98	0	100
15	-		0	99	0	98	0	100
16	-		0	99	0	98	0	100
17	-		0	99	0	98	0	100
18	-		0	99	0	98	0	100

Tab. 107i-D 5.2: Distraktoren-Normierung Subtestteil Lautsprache Adjektive								
Fehler- anzahl	HRP (6)		G (22)		M = MOD (26)		I (35)	
	f	PR	f	PR	f	PR	f	PR
0	23	11	10	5	13	7	3	2
1	30	36	17	18	11	18	15	10
2	28	64	10	31	16	31	14	24
3	19	86	19	44	16	46	10	35
4	5	97	13	59	12	59	7	43
5	1	100	15	73	15	72	12	52
6	0	100	11	85	9	83	9	62
7	-		7	93	7	91	11	72
8	-		3	98	5	96	10	82
9	-		1	100	1	99	8	90
10	-		0	100	1	100	6	97
11	-		0	100	0	100	1	99
12	-		0	100	0	100	1	100
13	-		0	100	0	100	0	100
14	-		0	100	0	100	0	100
15	-		0	100	0	100	0	100
16	-		0	100	0	100	0	100
17	-		0	100	0	100	0	100
18	-		0	100	0	100	0	100

Auch für das Distraktorenprofil soll mit Hilfe des Standardmessfehlers die Genauigkeit festgelegt werden, mit der aufgrund des Profilergebnisses Aussagen über die wahre Merkmalsausprägung einer Testperson getroffen werden können:

Tab. 108-D 5.2: Standardmessfehler im Distraktorenprofil					
Subtest- teil	HRP	MOD	G	M	I
	Standard- messfehler	Standard- messfehler	Standard- messfehler	Standard- messfehler	Standard- messfehler
GSN	1,30	1,37	1,37	1,22	1,49
GSV	0,84	1,07	1,07	0,74	1,08
GSA	0,90	1,43	1,43	1,03	1,50
SSN	1,07	1,04	1,28	1,29	1,53
SSV	1,08	1,52	1,59	1,53	1,69
SSA	0,60	1,34	1,47	1,53	1,70
LSN	1,20	1,50	1,36	1,50	1,43
LSV	1,08	1,50	1,41	1,50	1,54
LSA	1,01	1,50	1,50	1,50	1,68

Die Standardmessfehler im Distraktorenprofil weisen mit Werten zwischen 0,60 und 1,69 Punkten eine gute Messgenauigkeit der einzelnen Subtestteile aus.

6. Literaturverzeichnis des Anwenderhandbuchs

Ahrbeck, B. (1992): **Gehörlosigkeit und Identität. Probleme der Identitätsbildung aus der Sicht soziologischer und psychologischer Theorien.** Hamburg: Signum

Augst, G. (Hrsg. – 1984): **Kinderwort. Der aktive Wortschatz (kurz vor der Einschulung) nach Sachgebieten geordnet mit einem alphabetischen Register.** Frankfurt: Lang

Augst, G. (1989): **Schriftwortschatz. Untersuchungen und Wortlisten zum orthographischen Lexikon bei Schülern und Erwachsenen.** Frankfurt: Lang

Bortz, J. (1993): **Statistik für Sozialwissenschaftler.** 4. vollst. überarb. Aufl.; Berlin/Heidelberg/New York : Springer-Verlag

Günther, K.-B. (1989): **Übersicht über die je 100 häufigst gebrauchten Nomen, Verben, Adjektive von Grundschulkindern (mündlich/schriftlich - Pregel & Rickheit 1987), von Viertklässlern (schriftlich - Augst 1989) sowie ihr Vorkommen bei Vorschulkindern (mündlich - Augst 1984), verglichen mit den je 100 häufigsten Inhaltsworten Erwachsener (mündlich – Ruoff 1981).** Heidelberg: Unveröffentlicht.

Günther, K.-B. (1991): **Probleme der Diagnostik lexikalisch-semantischer Störungen.** In: Handbuch der Sprachtherapie. Band 3: Störungen der Semantik. M. Grohnfeld (Hrsg.), Marhold, S. 167-195.

Günther, K.-B. in Zusammenarbeit mit A. Staab, V. Thiel-Holtz, S. Tollgreif, H. Wudtke (1999): **Bilingualer Unterricht mit gehörlosen Grundschulern: Zwischenbericht zum Hamburger bilingualen Schulversuch.** Theorie und Praxis 3, K.-B. Günther (Hrsg.), Hamburg: Verlag hörgeschädigter Kinder GmbH

Leonhardt, A. (1996): **Didaktik des Unterrichts für Gehörlose und Schwerhörige.** Berlin: Luchterhand

Lienert, G. A., Raatz, U. (1994): **Testaufbau und Testanalyse.** 5. völlig neu bearb. u. erw. Aufl.; Weinheim: Beltz Psychologie Verlags Union

Pregel, D., Rickheit, G. (1987): **Der Wortschatz im Grundschulalter. Häufigkeitswörterbuch zum verbalen, substantivischen und adjektivischen Wortgebrauch.** Hildesheim/Zürich/New York: Olms

Prillwitz, S., Wudtke, H. (1988): **Gebärden in der vorschulischen Erziehung gehörloser Kinder. Zehn Fallstudien zur kommunikativ-sprachlichen Entwicklung gehörloser Kinder bis zum Einschulungsalter.** Hamburg: hörgeschädigte Kinder (Signum)

Ruoff, A. u. M., v. Fuchs, H., Gersbach, B., Graf R., Thiers S. (1981): **Häufigkeitswörterbuch gesprochener Sprache gesondert nach Wortarten alphabetisch, rückläufig alphabetisch und nach Häufigkeiten geordnet.** Tübingen: Niemeyer

Schmidtke, A., Schaller, S., Becker, P. (Deutsche Bearbeitung) (1980): **Coloured Progressive Matrices von J. C. Raven, J. Court and J. Raven Jr.** 2. verbesserte Aufl.; Weinheim: Beltz Test Gesellschaft

Tewes, U. (1985): **HAWIK-R Hamburg-Wechsler Intelligenztest für Kinder Revision 1983, Handbuch und Testanweisung.** 3. korrigierte Aufl.; Bern/Stuttgart/Toronto: Verlag Hans Huber

Verzeichnis der benutzten Literatur

Ahrbeck, B. (1992): **Gehörlosigkeit und Identität. Probleme der Identitätsbildung aus der Sicht soziologischer und psychologischer Theorien.** Hamburg: Signum

Augst, G. (Hrsg. - 1984): **Kinderwort. Der aktive Wortschatz (kurz vor der Einschulung) nach Sachgebieten geordnet mit einem alphabetischen Register.** Frankfurt: Lang

Augst, G. (1989): **Schriftwortschatz. Untersuchungen und Wortlisten zum orthographischen Lexikon bei Schülern und Erwachsenen.** Frankfurt: Lang

Bortz, J. (1993): **Statistik für Sozialwissenschaftler.** 4. vollst. überarb. Aufl., Berlin / Heidelberg / New York : Springer-Verlag

Bühl, A., Zöfel, P. (1996): **SPSS für Windows Version 6.1: Praxisorientierte Einführung in die moderne Datenanalyse.** Bonn: Addison-Wesley-Longman

Claussen, S. H. (1975): **Die soziale Problematik des Hörsprachgeschädigten und die aus ihr resultierende Bedeutung des Sprachunterrichts.** In: Claussen, S. H. u.a.: Psycholinguistik in der Sonderpädagogik. Zur sprachlichen und kognitiven Entwicklung normalsinniger und hörsprachgeschädigter Kinder. Schriften zur Hörgeschädigtenpädagogik, Heft 15, S. 1-21. Berlin: Marhold

Der Brockhaus multimedial 2000 Premium. Mannheim 1999: Bibliografisches Institut & F. A. Brockhaus AG. (permanent aktualisiert; hier abgefragt 2001)

Dunn, L. P. (1959): **Peabody Picture Vocabulary Test. PPVT.** Weinheim / Berlin / Basel: Verlag Julius Beltz

Dupuis, G., Kerkhoff, W. (Hrsg. - 1992): **Enzyklopädie der Sonderpädagogik, der Heilpädagogik und ihrer Nachbargebiete.** Berlin: Edition Marhold im Wissenschaftsverlag Volker Spiess

Freie und Hansestadt Hamburg, Behörde für Schule, Jugend und Berufsbildung (Hrsg. - 1975): **Richtlinien und Lehrpläne, Schule für Gehörlose.** In: Richtlinien und Lehrpläne, Band VI Sonderschulen. Regensburg: Walhalla und Pretoria Verlag KG Georg Zwickenpflug, S. 19-37

Furth, H. (1972): **Denkprozesse ohne Sprache.** Internationale Studien zur pädagogischen Anthropologie. Sprache und Lernen, Band 20. Düsseldorf: Schwann

Gage N. L., Berliner D. C. (1986): **Der Lehrer und standardisierte Tests.** In: Bach, G. (Hrsg. Und aus dem Amerikanischen übersetzt) Pädagogische Psychologie. 4. Aufl.; Weinheim und München: Beltz

Glück, C. W. (1998): **Kindliche Wortfindungsstörungen. Ein Bericht des aktuellen Erkenntnisstandes zu Grundlagen, Diagnostik und Therapie.** Frankfurt: Lang

Grubitzsch, S. (1978): **Konstruktion psychologischer Tests.** In: Grubitzsch, S., Rexilius, G. (Hrsg.): Testtheorie – Testpraxis. Voraussetzungen, Verfahren, Formen und Anwendungsmöglichkeiten psychologischer Tests im kritischen Überblick. Reinbek bei Hamburg: Rohwohlt, S. 75-106

Günther, K.-B. (1980): **Probleme der sprachlichen Entwicklung bei fünf- bis siebenjährigen Kindern. Sprach- und entwicklungspsychologische Grundlegung.** Weinheim / Basel: Beltz

Günther, K.-B. (1985): **Schriftsprache bei hör- und sprachgeschädigten Kindern. Bedeutung und Funktion für Sprachaufbau und Entwicklung, dargestellt am Beispiel gehörloser Kinder.** Hörgeschädigtenpädagogik, Beiheft 9; (2. verbesserte Aufl.); Heidelberg: Julius Groos Verlag

Günther, K.-B. (1986): **Zur Notwendigkeit besonderer konzeptioneller Überlegungen und didaktischer Konsequenzen für den Schriftspracherwerb gehörloser Kinder.** Hörgeschädigtenpädagogik, 40, S. 151-178

Günther, K.-B. (1987): **Zur Bedeutung von sprach- und entwicklungspsychologischen Erkenntnissen bei dysgrammatisch sprechenden Kindern.** In: dsg-Lg Rheinland (Hrsg.), S. 165-200

Günther, K.-B. (1988): **Probleme der Diagnostik lexikalisch-semantischer Entwicklungsstörungen am Beispiel des aktiven Wortschatzes für drei- bis sechsjährige Kinder (AWST 3-6).** In: Ders. (Hrsg.): Sprachstörungen. Probleme ihrer Diagnostik bei mentalen Retardierungen, Entwicklungsdysphasien und Aphasien. Heidelberg, S. 117-166

Günther, K.-B. (1989 a): **Die Bedeutung des Symbolverständnisses für die Sprachentwicklung des Kindes.** Heidelberg: Ms.

Günther, K.-B. (1989 b): **Übersicht über die je 100 häufigst gebrauchten Nomen, Verben, Adjektive von Grundschulkindern (mündlich/schriftlich - Pregel & Rickheit 1987), von Viertklässlern (schriftlich - Augst 1989) sowie ihr Vorkommen bei Vorschulkindern (mündlich - Augst 1984), verglichen mit den je 100 häufigsten Inhaltsworten Erwachsener (mündlich - Ruoff 1981).** Heidelberg: Unveröffentlicht. Im Anhang dieser Arbeit vorfindlich

Günther, K.-B. (1990): **Vergleich der symbolisch visuellen Wahrnehmungs- und visomotorischen Produktionsfähigkeit von sprachentwicklungsgestörten, gehörlosen und nichtbehinderten Kindern (VIS). Eine empirische Grundlagenuntersuchung zu den wahrnehmungsmässigen und feinmotorisch koordinativen Voraussetzungen für den Schriftspracherwerb.** Arbeiten zur Sprachanalyse, Band 20. Frankfurt am Main: Lang

Günther, K.-B. (1991): **Probleme der Diagnostik lexikalisch-semantischer Störungen.** In: Handbuch der Sprachtherapie. Band 3: Störungen der Semantik. Hrsg.: M. Grohnfeld. Marhold S. 167-195.

Günther, K.-B. in Zusammenarbeit mit A. Staab, V. Thiel-Holtz, S. Tollgreef, H. Wudtke (1999 a): **Bilingualer Unterricht mit gehörlosen Grundschulern: Zwischenbericht zum Hamburger bilingualen Schulversuch.** Theorie und Praxis 3, K.-B. Günther (Hrsg.), Hamburg: Verlag hörgeschädigter Kinder GmbH

Günther, K.-B. (1999 b): **Förderschwerpunkt Hören.** In Zeitschrift für Heilpädagogik 4/99, S. 171-181

Häcker, H., Stapf, H. (Hrsg. - 1998): **Dorsch Psychologisches Wörterbuch 1998.** 13. überarb. u. erw. Aufl.; Bern: Gruber

Ingenkamp, K. (Hrsg. - 1970): **Wortschatztest für Schulanfänger (WSS 1)**. Weinheim / Berlin: Verlag Julius Beltz

Ingenkamp, K. (1972 a): **Die Entwicklung von Schultests**. In: Ingenkamp, K. (Hrsg.): Tests in der Schulpraxis. Eine Einführung in Aufgabenstellung, Beurteilung und Anwendung von Tests. Weinheim: Beltz Verlag, S. 43-73

Ingenkamp, K. (1972 b): **Erläuterung einiger statistischer Begriffe und Vorgehensweisen als Hilfe zur Interpretation von Testergebnissen**. In: Ingenkamp, K. (Hrsg.): Tests in der Schulpraxis. Eine Einführung in Aufgabenstellung, Beurteilung und Anwendung von Tests. Weinheim: Beltz, S. 146-171

Jann, P. A. (1994): **Methoden der Sprachvermittlung beim gehörlosen Kind: Zur Grundlegung einer systematischen Spracherwerbsforschung**. Heidelberg: Ed. Schindele

Kiese, C., Kozielski, P.-M. (1979): **Aktiver Wortschatztest für drei- bis sechsjährige Kinder, AWST 3-6**. Weinheim: Beltz Test Gesellschaft.

Kleber, E. (1976): **Grundlagen Sonderpädagogischer Diagnostik**. Eine Einführung für Studienanfänger und interessierte Pädagogen. Berlin: Carl Marhold Verlagsbuchhandlung

Leonhardt, A. (1996): **Didaktik des Unterrichts für Gehörlose und Schwerhörige**. Berlin: Luchterhand

Lienert, G. A., Raatz, U. (1994): **Testaufbau und Testanalyse**. 5. völlig neu bearb. u. erw. Aufl.; Weinheim: Beltz Psychologie Verlags Union

Michelsen, U. A., Müller, J. (1988): **Eingrenzung der Häufigkeit einzelner Ereignisse beim Test auf Gleichverteilung. Ein Instrument zur Distraktorenanalyse**. Diagnostica

Mietzel, G. (unter Mitarbeit von Rüssmann-Stöhr, C.) (1993): **Einführung in die pädagogische Psychologie für Pädagogen und Psychologen**. Aus der Reihe Psychologie in Unterricht und Erziehung. 4. Aufl.; Göttingen: Hogrefe Verlag für Psychologie

Ogden, C. K., Richards I. A. (1974): **Die Bedeutung der Bedeutung**. Deutsche Ausgabe: Frankfurt, Suhrkamp. Originalausgabe: Ogden, C. K., Richards I. A. (1923): The Meaning of Meaning. London: Routledge and Kegan Paul

Orth, B. (1974): **Einführung in die Theorie des Messens**. Stuttgart: Kohlhammer

Piaget, J. (1969): **Nachahmung, Spiel und Traum. Die Entwicklung der Symbolfunktion beim Kinde**. Stuttgart

Pregel, D., Rickheit, G. (1987): **Der Wortschatz im Grundschulalter. Häufigkeitswörterbuch zum verbalen, substantivischen und adjektivischen Wortgebrauch**. Hildesheim / Zürich / New York: Olms

Prillwitz, S. (1975): **Lautsprache versus Gebärdensprache. Versuch einer Funktionsbestimmung**. In: Claussen, S. H. u.a.: Psycholinguistik in der Sonderpädagogik. Zur sprachlichen und kognitiven Entwicklung normalsinniger und hörsprachgeschädigter Kinder. Schriften zur Hörgeschädigtenpädagogik, Heft 15; Berlin: Marhold, S. 51-108

Prillwitz, S. (1982): **Zum Zusammenhang von Kognition, Kommunikation und Sprache mit Bezug auf die Gehörlosenproblematik.** Band 130 der Schriftenreihe des Bundesministers für Jugend, Familie und Gesundheit. Stuttgart: Verlag W. Kohlhammer

Prillwitz, S. (1986 a): **Grundlagenreferat: Die Gebärde der Gehörlosen. Ein Beitrag zur Deutschen Gebärdensprache und ihrer Grammatik.** In: Die Gebärde in Erziehung und Bildung Gehörloser. Internationaler Kongreß am 9. Und 10. November 1985 im Congress Centrum Hamburg. Tagungsbericht. Hamburg: Verlag hörgeschädigte Kinder GmbH, S. 55 - 78

Prillwitz, S. (1986 b): **Die Gebärde in Erziehung und Bildung Gehörloser. Situationsbestimmung und Perspektiven nach dem Hamburger Gebärdenkongreß.** In: Die Gebärde in Erziehung und Bildung Gehörloser. Internationaler Kongreß am 9. Und 10. November 1985 im Congress Centrum Hamburg. Tagungsbericht. Hamburg: Verlag hörgeschädigte Kinder GmbH, S. 175 - 189

Prillwitz, S., Wudtke, H. (1988): **Gebärden in der vorschulischen Erziehung gehörloser Kinder. Zehn Fallstudien zur kommunikativ-sprachlichen Entwicklung gehörloser Kinder bis zum Einschulungsalter.** Hamburg: hörgeschädigte Kinder (Signum)

Probst, H. (1981): **Zur Diagnostik und Didaktik der Oberbegriffsbildung.** Solms-Oberbiel: Jarick-Oberbiel

Probst, H. (1982): **Strukturbezogene Diagnostik.** In: Probst, H. (Hrsg.): Kritische Behindertenpädagogik in Theorie und Praxis. Vierteljahresschrift für Behindertenpädagogik in Praxis, Forschung und Lehre und Integration Behinderter, Beiheft 7 (2. Aufl.); Solms-Oberbiel: Jarick-Oberbiel, S.113-135

Probst, H. (1992): **Qualitative Testverfahren in der Entwicklung pädagogisch-diagnostischer Verfahren.** In: K.-H. Ingenkamp, R. S. Jäger, H. Petillon, B. Wolf (Hrsg.): Empirische Pädagogik 1970-1990. Band I; Weinheim: Deutscher Studien Verlag, S. 169-196

Raatz, U., Möhling, R. (1971): **Frankfurter Test für Fünfjährige – Wortschatztest (FTF-W).** Weinheim: Beltz Test GmbH

Rauchfleisch, U. (1994): **Testpsychologie – Eine Einführung in die Psychodiagnostik.** 3. Aufl.; Göttingen / Zürich: Vandenhoeck & Ruprecht

Rexilius, G. (1978): **Grenzen der Testerei.** In: Grubitzsch, S., Rexilius, G. (Hrsg.): Testtheorie – Testpraxis. Voraussetzungen, Verfahren, Formen und Anwendungsmöglichkeiten psychologischer Tests im kritischen Überblick. Reinbek bei Hamburg: Rohwohlt, S. 112-166

Ruoff, A. u. M.v. Fuchs, H., Gersbach, B., Graf R., Thiers S. (1981): **Häufigkeitswörterbuch gesprochener Sprache gesondert nach Wortarten alphabetisch, rückläufig alphabetisch und nach Häufigkeiten geordnet.** Tübingen: Niemeyer

Schiefele, H., Oerter, R. (Hrsg. - 1973): **Diagnostik in der Schule. Beiträge zu einer pädagogischen Orientierung der Schülerbeurteilung.** München: R. Oldenbourg Verlag

Schmidtke, A., Schaller, S., Becker, P. (Deutsche Bearbeitung) (1980): **Coloured Progressive Matrices von J.C. Raven, J. Court and J. Raven Jr.** 2. verbesserte Aufl.; Weinheim: Beltz Test Gesellschaft

Schuck, K.-D. (2000): **Diagnostik**. In: J. Borchert (Hrsg.): Handbuch der sonderpädagogischen Psychologie. Göttingen: Hogrefe-Verlag

Stern, C., Stern, W. (1975): **Die Kindersprache. Eine psychologische und sprachtheoretische Untersuchung**. Nachdruck der 4. Aufl.: Darmstadt

Stern, W. (1971): **Psychologie der frühen Kindheit**. 10. Aufl.: Heidelberg

Sturges, H. A. (1926): **The choice of a class interval**. J. Amer. Statist. Assoc.

Szagun, G. (1983): **Bedeutungsentwicklung beim Kind. Wie Kinder Wörter entdecken**. München / Weinheim: Psychologie Union – Urban und Schwarzenberg

Szagun, G. (1986): **Sprachentwicklung beim Kind. Eine Einführung**. München / Weinheim: Psychologie Union – Urban und Schwarzenberg

Tewes, U. (1985): **HAWIK-R Hamburg-Wechsler Intelligenztest für Kinder Revision 1983, Handbuch und Testanweisung**. 3. korrigierte Aufl.; Bern / Stuttgart / Toronto: Verlag Hans Huber

Textor, A. M. (2000): **Auf Deutsch – Das Fremdwörterlexikon**. Vollständig überarbeitete u. erw. Neuausgabe; Hamburg: Rowohlt Taschenbuch Verlag

Voit, H. (1992): **Von neuen Weichenstellungen in der Gehörlosenpädagogik. Entscheidende Überlegungen**. Heidelberg: Hörgeschädigtenpädagogik, Beiheft 29, S.167-185

Wersich, R. (1992): **Begriffsbildung bei hörbehinderten Vorschulkindern. Eine Studie zum Einfluß kommunikativer Kompetenz in Verbal- oder Gebärdensprache auf die kognitive Entwicklung**. Internationale Arbeiten zur Gebärdensprache und Kommunikation Gehörloser, Band 23, Hrsg.: S. Prillwitz. Hamburg: Signum

Wisch, F.-H. (1990): **Lautsprache und Gebärdensprache. Die Wende zur Zweisprachigkeit in Erziehung und Bildung Gehörloser**. Hamburg: Signum Verlag

Wolf, W. (1995): **Qualitative versus quantitative Forschung**. In: E. König, P. Zedler (Hrsg.): Bilanz qualitativer Forschung, Band I: Grundlagen qualitativer Forschung. Weinheim: Deutscher Studien Verlag, S. 309–329

Zimmer, D. E. (1988): **So kommt der Mensch zur Sprache. Über Spracherwerb, Sprachentstehung, Sprache und Denken**. Zürich: Haffmanns Taschenbuchverlag

Herzlichen Dank an die Illustratorin

**Birgit Wudtke
Daimlerstraße 32
22763 Hamburg**

Anhang 1

zu

Entwicklung eines Wortschatztests für gehörlose Kinder im Grundschulalter in Gebärden-, Schrift- und Lautsprache

Inauguraldissertation
zur Erlangung des akademischen Grades
des Doktors der Philosophie
am Fachbereich Erziehungswissenschaft
der Universität Hamburg

vorgelegt von

Sibylle Bizer und Anne-Katrin Karl

Hamburg, 2002

Inhalt Anhang 1

Seite	Tabelle / Abbildung	Bezeichnung
Anh. 1: B 1.3a, S. 1	Tab. 4-B 1.3	Wortliste Günther (1989 b)
Anh. 1: B 1.3b, S. 12	Tab. 6a/b/c-B 1.3	Distraktorenähnlichkeiten Pilotversion N/V/A
Anh. 1: B 2.3.1, S. 16	Tab. 7a/b-B 2.3.1	Klasseninterner Vergleich der Leistungen Gehörloser Klasse 3/4
Anh. 1: B 2.3.2, S. 18	Tab. 8a/b-B 2.3.2	Klasseninterner Vergleich der Zeiten Gehörloser Klasse 3/4
Anh. 1: B 2.3.3, S. 20	Tab. 9a/b-B 2.3.3	Gesamtergebnisse Gehörloser Klasse 3/4
Anh. 1: B 2.3.4, S. 21	Tab. 10a/b/c-B 2.3.4	Ergebnisse/Fehleranalyse Hörender
Anh. 1: B 2.4.1, S. 26	Tab. 11-B 2.4.1	Fehlerverteilung Gehörloser
Anh. 1: B 2.4.2, S. 27	Tab. 12a-k-B 2.4.2	Rating Testpersonen 1 - 11, Klasse 3/4
Anh. 1: B 2.4.3a, S. 33	Abb. 9a-i-B 2.4.3	Schwierigkeit–Trennschärfe Gehörlose Pilotstudie GSN - LSA
Anh. 1: B 2.4.3b, S. 42	Abb. 10a-f-B 2.4.3	Schwierigkeit–Trennschärfe Hörende Pilotstudie SSN - LSA
Anh. 1: B 2.4.5, S. 48	Tab. 13a/b/c-B 2.4.5	Fehleranalyse Gehörloser N/V/A
Anh. 1: B 2.4.7, S. 51	Tab. 14a/b/c-B 2.4.7	Itemrangliste N/V/A
Anh. 1: B 3.2.1, S. 53	Tab. 21-B 3.2.1	Testpunktwerte
Anh. 1: B 3.2.2, S. 56	Tab. 22a/b/c-B 3.2.2	Distraktorenverteilung N/V/A
Anh. 1: B 3.3a, S. 64	Tab. 23-B 3.3	Punkteverteilung in Kategorien
Anh. 1: B 3.3b, S. 65	Abb. 11a-i-B 3.3	Kategorisierte Verteilungskurven GSN - LSA
Anh. 1: B 3.4.2, S. 69	Abb. 12a-i-B 3.4.2	Schwierigkeit–Trennschärfe Version I GSN - LSA
Anh. 1: B 3.4.3, S. 78	Tab. 26-B 3.4.3	Interkorrelationen
Anh. 1: B 3.4.4, S. 79	Tab. 27a-i-B 3.4.4	Reliabilitätswerte der Testaufgaben GSN - LSA
Anh. 1: B 3.4.5a, S. 85	Tab. 29a-e-B 3.4.5	Distraktorentrennschärfewerte
Anh. 1: B 3.4.5b, S. 90	Tab. 30a-e-B 3.4.5	Korrelation Distraktor–Raven
Anh. 1: B 3.5.3, S. 95	Tab. 34-B 3.5.	Testergebnisse Version I-6
Anh. 1: B 3.5.4a, S. 98	Tab. 35a/b/c-B 3.5.4	Neue Platzverteilung aufgrund der Schwierigkeit N/V/A
Anh. 1: B 3.5.4b, S. 101	Tab. 36-B 3.5.4	Mittelwerte der Schwierigkeitsrangfolge für Testheft I
Anh. 1: B 3.5.4c, S. 102	Abb. 13-B 3.5.4	Gebärdenfotos und Ausführungsanweisungen
Anh. 1: B 3.5.6a, S. 111	Tab. 40a/b/c-B 3.5.6	Schwierigkeit und Trennschärfe Version II N/V/A
Anh. 1: B 3.5.6b, S. 114	Abb. 14a-i-B 3.5.6	Schwierigkeit–Trennschärfe Version II GSN - LSA

Seite	Tabelle / Abbildung	Bezeichnung
Anh. 1: B 3.5.7, S. 123	Tab. 42a-j-B 3.5.7	Modifiziertes Testprotokoll Deckblatt, GSN - LSA
Anh. 1: B 3.7.2a, S. 133	Tab. 46-B 3.7.2	Aufteilung der Items in Testhälften A und B
Anh. 1: B 3.7.2b, S. 134	Tab. 47-B 3.7.2	Gesamtwerte der Testhälften A und B nach Halbierung über statistische Kennwerte
Anh. 1: B 3.7.2c, S. 136	Tab. 48-B 3.7.2	Reliabilitätskoeffizienten der Split-half-Methode nach statistischen Kennwerten
Anh. 1: B 3.7.3, S. 137	Tab. 53-B 3.7.3	Lehrer-Rating der passiven Wortschatzkenntnis
Anh. 1: B 3.8.4, S. 139	Tab. 55a-i-B 3.8.4	Normierung über Flächentransformation GSN - LSA
Anh. 1: B 3.9.2a, S. 144	Tab. 59-B 3.9.2	Standardabweichung je zweier Subtestteile für die Beurteilung intraindividuelle Unterschiede
Anh. 1: B 3.9.2b, S. 145	Tab. 60-B 3.9.2	Reliabilität je zweier Subtestteile für die Beurteilung intraindividuelle Unterschiede
Anh. 1: B 3.10.3a, S. 146	Tab. 66a-e-B 3.10.3	Standardabweichung je zweier Subtestteile HRP - I
Anh. 1: B 3.10.3b, S. 148	Tab. 67a-e-B 3.10.3	Reliabilität je zweier Subtestteile HRP - I
Anh. 1: C 1.7.4, S. 150	Abb. 23-C 1.7.4	Profilverläufe Konzepte
Anh. 1: C 1.7.5, S. 151	Abb. 24-C 1.7.5	Profilverläufe Klassen
Anh. 1: C 1.8.1, S. 152	Abb. 25-C 1.8.1	Fehleranzahl
Anh. 1, S. 153		Beispielseite Testheft I (Gebärden-, Lautsprache)
Anh. 1, S. 154		Beispielseite Testheft II (Schriftsprache)

Tab. 4-B 1.3: Wortliste Günther (1989 b)

AUGST, G. (1989): *Schriftwortschatz. Untersuchungen und Wortlisten zum orthographischen Lexikon bei Schülern und Erwachsenen*. Frankfurt/Bern/ New York/Paris: Lang (= Vierthklassier in den Übersichten).

Stichprobe: 10 Viertklässler, 2 Jungen und 8 Mädchen, keine weiteren personalen und sozialen Angaben.

Methode der Sprachdatenerfassung: Auswertung sämtlicher Hefte (aller Fächer), die die 10 Schülerinnen während des 4. Schuljahres vollgeschrieben haben.

Sprachmodalität und Korpusumfang: Schriftlich. 142.020 fortlaufende Wörter (Tokens) auf 12.633 Types, davon 73943 (Tokens) auf 12283 verschiedene Inhaltswörter incl. Adverbien, aber ohne Modalverben sowie sein, haben, werden als Vollverben bzw. Kopula und einige Häufigkeitsadverbien und -adjektive.

Differenzierung der Wortlisten: Nicht nach Wortarten gegliederte alphabetische Gesamtliste mit Häufigkeitsangaben sowie Häufigkeitsrangliste bis N > 9 für die Inhaltswörter.

Anmerkungen: In gleicherweise liegt bei AUGST (1989) auch ein Korpus für 10.-Klässler vor.

Für beide Korpora hat AUGST auch eine alphabetische Liste der (Grund-) Lexeme erstellt.

In seinen den Listen vorangestellten Untersuchungen geht AUGST vergleichend auf den *Manheimer Corpus* ein, der knapp 9 Millionen fortlaufende Wortformen des erwachsenen Schriftwortschatzes vorwiegend aus Zeitungen seit 1960 enthält. Dieser Korpus ist leider bis heute nicht in einer verarbeitbaren Form veröffentlicht worden.

RUOFF, A. u. M. V. FUCHS, H./GERSBACH, B./GRAF, R. & THIERS, S. (1981): *Häufigkeitswörterbuch gesprochener Sprache gesondert nach Wortarten alphabetisch, rückläufig alphabetisch und nach Häufigkeiten geordnet*. Tübingen: Niemeyer, 1981 (= Erwachsene in den Übersichten).

Stichprobe: 2500 Erwachsene differenziert nach dem Geschlecht und 3 Altersgruppen aus dem Fränkischen (582), Schwäbischen (1237) und Alemannischen (681) Sprachraum.

Methode der Sprachdatenerfassung: Proben unvorgeleiteter freier Gespräche aus 400 Orten des südwestdeutschen Sprachraums: 1500 Tonbandaufnahmen von zusammen 350 Stunden mit über 2,5 Millionen Wörtern.

Sprachmodalität und Korpusumfang: Mündlich. 500000 Wörter (Tokens) auf 15676 Types.

Differenzierung der Wortlisten: Getrennte alphabetische, rückläufig-alphabetische und nach Häufigkeiten geordnete Gesamtlisten für alle Wortarten mit Häufigkeits- und Prozentangaben.

Klaus-B. Günther: Heidelberg, 3. im November 1989

Übersicht über die je 100 häufigst gebrauchten Nomen, Verben, Adjektive von Grundschulkindern (mündlich/schriftlich - PREGEL & RICKHEIT 1987), von Viertklässlern (schriftlich - AUGST 1989) sowie ihr Vorkommen bei Vorschulkindern (mündlich - AUGST 1984), verglichen mit den je 100 häufigsten Inhaltswörtern Erwachsener (mündlich - RUOFF 1981)

Angaben zu den berücksichtigten Korpora

PREGEL, D. & RICKHEIT, G. (1987): *Der Wortschatz im Grundschulalter. Häufigkeitswörterbuch zum verbalen, substantivischen und adjektivischen Wortgebrauch*. Hildesheim/Zürich/New York: Olms (= Grundschul in den Übersichten).

Stichprobe: 1048 Grundschüler der Klassen 1-4, Alter 6:0-10:0 aus verschiedenen, soziographisch differenzierten Regionen der Bundesrepublik.

Methode der Sprachdatenerfassung: 4 festgelegte Sprachproduktionsimpulse - Lustige Geschichte, Tiergeschichte, Unfall(-Bericht), Beschreibung von Personen bzw. Kasperpuppen - sowie als diesen nicht zuzuordnende Kategorie 1 freie Erlebniszerzählung.

Sprachmodalität und Korpusumfang: 1. Mündlich. 2400 Texte von 705 Schülern mit 5137 verschiedenen (=Types) und 61839 Verben, Nomen und Adjektiven insgesamt (=Tokens). 2. Schriftlich: 800 Texte von 341 3./4.-Klässlern mit 3923 Types und 28098 Tokens der 3 Inhaltswortarten. Mündlich + Schriftlich: 6921 Types und 89937 Tokens.

Differenzierung der Wortlisten: 3 getrennte alphabetische Gesamtlisten mit Häufigkeits- und Prozentangaben für die Verben, Nomen und Adjektive, differenziert nach Gesamtkorpus, mündlichem und schriftlichem Korpus sowie Alter und mündlich/schriftlicher Differenzierung.

Ranghäufigkeitslisten der 3 Inhaltswortarten für den Gesamtkorpus, den mündlichen und schriftlichen Korpus, das Alter, die Klassenstufe, das Geschlecht, die soziale Schichtung und das Thema (Sprachimpuls), jeweils differenziert nach mündlich/schriftlich.

AUGST, G. (Hg. - 1984): *Kinderwort. Der aktive Kinderwortschatz (kurz vor der Einschulung nach Sachgebieten geordnet mit einem alphabetischen Register*. Frankfurt: Lang. (= Vorschüler in den Übersichten).

Stichprobe: 10 Vorschulkinder kurz vor der Einschulung aus der mittleren bis oberen Mittelschicht.

Methode der Sprachdatenerfassung: Aufnahme der sprachlichen Äußerungen in 200 Stunden innerhalb eines Zeitraums von 4 Monaten per schriftliche Notizen bzw. Kassetteneintrag.

Sprachmodalität und Korpusumfang: Mündlich. Angegeben wird nur der nicht nach Wortarten differenzierte Umfang der Types für die einzelnen Kinder: Minimum-Wortschatz = 3110 - Maximum-Wortschatz = 5300.

Differenzierung der Wortlisten: Der Korpus ist als Sachwörterbuch nach WEHRLE/EGGERS aufbereitet u. mit einem alphabetischen Register versehen. Häufigkeitsangaben beschränken sich darauf, ob ein Wort von 1, 2-5 oder mehr als 5 der 10 beteiligten Kinder mindestens einmal gebraucht wurde.

lfd. Nr.	Lexen	Grundschüler An- zahl	Vorschüler An- zahl	1.-Klassler An- zahl	Erwachsene An- zahl	Rang Gruppe	Erwachsene An- zahl	Rang Gruppe
Nomen I		++	++	++	++	++	++	++
Summe	Noenen	32500	3772*		54052	9398*		
001 Mutter/Mutti		832	001	+++	245	005	296	014
002 Auto		503	002	+++	110	018	128	058
003 Haus		477	003	+++	223	006	460	006
004 Vater/Vati		461	004	+++	188	008	593	004
005 Haar		445	005	+++	24	++	9	+
006 Auge		399	006	+++	73	048	30	++
007 Hund		362	007	+++	110	018	15	++
008 Nase		350	008	+++	42	++	8	+
009 Katze		328	009	+++	47	098	10	++
010 Kleid		320	010	+++	34	++	40	++
011 Tag		320	011	+++	255	004	861	002
012 Mann		294	012	+++	126	014	442	007
013 Hand		280	013	+++	93	028	253	019
014 Schwester		272	014	+++	37	++	120	063
015 Bruder		257	015	+++	43	++	202	030
016 Frau		233	016	+++	124	015	242	021
017 Mund		228	017	+++	35	++	3	+
018 Oma/Omi		227	018	+++	84	033	26	++
019 Wasser		226	019	+++	136	011	140	051
020 Junge Bub		225	020	+++	100	024	6	2)
021 Kopf		221	021	+++	65	059	73	++
022 Kasper		209	022	+++	5	+	-	-
023 Straße		209	022	+++	115	017	130	055
024 Gesicht		197	024	+++	26	++	25	++
025 Freundin		196	025	++	22	++	10	++

lfd. Nr.	Lexen	Grundschüler An- zahl	Vorschüler An- zahl	1.-Klassler An- zahl	Erwachsene An- zahl	Rang Gruppe	Erwachsene An- zahl	Rang Gruppe
Nomen 2		++	++	++	++	++	++	++
026 Schule		196	025	+++	122	016	214	028
027 Bett		185	027	+++	80	037	121	062
028 Puppe		183	028	+++	29	++	4	+
029 Krone		181	029	+++	13	++	4	+
030 Vogel		178	030	+++	84	033	6	+
031 Freund		174	031	+++	81	035	20	++
032 Papa/Papi		171	032	+++	19	++	2	+
033 Rad		171	032	+++	38	++	43	++
034 Jahr		161	034	+++	153	009	1511	001
035 Baum		157	035	+++	74	046	134	054
036 Fahrrad		156	036	+++	94	027	42	++
037 Kind		150	037	+++	289	003	416	009
038 Mütze		146	038	+++	15	++	3	+
039 Polizei		143	039	+++	28	++	25	++
040 Tante		139	040	+++	31	++	17	++
041 Affe		135	041	+++	27	++	1	+
042 Wald		130	042	+++	69	055	430	008
043 Unfall		126	043	+++	14	++	23	++
044 Ohr		125	044	+++	50	089	4	+
045 Onkel		121	045	+++	27	++	26	++
046 Ball		120	046	+++	68	057	17	++
047 Wagen		120	046	+++	75	043	141	050
048 Becke		119	048	+++	-	-	2	+
049 Mama/Mami		119	048	+++	12	++	6	+
050 Bein		117	050	+++	39	++	6	+

Lfd. Nr.	Lexem	Grundschüler An- zahl	Rang Gruppe	Vor- chüler An- zahl	4.-Klässler An- zahl	Erwachsene An- zahl	Rang Gruppe
Nomen 3							
051	Fuß	116	051	+++	73	048	100 074
052	Arm	114	052	+++	-	-	22 ++
053	Tür	113	053	+++	46	100	28 ++
054	Seite	110	054	+++	80	037	91 080
055	Clown	107	055	+++	14	++	- -
056	Krankenhaus	106	056	+++	11	++	13 ++
057	Lippe	106	056	+++	6	+	3 +
058	Teufel	106	056	+++	24	++	27 ++
059	Maus	105	059	+++	51	087	5 +
060	Knie	100	060	+++	10	++	12 ++
061	Reh	99	061	+++	11	++	12 ++
062	Roller	99	061	+++	4	+	1 +
063	Kaninchen	95	063	+++	7	+	1 +
064	Hals	94	064	+++	15	++	16 ++
065	Opa	94	064	+++	26	++	5 +
066	Loch	90	066	+++	53	081	72 ++
067	Morgen	88	067	+++	58	071	182 035
068	Zipfelmütze	88	067	++	-	-	- -
069	Angst	85	069	+++	32	++	52 ++
070	König	85	069	+++	14	++	23 ++
071	Garten	84	071	+++	33	++	51 ++
072	Stein	84	071	+++	39	++	68 ++
073	Tier	83	073	+++	78	041	10 ++
074	Krankswagen	82	074	+	5	+	- -
075	Bommel	80	075	++	-	-	- -

Lfd. Nr.	Lexem	Grundschüler An- zahl	Rang Gruppe	Vor- chüler An- zahl	4.-Klässler An- zahl	Erwachsene An- zahl	Rang Gruppe
Nomen 4							
076	Eltern	80	075	++	62	065	38 ++
077	Hose	80	075	+++	20	++	26 ++
078	Kurve	80	075	+++	15	++	7 +
079	Zug	79	079	+++	31	++	66 ++
080	Pferd	78	080	+++	85	031	43 ++
081	Uhr	77	081	+++	147	010	38 ++
082	Ding	76	082	+++	7	+	284 016
083	Fernsehen	75	083	+++	?	?	15 ++
084	Mädchen Madlein	75	083	+++	53	081	(- 189 033)
085	Schuh	75	083	+++	39	++	55 ++
086	Tisch	75	083	+++	50	089	36 ++
087	Zoo	75	083	+++	10	++	- -
088	Kragen	74	088	+++	5	+	- -
089	Hase	72	089	+++	34	++	7 +
090	Berg	71	090	+++	65	059	130 055
091	Fenster	71	090	+++	43	++	52 ++
092	Käfig	70	092	+++	6	+	- -
093	Stall	70	092	+++	30	++	164 040
094	Zirkus	70	092	+++	21	++	3 +
095	Abend	69	095	+++	58	071	176 038
096	Bart	68	096	+++	5	+	6 +
097	Zahn	68	096	+++	54	079	6 +
098	Essen	67	098	+++	?	?	78 099
099	Leute	67	098	+++	50	088	701 003
100	Milch	67	098	+++	72	051	93 084
101	Stück	67	098	+++	62	065	209 029

Weitere Nomen: (67, 101) = Grundschüler

Lfd. Nr.	Lexem	Grundschüler An- zahl Gruppe	Vor- schüler An- zahl Gruppe	4.-Klässler An- zahl Gruppe	Erwachsene An- zahl Gruppe
Nomen 5					
102	Antwort	-	++	364 001	5 +
103	Rechnung	-	+	307 002	9 +
104	Frage	2	+	222 007	18 ++
105	Hausaufgabe	9	+	133 012	3 +
106	Wort	5	+	127 013	22 ++
107	Stadt	15	++	110 015	127 058
108	Fehler	4	+	107 021	13 ++
109	Stunde	43	++	107 021	236 024
110	Zeit	64	++	101 024	517 005
111	Buch	27	++	97 027	47 ++
112	Klasse	39	++	91 029	81 092
113	El	32	++	89 030	23 ++
114	Land	13	++	85 031	1 +
115	Rest	3	+	82 035	4 +
116	Monat	12	++	79 039	90 089
117	Verbesserung	-	-	79 039	-
118	Zahl	1	+	77 042	5 +
119	Geld	28	++	75 043	224 027
120	Übung	-	-	75 043	13 ++
121	Probe	-	-	74 046	18 ++
122	Berichtigung	-	-	72 051	-
123	Note	-	-	72 051	1 +
124	Kasten	11	++	71 054	7 +
125	Verb	-	-	69 055	-

Häufigkeitsnomen (Rang < 101) bei den Viertklässlern, die nicht unter den 100 häufigsten Nomen bei den Grundschülern verzeichnet sind.

Lfd. Nr.	Lexem	Grundschüler An- zahl Gruppe	Vor- schüler An- zahl Gruppe	4.-Klässler An- zahl Gruppe	Erwachsene An- zahl Gruppe
126	Weg	55	++	68 057	149 047
127	Boden	54	++	65 059	141 050
128	Brot	34	++	65 059	103 067
129	Schüler	-	-	65 059	15 ++
130	Grub	1	+	63 063	3 +
131	Luft	32	++	63 063	18 ++
132	Bild	61	++	61 067	19 ++
133	Mensch	33	++	61 067	70 ++
134	Spiel	9	+	61 067	30 ++
135	Fluß	5	+	59 070	2 +
136	Alarm	-	-	57 073	4 +
137	Ende	33	++	57 073	44 ++
138	Fisch	41	++	57 073	24 ++
139	Fuchs	41	++	57 073	7 +
140	Woche	47	++	57 073	253 019
141	Satz	10	++	56 078	2 +
142	Wind	8	+	54 079	19 ++
143	Bauer	27	++	53 081	321 011
144	Lehrer	8	+	52 084	75 ++
145	Nuß	22	++	52 084	3 +
146	Wiese	53	++	52 084	73 ++
147	Diktat	2	+	50 088	1 +
148	Gott	13	++	50 088	82 068
150	Meer	-	-	50 088	7 +

Lfd. Nr.	Lexem	Grundschüler An- zahl	Vorschüler An- zahl	4.-Klässler An- zahl	Erwachsene An- zahl
		++	++	++	++
Nomen I					
151	Brief	8	+	49	094
152	Erde	40	++	48	095
153	Nacht	26	++	48	095
154	Person	1	+	48	095
155	Holz	19	++	47	098
156	Arbeit	14	++	46	100
157	Platz	34	++	46	100
158	Stelle	17	++	46	100
159	Schloß	5	+	46	100
Weitere Nomen (= Viertklässler)					
Nomen über Rangplatz 100 (= Erwachsene)					
				<46	>104
					<76 >100

Lfd. Nr.	Lexem	Grundschüler An- zahl	Vorschüler An- zahl	4.-Klässler An- zahl	Erwachsene An- zahl
		++	++	++	++
Verben I					
Summe	Verben	41899	23452	21537	28307b 71581 44091c
001	sein	6060	001	+++	o.A. 001:b 11072 001
002	haben	3633	002	+++	o.A. 002:b 4926 002
003	kommen	1448	003	+++	408 004 2246 005
004	sagen	1400	004	+++	419 003 2370 004
005	gehen	1238	005	+++	396 005 1533 008
006	wollen	1163	006	+++	o.A. 006:b 883 011
007	müssen	1105	007	+++	o.A. 007:c 3432 003
008	fahren	952	008	+++	264 011 405 019
009	sehen	833	009	+++	253 012 463 018
010	können	764	010	+++	o.A. 010 2131 007
011	spielen	724	011	+++	217 014 275 029
012	lachen	634	012	+++	23 ++ 41 ++
013	machen	631	013	+++	310 008 2192 006
014	heißen	451	014	+++	92 040 543 015
015	wissen	418	015	+++	122 026 1285 009
016	holen	330	016	+++	97 035 323 025
017	anhaben	328	017	+++	2 + 30 ++
018	kriegen:d	309	018	+++	39 ++ 780 012
019	laufen	305	019	+++	138 021 337 024
020	geben	286	020	+++	296 009 979 010
021	hinfallen	261	021	+++	11 ++ 17 ++
022	stehen	257	022	+++	224 013 241 034
023	liegen	255	023	+++	171 017 211 040
024	nehmen	251	024	+++	125 024 300 027
025	sollen	245	025	+++	o.A. 025:b 507 017

Lfd. Nr.	Lexem	Grundschrler				4.-Klssler				Vorschrlr				Erwachsene					
		An- zahl	Rang ++	+++hufg zahl	An- Rang ++	An- zahl	Rang ++	+++hufg zahl	An- Rang ++	+++mittel	+++selten	An- zahl	Rang ++	+++mittel	+++selten	An- zahl	Rang ++		
Verben 2																			
026	fallen	198	026	+++	105	030	69	++			051	wohnen	108	049	+++	49	079	83	088
027	lassen	193	027	+++		o.A.	031*b	360	023		052	schlafen	106	052	+++	71	057	55	++
028	dürfen	192	028	+++		o.A.	032*b	380	021		053	denken	101	053	+++	97	035	220	038
029	bekommen+d	187	029	+++	135	022	20	++			054	schreiben	96	054	+++	81	048	61	++
030	aussehen	176	030	+++	46	038	36	++			055	hinsetzen	95	055	+++	3	+	11	++
031	tun	173	031	+++	88	045	388	020			056	rausgehen	95	055	+++f	8	+	-	-
032	sitzen	170	032	+++	106	028	76	097			057	reingehen	95	055	+++f	3	+	-	-
033	werden	169	033	+++		o.A.	033*b	578	014		058	runterfallen	91	058	+++f	-	-	-	-
034	essen	168	034	+++	183	015	184	046			059	wehtun	91	058	+++	8	+	4	+
035	fragen	165	035	+++	171	017	126	060			060	anziehen	90	060	+++	20	++	25	++
036	fliegen	155	036	+++	78	052	8	+			061	beißen	90	060	+++	36	++	7	+
037	gucken	154	037	+++	26	++	205	041			062	rauskommen	90	060	+++	6	+	-	-
038	finden	149	038	+++	132	023	85	083			063	passieren	88	063	+++	27	++	94	077
039	erzählen	126	039	+++	30	++	36	++			064	mitnehmen	85	064	+++	15	++	85	083
040	bleiben	123	040	+++	93	039	201	043			065	spazierengehen	85	064	+++	19	++	2	+
041	bringen	123	040	+++	91	042	11	++			066	wiederkommen	85	064	+++	9	+	10	++
042	suchen	123	040	+++	70	060	67	++			067	glauben	84	067	+++	47	085	363	022
043	setzen	118	043	+++	62	062	83	088			068	füttern	83	068	+++	16	++	76	097
044	rufen	117	044	+++	125	024	51	++			069	fressen	82	069	+++	56	069	37	++
045	legen	115	045	+++	74	055	50	++			070	hören	82	069	+++	106	027	155	052
046	fangen	114	046	+++	50	078	21	++			071	trinken	81	071	+++	71	057	108	068
047	weglaufen	112	047	+++	6	+	3	+			072	ankommen	80	072	+++	36	++	85	083
048	bauen	109	048	+++	84	061	180	047			073	aufhaben	79	073	+++	5	+	6	+
049	anfangen	108	049	+++	48	081	264	032			074	hingehen	79	073	+++	7	+	69	++
050	springen	108	049	+++	77	054	31	++			075	aufstehen	78	075	+++	27	++	118	064

Lfd. Nr.	Lexem	Grundschüler An- zahl Gruppe	Vorschüler An- zahl Gruppe	4.-Klässler An- zahl Gruppe	Erwachsene An- zahl Gruppe
		++	++	++	++
076	mäßen	75	076	+++	86
077	bluten	67	077	+++	4
078	reinkommen	67	077	+++f	-
079	warten	66	079	+++	60
080	kaufen	65	080	+++	240
081	verstecken	65	080	+++	7
082	meinen	64	082	+++	319
083	weinen	64	082	+++	4
084	reiten	63	084	+++	32
085	tragen	63	084	+++	98
086	ziehen	63	084	+++	85
087	unterfahren	61	087 (he-)	++	-
088	stellen	61	087	+++	54
089	merken	60	089	+++	84
090	schimpfen	60	089	+++	29
091	sich freuen	60	089	+++	40
092	halten	58	092	+++	96
093	wegfahren	57	093	+++	22
094	reinfahren	56	094	+++f	-
095	weitergehen	56	094	+++	58
096	brechen	55	096	+++	42
097	rennen	55	096	+++	10
098	werfen	55	096	+++	10
099	weiterfahren	53	099	++	16
100	sprechen	52	100	+++	94
101	weggehen	52	100	++	29
weitere Verben <50 >100 (= Grundschüler)					
Lfd. Nr.	Lexem	Grundschüler An- zahl Gruppe	Vorschüler An- zahl Gruppe	4.-Klässler An- zahl Gruppe	Erwachsene An- zahl Gruppe
		++	++	++	++
102	schreiben	22	++	+++	84
103	bezahlen	10	++	+++	19
104	lesen	19	++	+++	45
105	singen	36	++	+++	102
106	brauchen	37	++	+++	272
107	laden	2	+	++	63
108	treffen	27	++	+++	37
109	helfen	45	++	+++	286
110	gehören	40	++	+++	99
111	leben	20	++	+++	96
112	vergessen	36	++	+++	24
113	üben	7	+	+++	9
114	nennen	39	++	+++	30
115	schwimmen	40	++	+++	7
116	schauen	41	++	++	11
117	zeigen	26	++	+++	32
118	fließen	5	+	++	6
119	schlagen	44	++	+++	46
120	antworten	22	++	+++	3
121	öffnen	11	++	+++	1
122	betrügen	1	+	-	-
123	fassen	8	+	+	12
124	gießen	6	+	++	-
125	schließen	2	+	++	8

Häufigkeitsverben (Rang < 101) bei den Viertklässlern, die nicht unter den 100 häufigsten Verben bei den Grundschulern verzeichnet sind.

Lfd. Nr.	Lexem	Grundschüler An- zahl	Grundschüler Rang ++	Vorschüler +++häufig An- zahl	Vorschüler Rang ++	4.-Klässler An- zahl	4.-Klässler Rang ++	Erwachsene An- zahl	Erwachsene Rang ++
Verben 5		++		++		++		++	
126	stoßen	24	++	+++	47	085	1	+	
127	verbieten	2	+	+++	47	085	10	++	
126	verkaufen	21	++	+++	45	091	125	061	
127	messen	3	+	+++	44	092	14	++	
128	schenken	31	++	+++	44	092	8	+	
129	waschen	28	++	+++	44	092	19	++	
130	brennen	15	++	+++	43	095	79	094	
131	beginnen	1	+	++	42	096	4	+	
132	knallen	18	++	++	42	096	1	+	
133	öffnen	11	++	+++	42	096	1	+	
134	schieben	22	++	+++	42	096	10	++	
135	hassen	-	-	-	41	100	3	+	
Weitere Verben (= Viertklässler)									
Verben über Rangplatz 100 (= Erwachsene)									
								<75	>100

Lfd. Nr.	Lexem	Grundschüler An- zahl	Grundschüler Rang ++	Vorschüler +++häufig An- zahl	Vorschüler Rang ++	4.-Klässler An- zahl	4.-Klässler Rang ++	Erwachsene An- zahl	Erwachsene Rang ++
Adjektive 1		++		++		++		++	
Summe Adjektive 14783		799fg		13955		1462+			
001	ganz	1338	001	+++	143	008	411	006	
002	rot	1025	002	+++	102	014	38	067	
003	klein	809	003	+++	224	003	391	007	
004	blau	558	004	+++	42	040	12	++	
005	schwarz	550	005	+++	61	023	27	087	
006	ander(er)	608	006	+++	o.A. 006sb		138	018	
007	groß	532	007	+++	296	001	615	003	
008	schön	518	008	+++	157	005	517	004	
009	weiß	448	009	+++	66	020	52	045	
010	schnell	413	010	+++	134	010	172	013	
011	lang, -e	405	011	++	178	004	483	005	
012	viel	355	012	+++	o.A. 012sb		104	023	
013	gelb	321	013	+++	39	043	7	+	
014	braun	261	014	+++	34	053	2	+	
015	gut	217	015	+++	246	002	1052	001	
016	golden	191	016	+++	11	++	9	+	
017	alt	154	017	+++	150	007	681	002	
018	richtig	154	017	++	61	028	152	016	
019	dick	152	019	+++	51	034	39	063	
020	kaputt	143	020	+++	17	++	57	042	
021	grün	141	021	+++	64	022	17	++	
022	beide	135	022	+++	57	029	o.A. sh		
023	lustig	126	023	+++	54	032	35	075	
024	doll	121	024	++	-	-	-	-	
025	grau	116	025	+++	18	++	1	+	

Lfd. Nr.	Lexem	Grundschüler				Vorschüler				4.-Klässler				Erwachsene					
		Anzahl	Rang	Gruppe	++	+	+++	häufig	Anzahl	Rang	Gruppe	++	+	Anzahl	Rang	Gruppe	++	+	
Adjektive 2		Anzahl		Rang		Gruppe		++		+		+++		Anzahl		Rang		Gruppe	
026	nächste,r	108	026	44	039	9	+	+++	051	rechter/rechts	42	051	+++	50	027	9	+		
027	voll	104	027	57	029	92	027	+++	052	rund	41	052	+++	19	++	19	++		
028	genau	99	028	23	082	132	020	+++	053	krumm	38	053	+++	6	+	7	+		
029	mehr/meiste(n)	99	028	143	008	o.A.	zh	+++	054	fest	37	054	+++	44	039+n	66	035		
030	blond	94	030	2	+	-	-	+++	055	hellblau,fj	37	054	++	-	-	-	-		
031	neu	94	030	116	011	134	019	+++	056	orange	36	056	-	8	++n	-	-		
032	komisch	92	032	11	++	10	++	+++	057	voriger	36	056	++	1	+	-	-		
033	hoch	91	033	107	013	120	021	+++	058	dünn	35	058	+++	18	++	3	+		
034	rosa	78	034	5	+	1	+	+++	059	langsam	34	059	+++	29	062	58	041		
035	spitz	77	035	17	++	1	+	+++	060	krank	34	059	+++	29	062	80	029		
036	tot	70	036	10	++	24	095	+++	061	nett	33	061	+++	12	++	82	028		
037	fertig	62	037	31	057	236	008	+++	062	natürlich	32	062	+++	35	047	3	+		
038	laut	62	037	54	030	5	+	+++	063	offen	31	063	+++	22	086	16	++		
039	weit	60	039	70	019	189	012	+++	064	schwer	29	064	+++	56	031	158	014		
040	lieb	56	040	96	015	62	038	+++	065	dunkel	28	065	+++	16	++	7	+		
041	naß	56	040	23	082	51	051	+++	066	kariert	28	065	++	-	-	-	-		
042	besser/beste,r	53	042	81	017	o.A.	zi	+++	067	silbern	28	065	++	5	+	1	+		
043	jung	52	043	42	041	224	009	+++	068	tief	28	065	+++	31	057	26	087		
044	halb	51	044	37	048	o.A.	zi	+++	069	tüchtig	28	065	++	8	+	11	++		
045	spät	50	045	50	036	77	030	+++	070	nah	27	070	+++	27	070	24	092		
046	böse	47	046	38	046	41	058	+++	071	letzte,r	26	071	+++	46	037	154	015		
047	lauter	47	046	1	+	o.A.	zi	+++	072	froh	26	071	+++	14	++	85	037		
048	schlimm	46	048	10	++	52	045	+++	073	hell	25	073	+++	19	++	14	++		
049	einfach	43	049	28	067	72	032	+++	074	kalt	25	073	+++	40	043	41	058		
050	kurz	43	049	61	023	54	044	+++	075	lieber/liebsten	24	075	++	5	+	62	038		

Weitere Adjektive <15 >101 (= Grundschüler)									
Lfd. Nr.	Lexem	Grundschüler An- zahl Gruppe	Vorschüler An- zahl Gruppe	4.-Klässler An- zahl Gruppe	Erwachsene An- zahl Gruppe				
Adjektive 4									
076	hellbraun	22	076	++	2	+	-	-	-
077	traurig	22	076	+++	34	053	3	+	+
078	bunt	20	078	+++	38	046	9	+	+
079	dumm	20	078	+++	95	016	46	055	
080	dunkelblau	20	078	++	1	+	-	-	-
081	frech	20	078	+++	14	++	5	+	+
082	plötzlich	20	078	+++	65	021	3	+	+
083	verletzt	20	078	-	-	-	-	-	-
084	blau-weiß	19	084	-	-	-	-	-	-
085	echt	19	085	+++	6	+	5	+	+
086	leer	19	085	+++	17	++	34	078	
087	niedlich	18	087	+++	7	+	-	-	-
088	warm	18	087	+++	25	077	36	073	
089	dunkelbraun	17	089	-	2	+	-	-	-
090	glücklich	17	089	++	14	++	12	++	
091	ruhig	17	089	+++	21	092	61	039	
092	schief	17	089	+++	5	+	3	+	+
093	steil	17	089	+++	2	+	43	057	
094	breit	16	094	+++	26	070	39	063	
095	furchtbar	16	094	+++	4	+	-	-	-
096	lose	16	094	-	2	+	-	-	-
097	scharf	16	094	+++	28	066	14	++	
098	wütend	16	094	+++	20	094	1	+	+
099	fröh	15	099	+++	24	080	56	043	
100	glatt	15	099	+++	12	++	12	++	
101	still	15	099	+++	20	094	7	+	+
102	wenig	15	099	+++	53	034	73	031	

Häufigkeitsadjektive (Rang < 101) bei den Viertklässlern, die nicht unter den 100 häufigsten Adjektiven der Grundschulern verzeichnet sind.									
Lfd. Nr.	Lexem	Grundschüler An- zahl Gruppe	Vorschüler An- zahl Gruppe	4.-Klässler An- zahl Gruppe	Erwachsene An- zahl Gruppe				
Adjektive 5									
103	gleich	9	++	++	71	018	103	024	
104	spannend	3	+	+++	61	023	-	-	
105	links/linke, r	14	++	+++	61	023	9	+	
106	frei	11	++	??	46	037	67	034	
107	wichtig	3	3	+++	40	043	25	091	
108	stark	11	++	+++	36	049	96	026	
109	bestimmt	1	+	+++	35	050	20	++	
110	männlich	-	-	-	35	050	3	+	
111	schlecht	13	++	??	33	055	112	022	
112	herzlich	1	+	++	32	056	-	-	
113	leicht	11	++	++	31	057	66	035	
114	heiß	6	+	+++	30	060	35	075	
115	süß	5	+	+++	30	060	4	+	
116	dicht	11	++	??	29	062	1	+	
117	gesund	10	++	+++	29	062	61	039	
118	freundlich	4	+	++	28	066	7	+	
119	weiblich	1	+	-	28	066	2	+	
120	fröhlich	1	+	+++	26	071	6	+	
121	gar	2	+	++	26	071	1	+	
122	sächlich	-	-	-	26	071	-	-	
123	teuer	2	+	+++	26	071	41	058	
124	wohl	1	+	+++	26	071	10	++	
125	frisch	7	+	+++	25	077	56	043m	

Lfd. Nr.	Lexem	Grundschüler An- zahl Gruppe	Vorschüler +++häufig ++mittel ++seiten	4.-Klässler An- zahl Gruppe	Erwachsene An- zahl Gruppe
	Adjektive 5				
126	wild	13	++	25 077	13 ++
127	falsch	13	++	23 081	7 +
128	fleißig	1	+	23 081	11 ++
129	gefährlich	10	++	23 081	51 048
130	deutlich	1	+	22 086	1 +
131	eigentlich	22	+++1	22 086	6 +
132	geboren	1	+	22 086	2 +
133	klug	3	+	22 086	- -
134	sicher	12	++	22 086	24 092
135	kräftig	6	+	21 092	5 +
136	schmutzig	13	++	20 094	4 +
137	toll	8	+	20 094	4 +
138	weich	10	++	20 094	8 +
139	leise	14	++	19 098	2 +
140	sauber	10	++	19 098	50 051
141	vorsichtig	7	+	19 098	5 +
Weitere Adjektive (= Viertklässler)					<19 >101
Adjektive über Rangplatz >100 (= Erwachsene					<22 >102

Anmerkungen

- 1a) Häufigkeit unbestimmbar, da Rosen und Verb bei AUGST (1985) nicht getrennt notiert sind.
- 1b) Bei AUGST (1985) ohne Häufigkeitsangaben als Häufigkeitsverfälschter angegeben. Um eine Vergleichbarkeit mit den Korpora von PEGEL & EICHERT (1987) sowie auch BUDFF (1981) zu erreichen, wurden diese Verben in der Übersicht den Rangplätze nach den Angaben von PEGEL & EICHERT gleichgesetzt.
- 1c) Angaben für BUDFF (1981) ohne Hilfverben.
- 1d) Das Verb bezaubern (Rang 18) ist synonym mit den magischsprachlichen Begriffen (Rang 18)
- 1e) Die Kennzeichnung "1+" bedeutet, daß es mehrere nach-semantische Belege für das betreffende Lexem gibt, die jedoch jeweils nicht von der Mehrheit der Kinder gebraucht werden. Aufgrund der Notationsform von AUGST (1984) ist nicht zu bestimmen, von wievielen Kindern das Verben insgesamt verwendet wurde.
- 1f) Bei AUGST (1984) ausnahmslos aufzeichnungsbefreiung mit der Vorbildbe- notiert.
- 1g) Aus der Häufigkeitsliste von PEGEL & EICHERT (1987, 43) wurden die Zahlwörter erster, zweiter, dritter sowie der Zweifelsfall alle herausgenommen, da sie bei den anderen Korpora nicht mit Häufigkeitsangaben unter der Wortart Adjektive notiert sind.
- 1h) Bei BUDFF (1981) als andere Wortart ohne Häufigkeitsangabe notiert.
- 1i) Bei BUDFF (1981) nicht angeführte Wortform.
- 1j) In den anderen Korpora sind jedoch verschiedene hell- und dunkel-Farben sowie Doppel-Farben verzeichnet.
- 1k) Bei PEGEL & EICHERT (1987, 434) fälschlich als breit notiert (Beleg in der vollständigen alphabetischen Liste).
- 1l) Bei PEGEL & EICHERT (1987, 432) in der Sonderliste: Zweifelsfälle verzeichnet.
- 1m) Zusammenfassung von zwei bei BUDFF (1981) getrennten semantischen Varianten.
- 1n) Nur als Adjektiv gewertet, bei AUGST (1984) undifferenziert als Adjektiv/Nomen notiert.

Anh. 1: B 1.3b:

Tab. 6a-B 1.3: Distraktorenähnlichkeiten Nomen Pilotversion					
Nr.	Item	Distraktor	Gebärde	Mundbild / Schrift	Inhalt / Zeichnung
1	Mutter	Bauer	X		
		Mädchen			X
		Motte		X	
2	Auto	Bus			X
		Laster	X		X
		Zug			X
3	Haus	Hand		X	
		Kirche	X		X
		Maus		X	
4	Kopf	Junge	X		
		Topf		X	
		Zopf		X	X
5	Straße	Blase		X	
		Strauß		X	
		Weg	X		X
6	Bett	Brett		X	
		Decke	X		X
		Sofa			X
7	Angst	Freude			X
		Trauer	X		X
		Wut	X		X
8	Stein	Kaffe	X		
		Stange		X	
		Stirn		X	
9	Tier	Bier		X	
		Tor		X	
		Verkehr	X		
10	Bart	Bett		X	
		Kinn	X		X
		Post	X	X	
11	Zahn	Hahn		X	
		Mund	X		X
		Zaun		X	
12	Leute	Herde			X
		Mensch	X		X
		Paar			X
13	Zahl	Buchstabe			X
		Form			X
		Zeichen		X	X
14	Geld	Feld		X	
		Gold		X	X
		Salz	X		

Nr.	Item	Distraktor	Gebärde	Mundbild / Schrift	Inhalt / Zeichnung
15	Kasten	Kasse		X	X
		Kissen		X	
		Schrank	X	X	X
16	Holz	Brot	X	X	
		Gold		X	
		Höhle			
17	Arbeit	Abend		X	
		Eimer		X	X
		Kaffee	X		
18	Schloss	Haus			X
		Kirche	X		X
		Schlüssel		X	X

Tab. 6b-B 1.3: Distraktorenähnlichkeiten Verben Pilotversion					
Nr.	Item	Distraktor	Gebärde	Mundbild / Schrift	Inhalt / Zeichnung
1	sagen	fragen	X	X	X
		schreien	X		X
		schweigen	X		X
2	gehen	nagen	X		
		sehen		X	
		stehen	X	X	X
3	fahren	fliegen			X
		frieren		X	
		schenken	X		
4	essen	schmieren			X
		trinken	X		X
		zähneputzen	X		X
5	gucken	backen		X	
		ducken		X	
		küssen		X	
6	fliegen	fragen		X	
		liegen		X	
		winken	X		
7	schlafen	liegen			X
		schlagen		X	
		träumen			X
8	denken	danken		X	
		lenken		X	
		schenken		X	
9	hinsetzen	anspitzen		X	
		schwitzen		X	
		sitzen	X	X	X

Nr.	Item	Distraktor	Gebärde	Mundbild / Schrift	Inhalt / Zeichnung
10	warten	braten		X	
		halten		X	
		wandern	X	X	
11	verstecken	erschrecken		X	X
		schlafen	X		
		stecken		X	
12	weinen	regnen			X
		schwitzen			X
		warnen		X	
13	schreiben	bohren	X		
		reiben	X	X	
		schlafen		X	
14	lesen	blasen		X	
		fegen		X	
		halten	X		
15	singen	geigen			X
		sagen		X	X
		schweigen			X
16	waschen	wanken		X	
		warten		X	
		wischen	X	X	
17	brennen	brechen		X	
		rechnen	X	X	
		rennen		X	
18	schieben	heben			X
		schließen	X	X	X
		ziehen			X

Tab. 6c-B 1.3: Distraktorenähnlichkeiten Adjektive Pilotversion					
Nr.	Item	Distraktor	Gebärde	Mundbild / Schrift	Inhalt / Zeichnung
1	ganz	gerade		X	X
		halb			X
		rund	X		X
2	klein	groß			X
		klug	X	X	
		schlecht	X		
3	blau	braun		X	X
		grau	X		X
		schwarz			X
4	voll	falsch		X	
		stolz	X		
		viel	X	X	X

Nr.	Item	Distraktor	Gebärde	Mundbild / Schrift	Inhalt / Zeichnung
5	blond	braun			X
		grau			X
		rot			X
6	komisch	gemischt		X	
		kurz			X
		neu	X		
7	krumm	kaputt	X		
		krank		X	
		stumm		X	
8	hellblau	dunkelblau	X	X	X
		hellgrau	X	X	X
		hellgrün	X	X	X
9	orange	lila			X
		rosa	X	X	X
		rot			X
10	traurig	ernst			X
		erstaunt		X	X
		fröhlich			X
11	verletzt	besetzt		X	
		gesund			X
		sauber	X		
12	warm	heiß	X		X
		kalt			X
		weich		X	
13	glatt	platt		X	X
		satt		X	
		sauber	X		
14	links	oben			X
		rechts	X		X
		unten			X
15	frei	erschrocken	X		
		frech	X	X	
		klein		X	
16	schmutzig	sauber			X
		zerknüllt			X
		zerrissen			X
17	weich	wach		X	X
		warm		X	
		weiß		X	
18	leise	heiß		X	
		leicht		X	
		links		X	

Anh. 1: B 2.3.1:

Tab. 7a-B 2.3.1: Klasseninterner Vergleich der Leistungen Gehörloser											
Klasse 3											
Gebärdensprache Nomen			Gebärdensprache Verben			Gebärdensprache Adjektive			Gebärdensprache gesamt		
Pkte	%	Name	Pkte	%	Name	Pkte	%	Name	Pkte	%	Name
18	100	A	17	94	A	13	72	A	48	89	A
18	100	F	17	94	F	13	72	F	48	89	F
18	100	V*	17	94	M	12	67	M	46	85	M
17	94	M*	16	89	V	11	61	V	45	83	V
17	94	T	16	89	T	11	61	T	44	81	T
16	89	O	14	78	O	9	50	O	39	72	O
Schriftsprache Nomen			Schriftsprache Verben			Schriftsprache Adjektive			Schriftsprache gesamt		
Pkte	%	Name	Pkte	%	Name	Pkte	%	Name	Pkte	%	Name
18	100	F	14	78	A	15	83	F	46	85	F
17	94	M	13	72	F	13	72	M	42	78	M
15	83	A	12	67	M	11	61	A	40	74	A
11	61	T	10	56	T	8	44	T	29	54	T
11	61	V	9	50	V	6	33	V	26	48	V
3	17	O	5	28	O	3	17	O	11	20	O
Lautsprache Nomen			Lautsprache Verben			Lautsprache Adjektive			Lautsprache gesamt		
Pkte	%	Name	Pkte	%	Name	Pkte	%	Name	Pkte	%	Name
14	78	A	14	78	V*	10	56	F	36	67	F
13	72	F	13	72	F	8	44	A	30	56	A
13	72	M	11	61	M	6	33	M	30	56	M
11	61	T	8	44	A*	5	28	V	30	56	V
11	61	V	8	44	T	4	22	T	23	43	T
5	28	O	6	33	O	3	17	O	14	26	O

Tab. 7b-B 2.3.1: Klasseninterner Vergleich der Leistungen Gehörloser											
Klasse 4											
Gebärdensprache Nomen			Gebärdensprache Verben			Gebärdensprache Adjektive			Gebärdensprache gesamt		
Pkte	%	Name	Pkte	%	Name	Pkte	%	Name	Pkte	%	Name
18	100	K	17	94	F*	15	83	K	49	91	K
18	100	L	16	89	K	13	72	L	47	87	L
16	89	F	16	89	L	11	61	F	44	81	F
16	89	S	15	83	S	11	61	S	39	72	S
14	78	E	14	78	E	8	44	E	39	72	E
Schriftsprache Nomen			Schriftsprache Verben			Schriftsprache Adjektive			Schriftsprache gesamt		
Pkte	%	Name	Pkte	%	Name	Pkte	%	Name	Pkte	%	Name
17,25	99	K	13,75	76	K	14	78	L	43	80	K
14	78	L	13	72	L	12	67	K	41	76	L
12	67	F	11	61	S	7	39	F	29	54	F
11	61	S	10	56	F	3	17	S	25	46	S
6	33	E	8	44	E	3	17	E	17	31	E
Lautsprache Nomen			Lautsprache Verben			Lautsprache Adjektive			Lautsprache gesamt		
Pkte	%	Name	Pkte	%	Name	Pkte	%	Name	Pkte	%	Name
13,25	74	S*	16,5	92	K	12,75	71	K	41,25	76	K
12	67	K	13	72	L	11	61	L	35	65	L
11	61	L*	12,25	68	S	7	39	S	32,5	60	S
9	50	E	8	44	E	3	17	E	20	37	E
7	39	F	6	33	F	3	11	F	15	28	F

Anh. 1: B 2.3.2:

Tab. 8a-B 2.3.2: Klasseninterner Vergleich der Zeiten Gehörloser Klasse 3							
Gebärdensprache Nomen		Gebärdensprache Verben		Gebärdensprache Adjektive		Gebärdensprache gesamt	
Zeit	Name	Zeit	Name	Zeit	Name	Zeit	Name
2 : 07	A	1 : 51	M	2 : 19	M	6 : 43	M
2 : 17	O	2 : 14	A	2 : 30	A	6 : 51	A
2 : 29	F	2 : 26	F	2 : 39	T	7 : 44	F
2 : 33	M	2 : 30	O	2 : 49	F	7 : 50	O
2 : 36	T	2 : 37	T	3 : 03	O	7 : 52	T
2 : 43	V	2 : 55	V	3 : 20	V	8 : 58	V
Schriftsprache Nomen		Schriftsprache Verben		Schriftsprache Adjektive		Schriftsprache gesamt	
Zeit	Name	Zeit	Name	Zeit	Name	Zeit	Name
1 : 37	T	1 : 40	T	2 : 04	T	5 : 21	T
1 : 42	M	2 : 20	M	2 : 06	O	6 : 23	M
2 : 06	F	2 : 31	F	2 : 21	M	7 : 05	F
2 : 09	A	2 : 33	O	2 : 28	F	7 : 09	O
2 : 28	V	2 : 34	A	2 : 30	V	7 : 32	V
2 : 30	O	2 : 34	V	3 : 17	A	8 : 00	A
Lautsprache Nomen		Lautsprache Verben		Lautsprache Adjektive		Lautsprache gesamt	
Zeit	Name	Zeit	Name	Zeit	Name	Zeit	Name
2 : 22	T	2 : 32	T	2 : 01	A	7 : 31	A
2 : 25	V	2 : 46	V	2 : 12	M	7 : 42	M
2 : 37	F	2 : 50	A	2 : 42	O	8 : 02	V
2 : 38	M	2 : 51	F	2 : 51	V	8 : 12	T
2 : 40	O	2 : 52	M	3 : 16	F	8 : 44	F
3 : 10	A	3 : 34	O	3 : 18	T	9 : 26	O

Tab. 8b-B 2.3.2: Klasseninterner Vergleich der Zeiten Gehörloser Klasse 4							
Gebärdensprache Nomen		Gebärdensprache Verben		Gebärdensprache Adjektive		Gebärdensprache gesamt	
Zeit	Name	Zeit	Name	Zeit	Name	Zeit	Name
1 : 55	K	1 : 47	E	2 : 19	L	6 : 33	L
2 : 04	L	1 : 57	K	2 : 27	E	6 : 39	K
2 : 28	S	2 : 10	L	2 : 47	K	6 : 43	E
2 : 29	E	2 : 27	F	3 : 29	F	9 : 03	F
1 : 07	F	2 : 38	S	4 : 22	S	9 : 28	S
Schriftsprache Nomen		Schriftsprache Verben		Schriftsprache Adjektive		Schriftsprache gesamt	
Zeit	Name	Zeit	Name	Zeit	Name	Zeit	Name
2 : 31	E	1 : 57	E	2 : 15	E	6 : 43	E
2 : 40	F	2 : 12	F	2 : 21	L	7 : 51	K
2 : 42	K	2 : 25	K	2 : 44	K	8 : 00	F
3 : 05	L	2 : 45	L	3 : 08	F	8 : 11	L
3 : 24	S	2 : 47	S	3 : 51	S	10 : 02	S
Lautsprache Nomen		Lautsprache Verben		Lautsprache Adjektive		Lautsprache gesamt	
Zeit	Name	Zeit	Name	Zeit	Name	Zeit	Name
2 : 01	E	1 : 58	E	1 : 56	E	5 : 55	E
2 : 37	L	2 : 20	L	2 : 52	L	7 : 49	L
2 : 42	K	2 : 35	K	3 : 13	F	8 : 33	K
3 : 43	F	2 : 55	F	3 : 16	K	9 : 51	F
3 : 49	S	4 : 33	S	4 : 44	S	13 : 06	S

Anh. 1: B 2.3.3:

Tab. 9a-B 2.3.3: Gesamtergebnisse Gehörloser Klasse 3											
Gebärdensprache			Schriftsprache			Lautsprache			gesamt		
Pkte	%	Name	Pkte	%	Name	Pkte	%	Name	Pkte	%	Name
48	89	F	46	85	F	36	67	F	130	80	F
48	89	A	42	78	M	30	56	A	118	73	A
46	85	M	40	74	A	30	56	M	118	73	M
45	83	V	29	54	T	30	56	V	101	62	V
44	81	T	29	48	V	23	43	T	96	59	T
39	72	O	11	20	O	14	26	O	64	40	O
Zeit		Name	Zeit		Name	Zeit		Name	Zeit		Name
6 : 43		M	5 : 21		T	7 : 31		A	20 : 47		M
6 : 51		A	6 : 23		M	7 : 42		M	21 : 25		T
7 : 44		F	7 : 05		F	8 : 02		V	22 : 22		A
7 : 50		O	7 : 09		O	8 : 12		T	23 : 33		F
7 : 52		T	7 : 32		V	8 : 44		F	24 : 25		O
8 : 58		V	8 : 00		A	9 : 26		O	24 : 32		V

Tab. 9b-B 2.3.3: Gesamtergebnisse Gehörloser Klasse 4											
Gebärdensprache			Schriftsprache			Lautsprache			gesamt		
Pkte	%	Name	Pkte	%	Name	Pkte	%	Name	Pkte	%	Name
49	91	K	43	80	K	41,25	76	K	133,25	82	K
47	87	L	41	76	L	35	65	L	123	76	L
44	81	F	29	54	F	32,5	60	S	96,5	60	S
39	72	E	25	46	S	20	37	E	88	54	F
39	72	S	17	31	E	15	28	F	76	47	E
Zeit		Name	Zeit		Name	Zeit		Name	Zeit		Name
6 : 33		L	6 : 43		E	5 : 55		E	19 : 21		E
6 : 39		K	7 : 51		K	7 : 49		L	23 : 03		K
6 : 43		E	8 : 00		F	8 : 33		K	22 : 22		L
9 : 03		F	8 : 11		L	9 : 51		F	26 : 54		F
9 : 28		S	10 : 02		S	13 : 06		S	32 : 36		S

Anh. 1: B 2.3.4:

Die folgenden Tabellen beschreiben die Summe der Antworten aller hörenden Testpersonen zu jeder Aufgabe. Zu jeder Aufgabennummer gehören jeweils drei Zeilen für die Sprachmodalitäten Gebärdensprache (GS), Schriftsprache (SS) und Lautsprache (LS). Spaltenweise hintereinander werden die Wahlhäufigkeiten des Items selbst, der drei Distraktoren und schließlich der fehlenden Antworten aufgeführt. In der ersten Tabelle befindet sich die Darstellung der Nomen, in der zweiten die der Verben und in der dritten die der Adjektive.

Tab. 10a-B 2.3.4: Ergebnisse/Fehleranalyse Hörender Nomen						
Nr.	Item	Distr. I	Distr. II	Distr. III	weiß nicht	nicht beantw.
1	Mutter	Mädchen	Bauer	Motte	w. n.	fehlt
SS	79	1	1	-	-	-
LS	83	2	-	-	-	1
2	Auto	Bus	Laster	Zug	w. n.	fehlt
SS	80	1	-	-	-	-
LS	85	-	1	-	-	-
3	Haus	Kirche	Hand	Maus	w. n.	fehlt
SS	79	1	-	1	-	-
LS	84	2	-	-	-	-
4	Kopf	Zopf	Junge	Topf	w. n.	fehlt
SS	71	9	1	-	-	-
LS	83	3	-	-	-	-
5	Straße	Weg	Strand	Strauß	w. n.	fehlt
SS	80	-	-	1	-	-
LS	85	1	-	-	-	-
6	Bett	Decke	Brett	Sofa	w. n.	fehlt
SS	80	-	1	-	-	-
LS	85	-	-	-	-	-
7	Angst	Wut	Trauer	Freude	w. n.	fehlt
SS	77	2	2	-	-	-
LS	84	2	-	-	-	-
8	Stein	Stirn	Kaffee	Stange	w. n.	fehlt
SS	80	4	-	-	-	-
LS	85	-	1	-	-	-
9	Tier	Bier	Tor	Verkehr	w. n.	fehlt
SS	81	-	1	2	-	-
LS	86	-	-	-	-	-

Nr.	Item	Distr. I	Distr. II	Distr. III	weiß nicht	nicht beantw.
10	Bart	Bett	Kinn	Post	w. n.	fehlt
SS	77	-	-	3	-	-
LS	85	1	-	-	-	-
11	Zahn	Mund	Hahn	Zange	w. n.	fehlt
SS	80	-	1	-	-	-
LS	86	-	-	-	-	-
12	Leute	Paar	Mensch	Herde	w. n.	fehlt
SS	70	-	8	3	-	-
LS	84	-	-	1	1	-
13	Zahl	Form	Buchstabe	Zeichen	w. n.	fehlt
SS	79	1	1	-	-	-
LS	85	-	1	-	-	-
14	Geld	Gold	Salz	Feld	w. n.	fehlt
SS	80	1	-	-	-	-
LS	67	19	-	-	-	-
15	Kasten	Kasse	Kissen	Schrank	w. n.	fehlt
SS	67	13	1	-	-	-
LS	84	2	-	-	-	-
16	Holz	Höhle	Brot	Gold	w. n.	fehlt
SS	81	-	-	-	-	-
LS	84	-	1	1	-	-
17	Arbeit	Eimer	Abend	Kaffee	w. n.	fehlt
SS	81	-	-	-	-	-
LS	85	1	-	-	-	-
18	Schloß	Schlüssel	Kirche	Haus	w. n.	fehlt
SS	65	-	7	8	-	-
LS	65	7	14	-	-	-

Tab. 10b-B 2.3.4: Ergebnisse/Fehleranalyse Hörender Verben						
Nr.	Item	Distr. I	Distr. II	Distr. III	weiß nicht	nicht beantw.
1	sagen	fragen	schweigen	schreien	w. n.	fehlt
SS	61	13	4	2	1	-
LS	54	7	2	22	1	-
2	gehen	stehen	nagen	sehen	w. n.	fehlt
SS	81	-	-	-	-	-
LS	86	-	-	-	-	-

Nr.	Item	Distr. I	Distr. II	Distr. III	weiß nicht	nicht beantw.
3	fahren	fliegen	schenken	frieren	w. n.	fehlt
SS	81	-	-	-	-	-
LS	86	-	-	-	-	-
4	essen	trinken	zähneputzen	schmieren	w. n.	fehlt
SS	78	-	2	1	-	-
LS	77	-	5	4	-	-
5	gucken	backen	küssen	ducken	w. n.	fehlt
SS	76	-	3	2	-	-
LS	78	1	-	7	-	-
6	fliegen	fragen	liegen	winken	w. n.	fehlt
SS	81	-	-	-	-	-
LS	86	-	-	-	-	-
7	schlafen	schlagen	träumen	liegen	w. n.	fehlt
SS	76	-	4	-	1	-
LS	78	-	-	8	-	-
8	denken	danken	lenken	schenken	w. n.	fehlt
SS	79	2	-	-	-	-
LS	85	-	1	-	-	-
9	hinsetzen	sitzen	anspitzen	schwitzen	w. n.	fehlt
SS	52	29	-	-	-	-
LS	80	6	-	-	-	-
10	warten	halten	wandern	braten	w. n.	fehlt
SS	79	1	-	-	1	-
LS	84	-	2	-	-	-
11	verstecken	schlafen	erschrecken	stecken	w. n.	fehlt
SS	73	5	-	2	1	-
LS	84	-	2	-	-	-
12	weinen	regnen	schwitzen	warnen	w. n.	fehlt
SS	78	-	3	-	-	-
LS	80	1	5	-	-	-
13	schreiben	bohren	reiben	schlafen	w. n.	fehlt
SS	76	-	5	-	-	-
LS	86	-	-	-	-	-
14	lesen	fegen	halten	blasen	w. n.	fehlt
SS	80	-	1	-	-	-
LS	85	1	-	-	-	-
15	singen	sagen	schweigen	geigen	w. n.	fehlt
SS	80	1	-	-	-	-
LS	84	-	1	1	-	-

Nr.	Item	Distr. I	Distr. II	Distr. III	weiß nicht	nicht beantw.
16	waschen	wischen	wanken	warten	w. n.	fehlt
SS	78	3	-	-	-	-
LS	81	4	1	-	-	-
17	brennen	brechen	rechnen	rennen	w. n.	fehlt
SS	79	-	-	1	1	-
LS	82	1	-	3	-	-
18	schieben	schließen	heben	ziehen	w. n.	fehlt
SS	77	-	-	4	-	-
LS	79	1	2	4	-	-

Tab. 10c-B 2.3.4: Ergebnisse/Fehleranalyse Hörender Adjektive						
Nr.	Item	Distr. I	Distr. II	Distr. III	weiß nicht	nicht beantw.
1	ganz	rund	gerade	halb	w. n.	fehlt
SS	73	-	6	1	1	-
LS	83	3	-	-	-	-
2	klein	klug	groß	schlecht	w. n.	fehlt
SS	77	-	4	-	-	-
LS	79	1	4	1	1	-
3	blau	grau	braun	schwarz	w. n.	fehlt
SS	81	-	-	-	-	-
LS	86	-	-	-	-	-
4	voll	viel	falsch	stolz	w. n.	fehlt
SS	62	27	11	6	-	-
LS	72	10	-	3	1	-
5	blond	braun	rot	grau	w. n.	fehlt
SS	79	-	1	1	-	-
LS	81	3	-	2	-	-
6	komisch	gemischt	kurz	neu	w. n.	fehlt
SS	77	1	-	-	1	-
LS	63	-	20	2	1	-
7	krumm	kaputt	krank	stumm	w. n.	fehlt
SS	79	-	-	1	1	-
LS	79	1	-	5	1	-
8	hellblau	dunkelblau	hellgrau	hellgrün	w. n.	fehlt
SS	79	1	-	1	-	-
LS	86	-	-	-	-	-

Nr.	Item	Distr. I	Distr. II	Distr. III	weiß nicht	nicht beantw.
9	orange	rosa	lila	rot	w. n.	fehlt
SS	78	1	-	2	-	-
LS	84	1	1	-	-	-
10	traurig	ernst	erstaunt	fröhlich	w. n.	fehlt
SS	78	3	-	-	-	-
LS	86	-	-	-	-	-
11	verletzt	besetzt	gesund	sauber	w. n.	fehlt
SS	81	-	-	-	-	-
LS	83	-	3	-	-	-
12	warm	heiß	kalt	weich	w. n.	fehlt
SS	45	25	10	-	1	-
LS	7	76	1	1	1	-
13	glatt	sauber	satt	platt	w. n.	fehlt
SS	58	5	5	13	-	-
LS	28	8	-	48	2	-
14	links	rechts	oben	unten	w. n.	fehlt
SS	64	13	3	-	1	-
LS	34	31	17	4	-	-
15	frei	erschrocken	frech	klein	w. n.	fehlt
SS	74	2	-	5	-	-
LS	82	2	2	-	-	-
16	schmutzig	sauber	zerknüllt	zerrissen	w. n.	fehlt
SS	68	1	3	11	-	-
LS	75	-	2	9	-	-
17	weich	weiß	wach	warm	w. n.	fehlt
SS	75	4	-	1	1	-
LS	64	-	18	4	-	-
18	leise	heiß	leicht	links	w. n.	fehlt
SS	73	2	1	5	-	-
LS	69	-	14	1	2	-

Anh. 1: B 2.4.1:

Tab. 11-B 2.4.1: Fehlerverteilung Gehörloser				
	Gebärdenspr.	Schriftspr.	Lautspr.	ges.
Nomen	12	63,5	79,5	155
%	6	32	40	26
Verben	22	79,25	83,5	184,75
%	11	40	42	31
Adjektive	71	103	129,25	303,25
%	36	52	65	51
ges.	105	245,75	292,25	643
%	18	41	49	36

Anh. 1: B 2.4.2:

Tab. 12a-B 2.4.2: Rating Anna Klasse 3

GS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös						X	ruhig
spielerisch						X	ernst

SS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös						X	ruhig
spielerisch						X	ernst

LS	1	2	3	4	5	6	
flüchtig					X		konzentriert
lustlos					X		motiviert
nervös						X	ruhig
spielerisch						X	ernst

1 2 3 4 5 6

Tab. 12b-B 2.4.2: Testergebnis und Rating Florian Klasse 3

GS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös						X	ruhig
spielerisch				X			ernst

SS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös						X	ruhig
spielerisch					X		ernst

LS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos					X		motiviert
nervös						X	ruhig
spielerisch						X	ernst

1 2 3 4 5 6

Tab. 12c-B 2.4.2: Testergebnis und Rating Michael Klasse 3

GS	1	2	3	4	5	6	
flüchtig			X				konzentriert
lustlos						X	motiviert
nervös		X					ruhig
spielerisch		X					ernst

SS	1	2	3	4	5	6	
flüchtig				X			konzentriert
lustlos						X	motiviert
nervös			X				ruhig
spielerisch		X					ernst

LS	1	2	3	4	5	6	
flüchtig					X		konzentriert
lustlos						X	motiviert
nervös			X				ruhig
spielerisch		X					ernst

1 2 3 4 5 6

Tab. 12d-B 2.4.2: Testergebnis und Rating Orhan Klasse 3

GS	1	2	3	4	5	6	
flüchtig					X		konzentriert
lustlos					X		motiviert
nervös					X		ruhig
spielerisch				X			ernst

SS	1	2	3	4	5	6	
flüchtig		X					konzentriert
lustlos			X				motiviert
nervös					X		ruhig
spielerisch				X			ernst

LS	1	2	3	4	5	6	
flüchtig		X					konzentriert
lustlos			X				motiviert
nervös					X		ruhig
spielerisch			X				ernst

1 2 3 4 5 6

Tab. 12e-B 2.4.2: Testergebnis und Rating Thomas Klasse 3

GS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös						X	ruhig
spielerisch						X	ernst

SS	1	2	3	4	5	6	
flüchtig				X			konzentriert
lustlos					X		motiviert
nervös					X		ruhig
spielerisch					X		ernst

LS	1	2	3	4	5	6	
flüchtig					X		konzentriert
lustlos					X		motiviert
nervös				X			ruhig
spielerisch		X					ernst

1 2 3 4 5 6

Tab. 12f-B 2.4.2: Testergebnis und Rating Vanessa Klasse 3

GS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös					X		ruhig
spielerisch						X	ernst

SS	1	2	3	4	5	6	
flüchtig					X		konzentriert
lustlos					X		motiviert
nervös						X	ruhig
spielerisch						X	ernst

LS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös						X	ruhig
spielerisch						X	ernst

1 2 3 4 5 6

Tab. 12g-B 2.4.2: Testergebnis und Rating Erik Klasse 4

GS	1	2	3	4	5	6	
flüchtig				X			konzentriert
lustlos			X				motiviert
nervös				X			ruhig
spielerisch						X	ernst

SS	1	2	3	4	5	6	
flüchtig			X				konzentriert
lustlos			X				motiviert
nervös				X			ruhig
spielerisch						X	ernst

LS	1	2	3	4	5	6	
flüchtig		X					konzentriert
lustlos		X					motiviert
nervös				X			ruhig
spielerisch						X	ernst

1 2 3 4 5 6

Tab. 12h-B 2.4.2: Testergebnis und Rating Farid Klasse 4

GS	1	2	3	4	5	6	
flüchtig				X			konzentriert
lustlos					X		motiviert
nervös					X		ruhig
spielerisch					X		ernst

SS	1	2	3	4	5	6	
flüchtig					X		konzentriert
lustlos				X			motiviert
nervös					X		ruhig
spielerisch					X		ernst

LS	1	2	3	4	5	6	
flüchtig		X					konzentriert
lustlos			X				motiviert
nervös					X		ruhig
spielerisch				X			ernst

1 2 3 4 5 6

Tab. 12i-B 2.4.2: Testergebnis und Rating Katharina Klasse 4

GS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös						X	ruhig
spielerisch						X	ernst

SS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos						X	motiviert
nervös						X	ruhig
spielerisch						X	ernst

LS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos					X		motiviert
nervös						X	ruhig
spielerisch						X	ernst

1 2 3 4 5 6

Tab. 12j-B 2.4.2: Testergebnis und Rating Lars Klasse 4

GS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos					X		motiviert
nervös						X	ruhig
spielerisch					X		ernst

SS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos					X		motiviert
nervös						X	ruhig
spielerisch					X		ernst

LS	1	2	3	4	5	6	
flüchtig						X	konzentriert
lustlos					X		motiviert
nervös						X	ruhig
spielerisch					X		ernst

1 2 3 4 5 6

Tab. 12k-B 2.4.2: Testergebnis und Rating Simon Klasse 4

GS	1	2	3	4	5	6	
flüchtig					X		konzentriert
lustlos				X			motiviert
nervös					X		ruhig
spielerisch					X		ernst

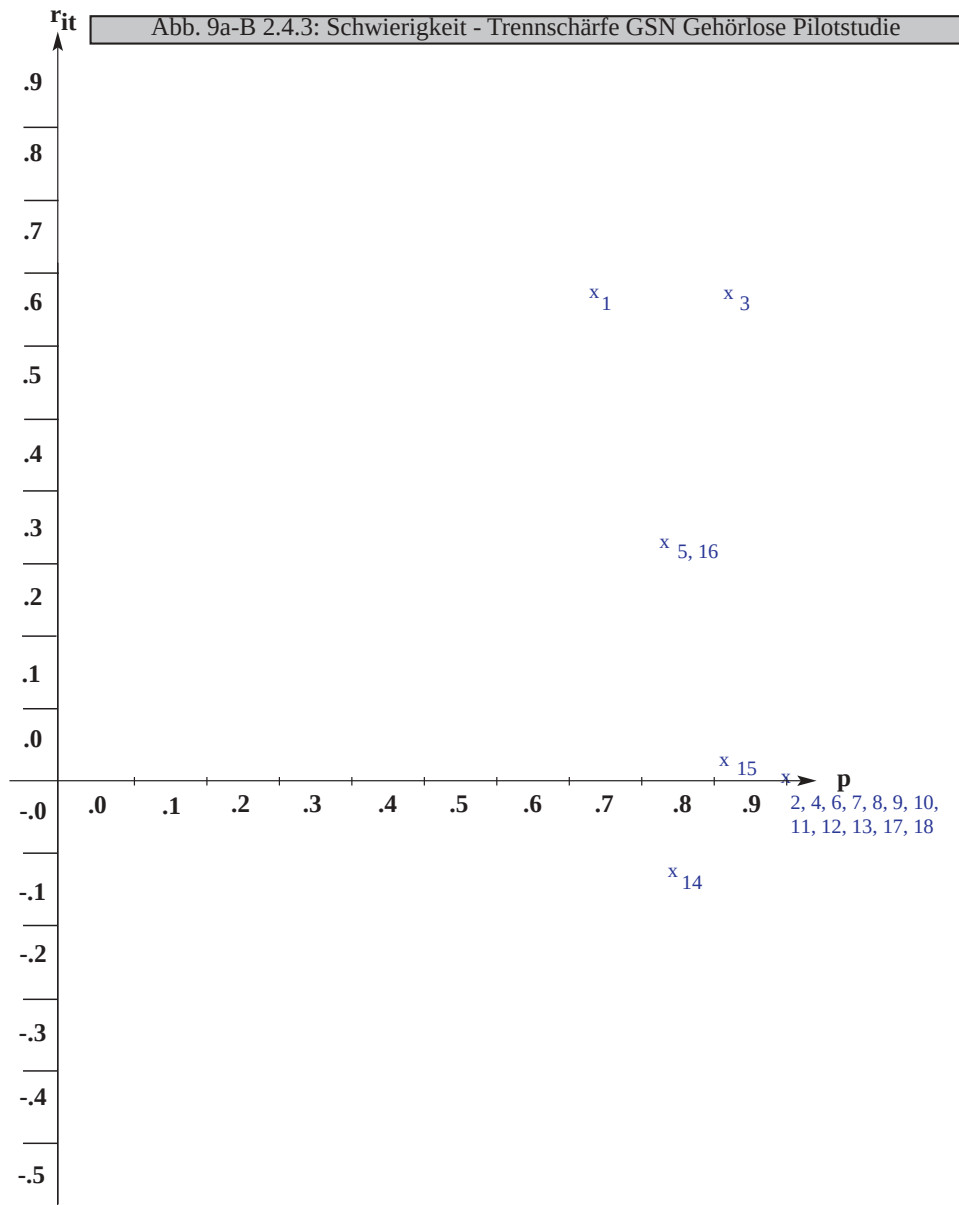
SS	1	2	3	4	5	6	
flüchtig		X					konzentriert
lustlos	X						motiviert
nervös	X						ruhig
spielerisch	X						ernst

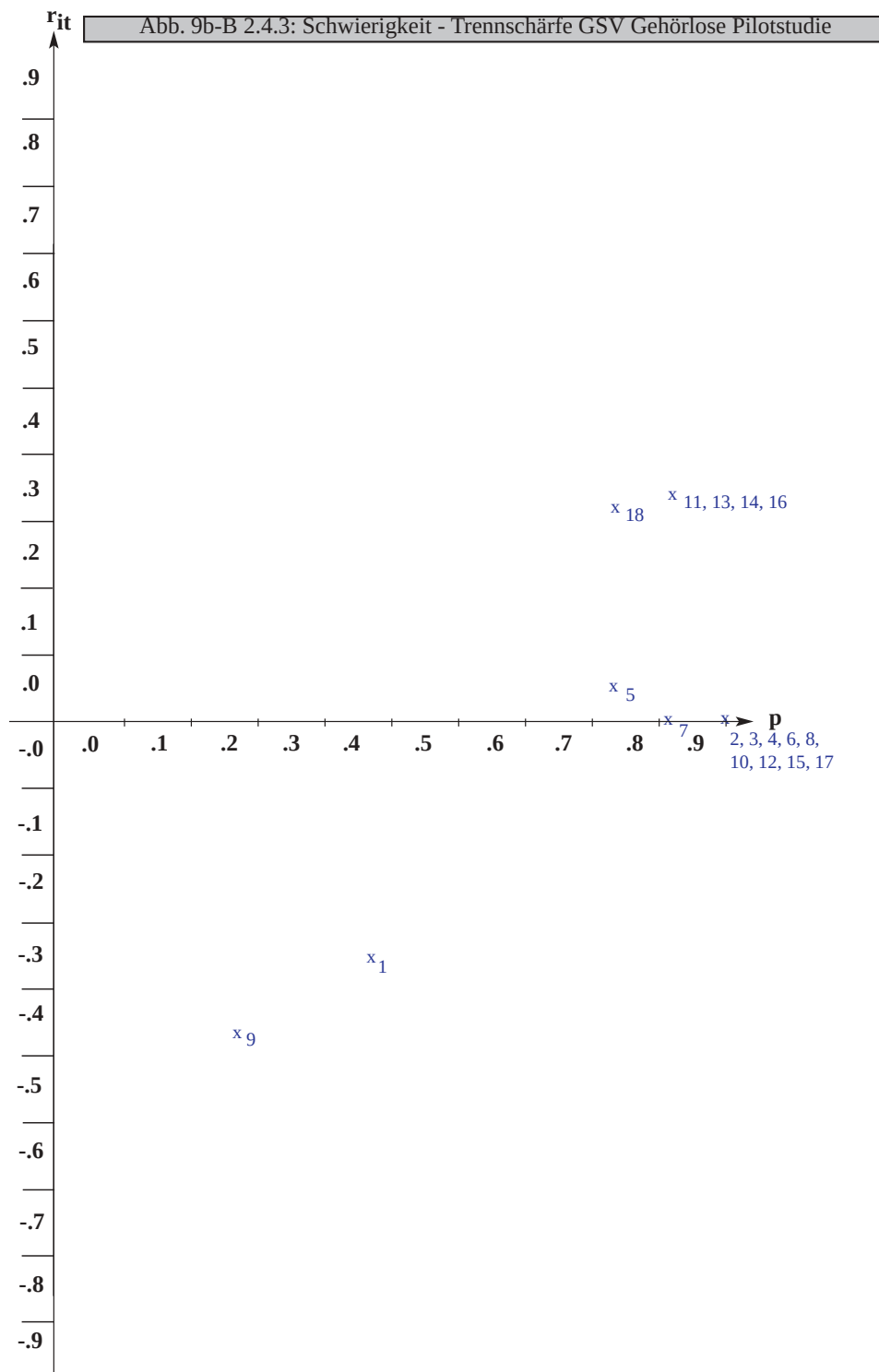
LS	1	2	3	4	5	6	
flüchtig		X					konzentriert
lustlos		X					motiviert
nervös	X						ruhig
spielerisch		X					ernst

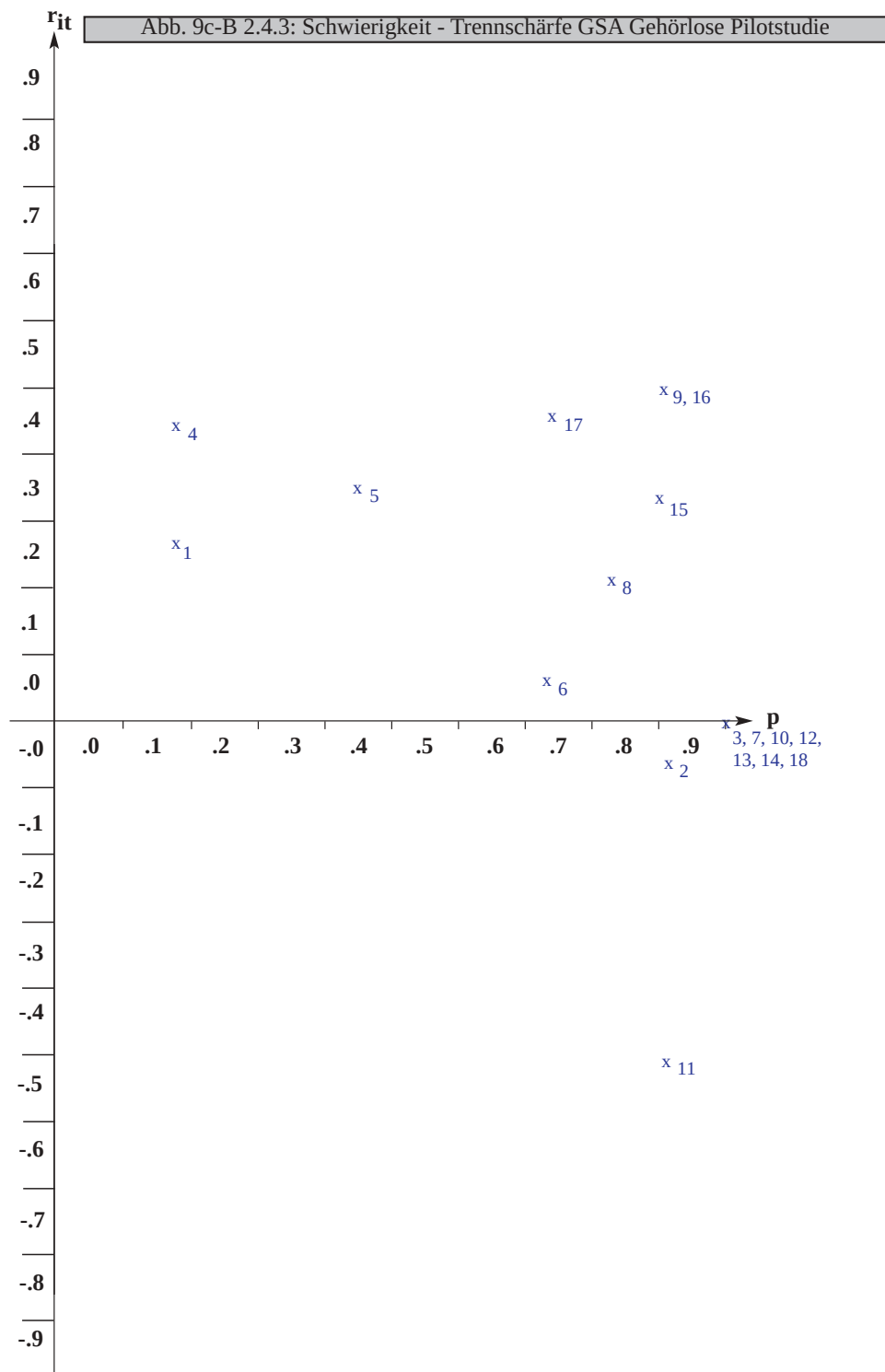
1 2 3 4 5 6

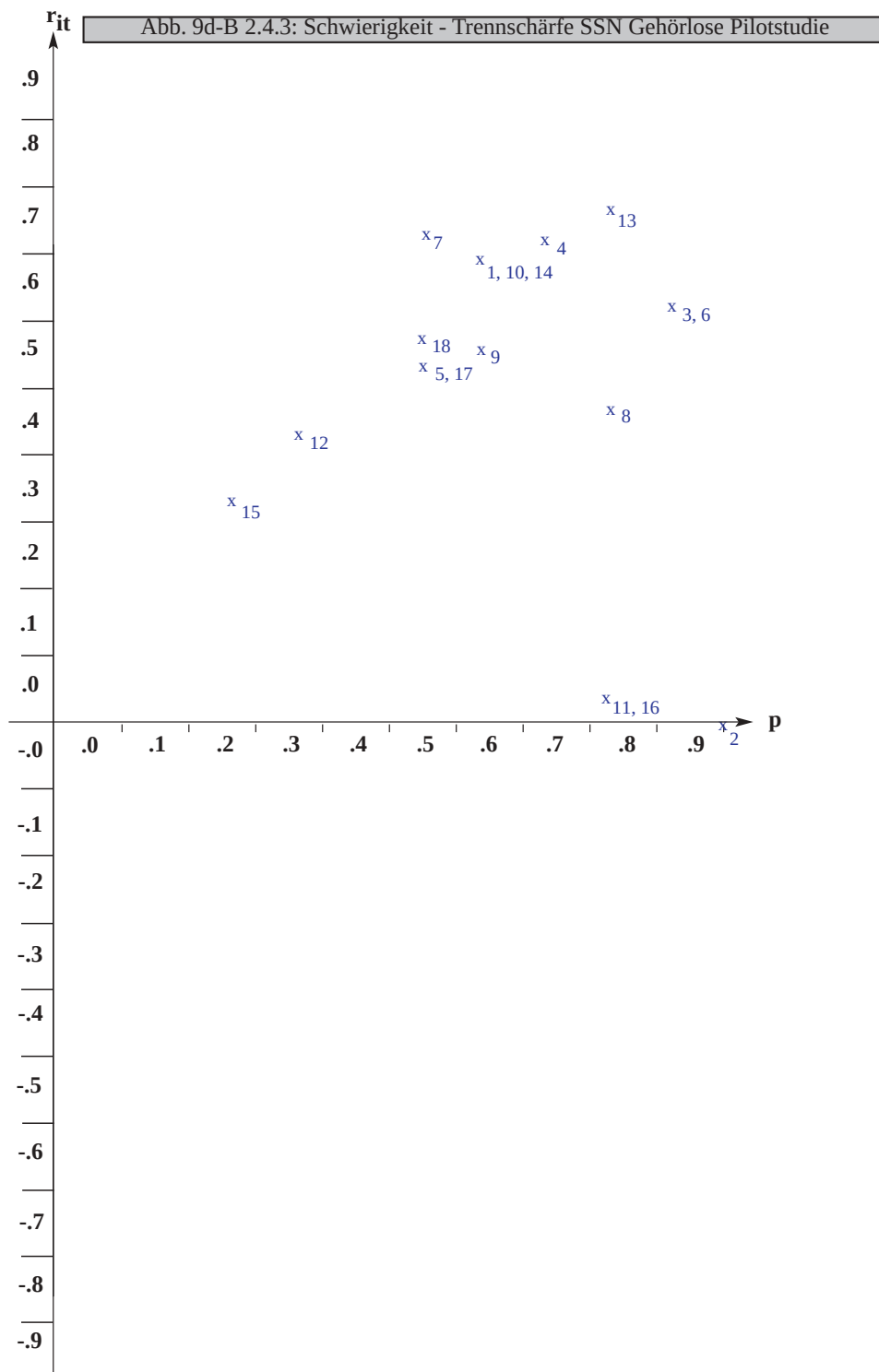
Anh. 1: B 2.4.3a:

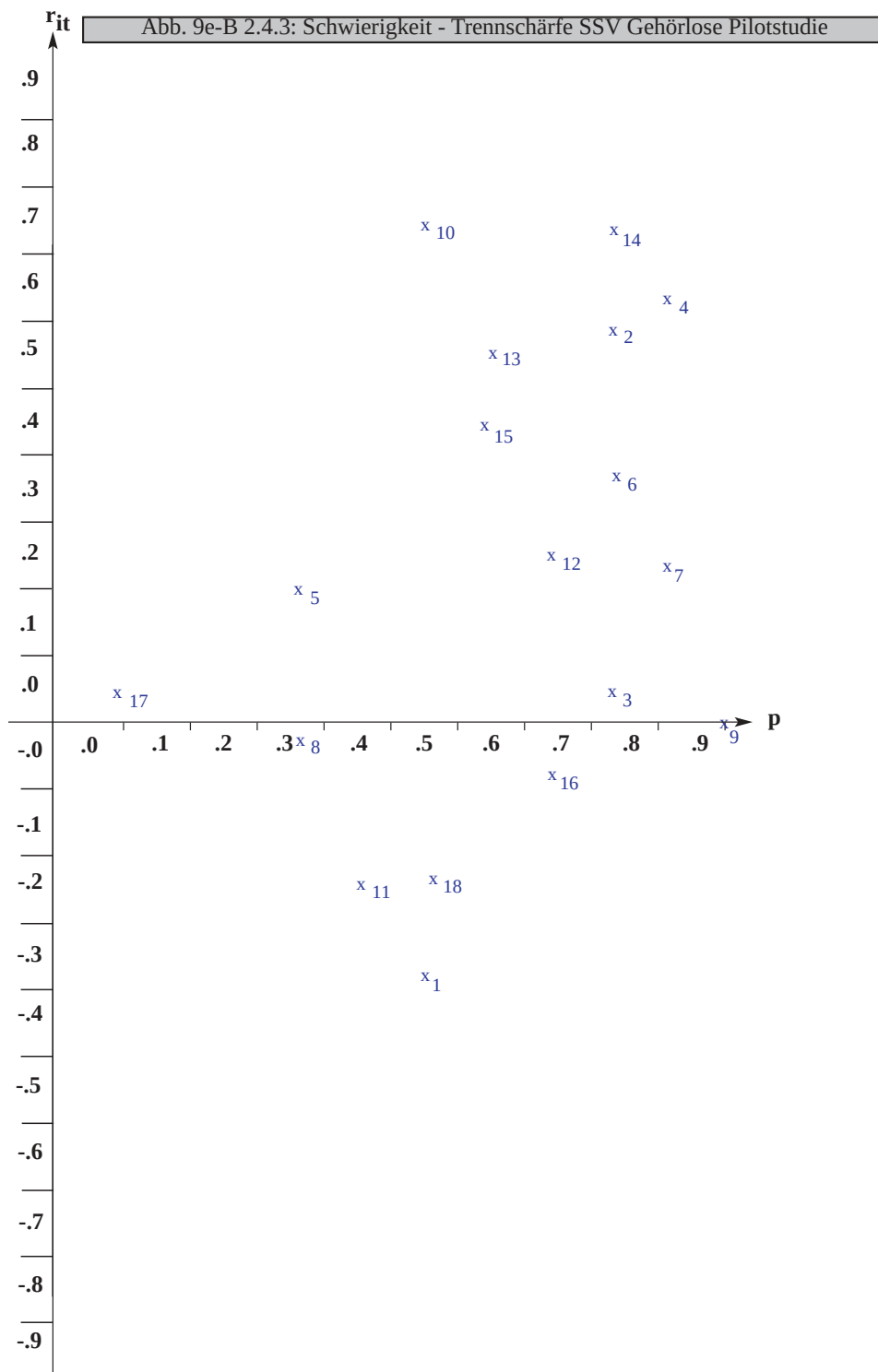
Die folgenden Koordinatensysteme verdeutlichen anschaulich die Beziehungen der Trennschärfe- und Schwierigkeitsindizes der ursprünglichen 18 Aufgaben eines jeden Subtestteils aus der Pilotversion des Tests, erhoben an den elf Schülern der bilingualen Klassen der Hamburger Gehörlosenschule. Die Schwierigkeit ist dabei auf der Abzisse, die Trennschärfe auf der Ordinate abgetragen.

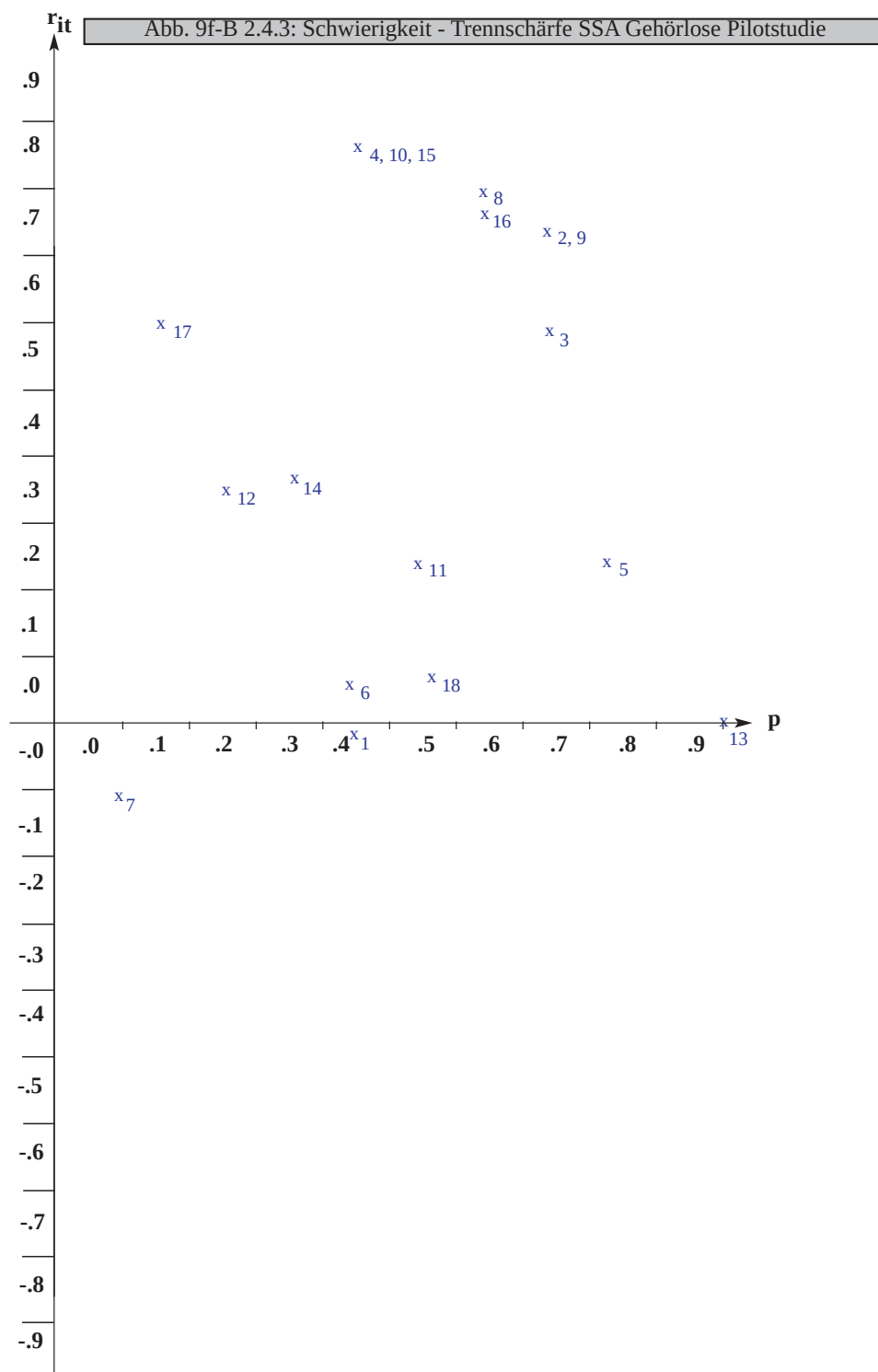


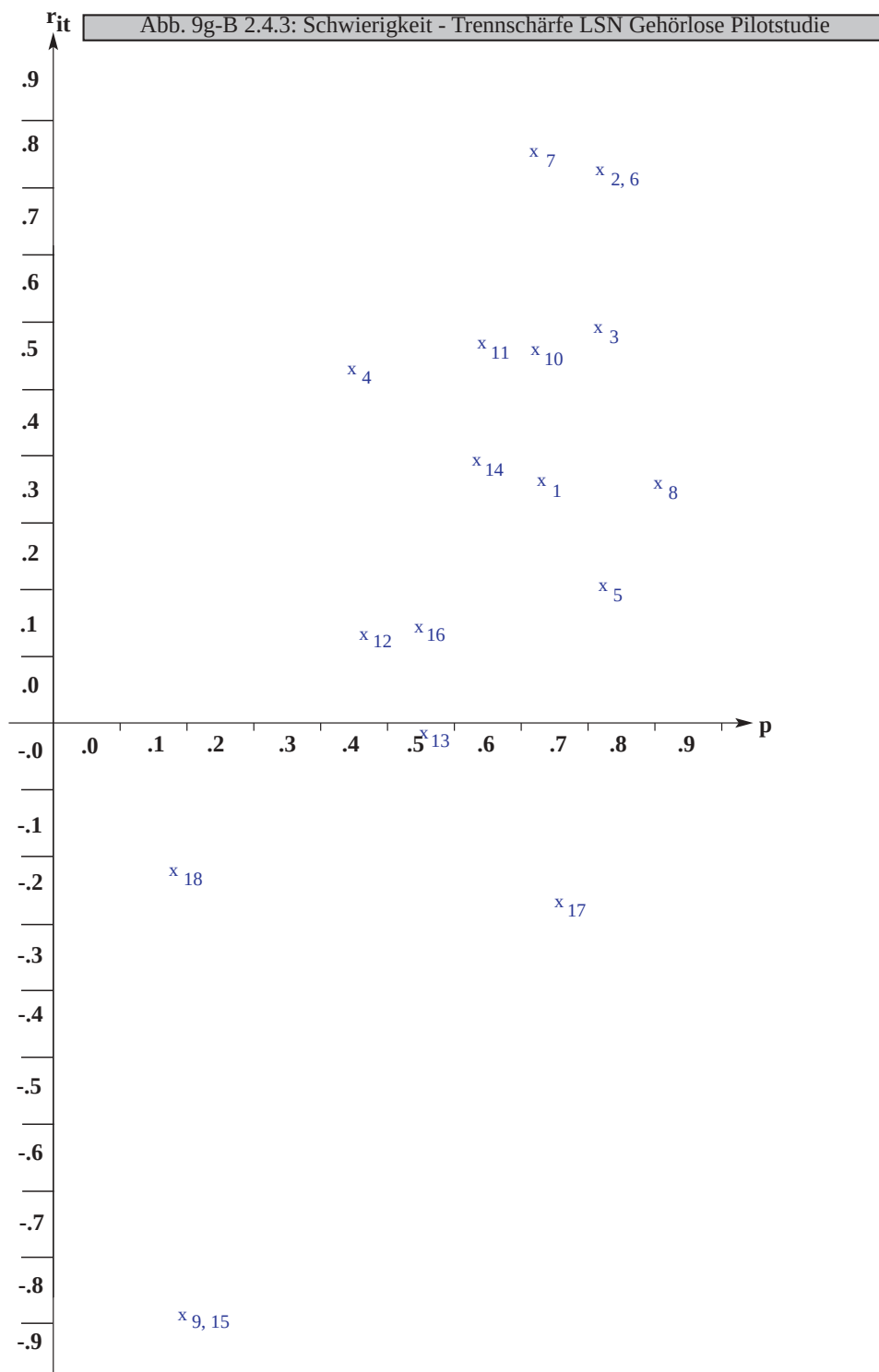


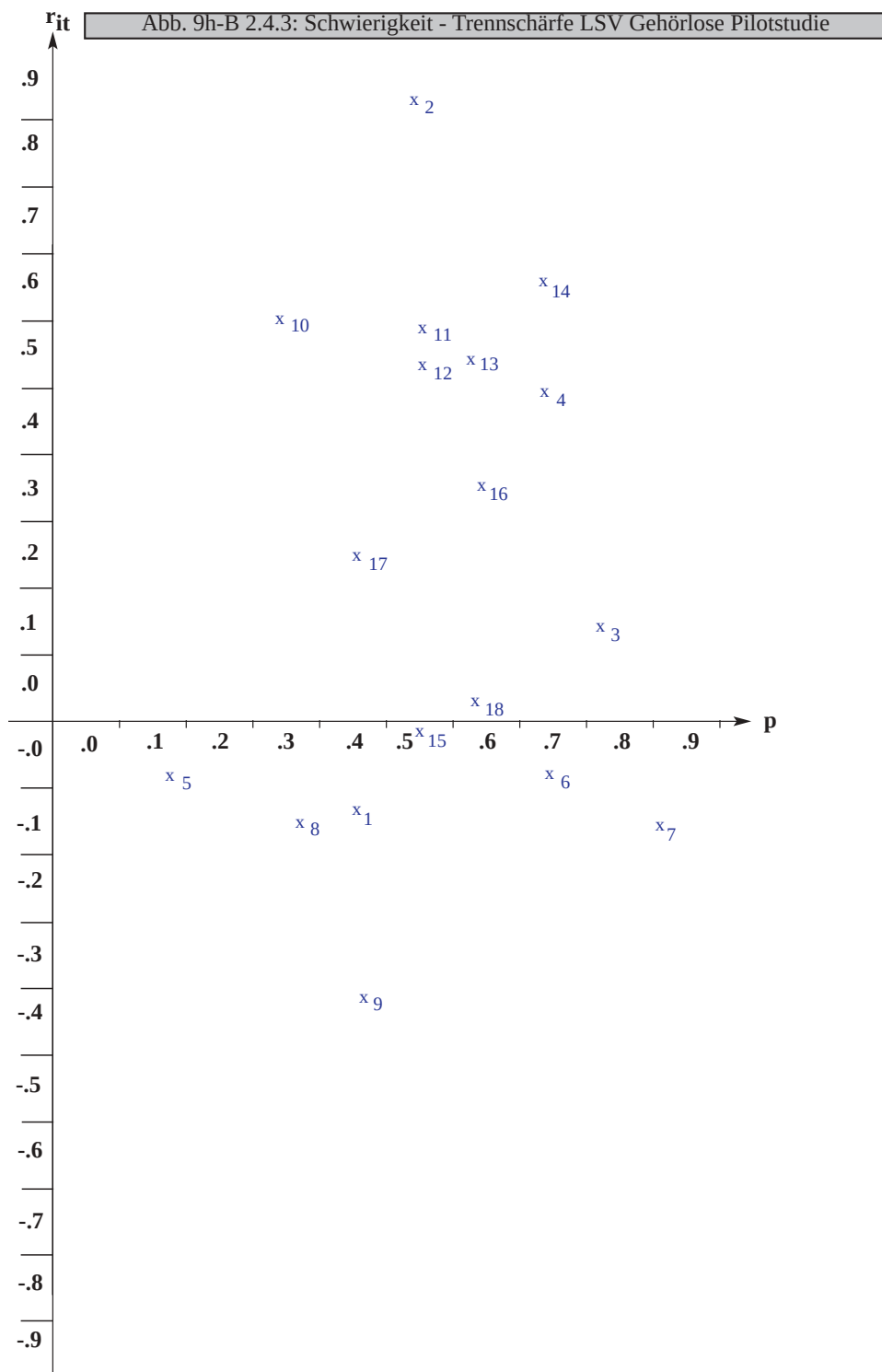


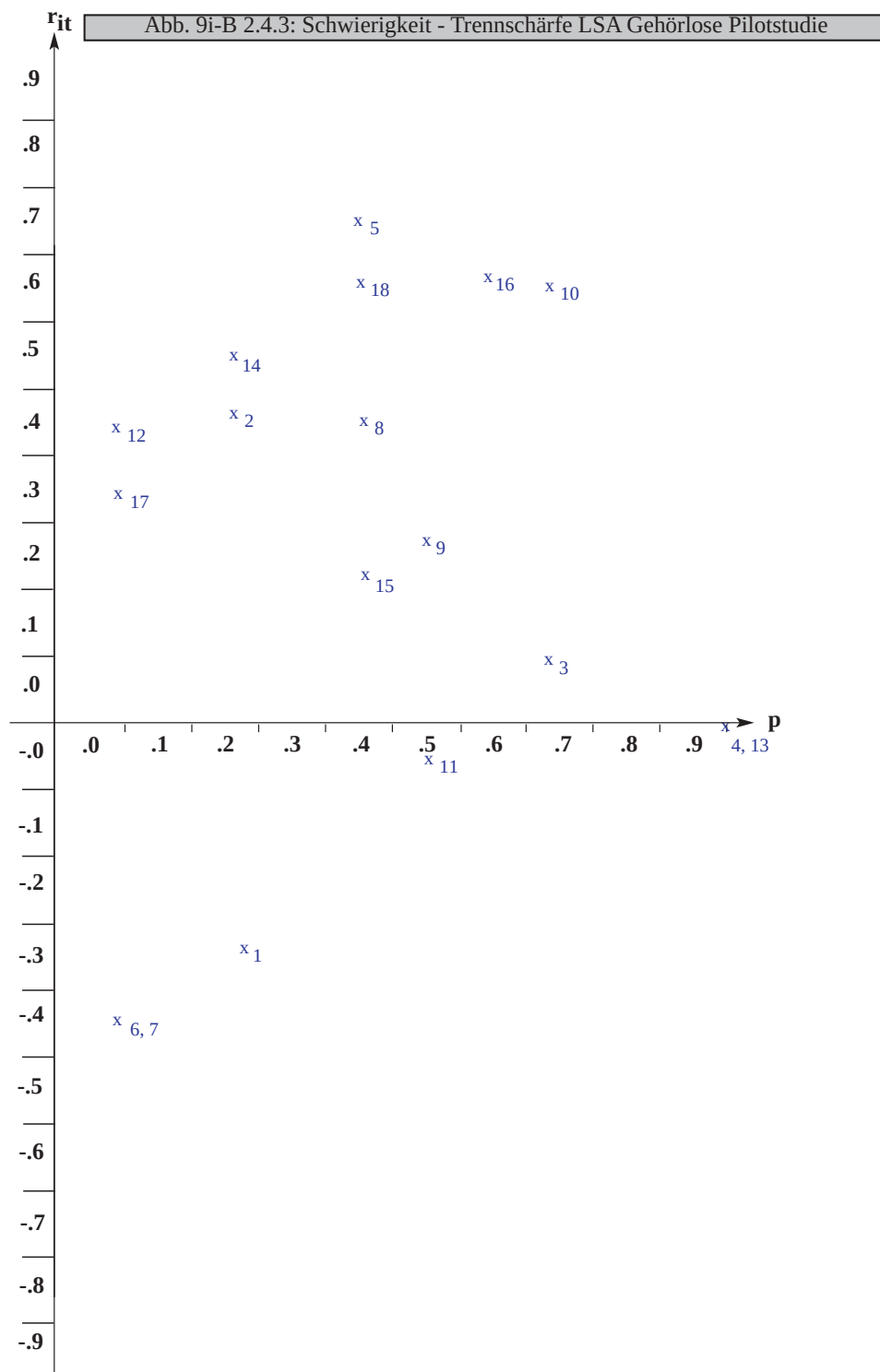






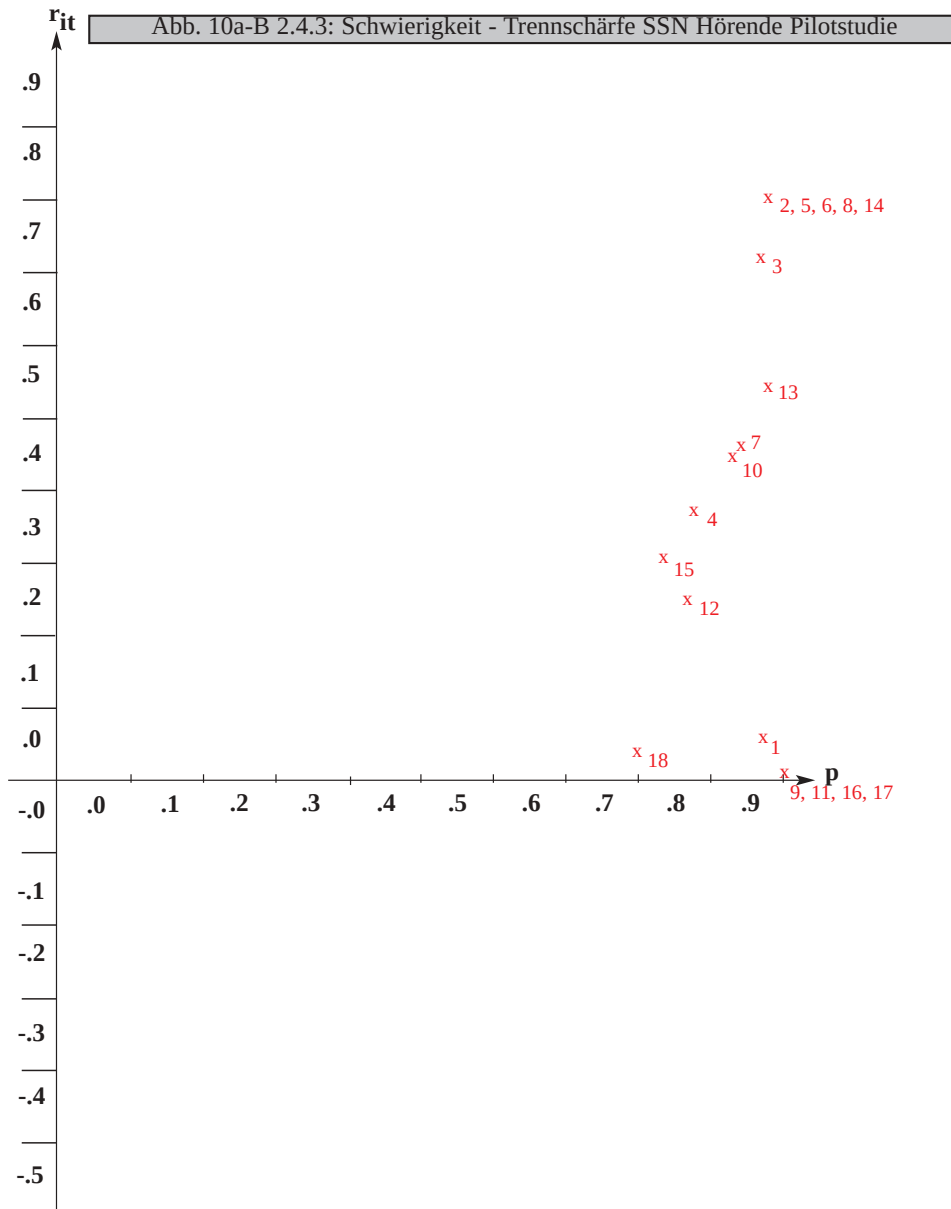


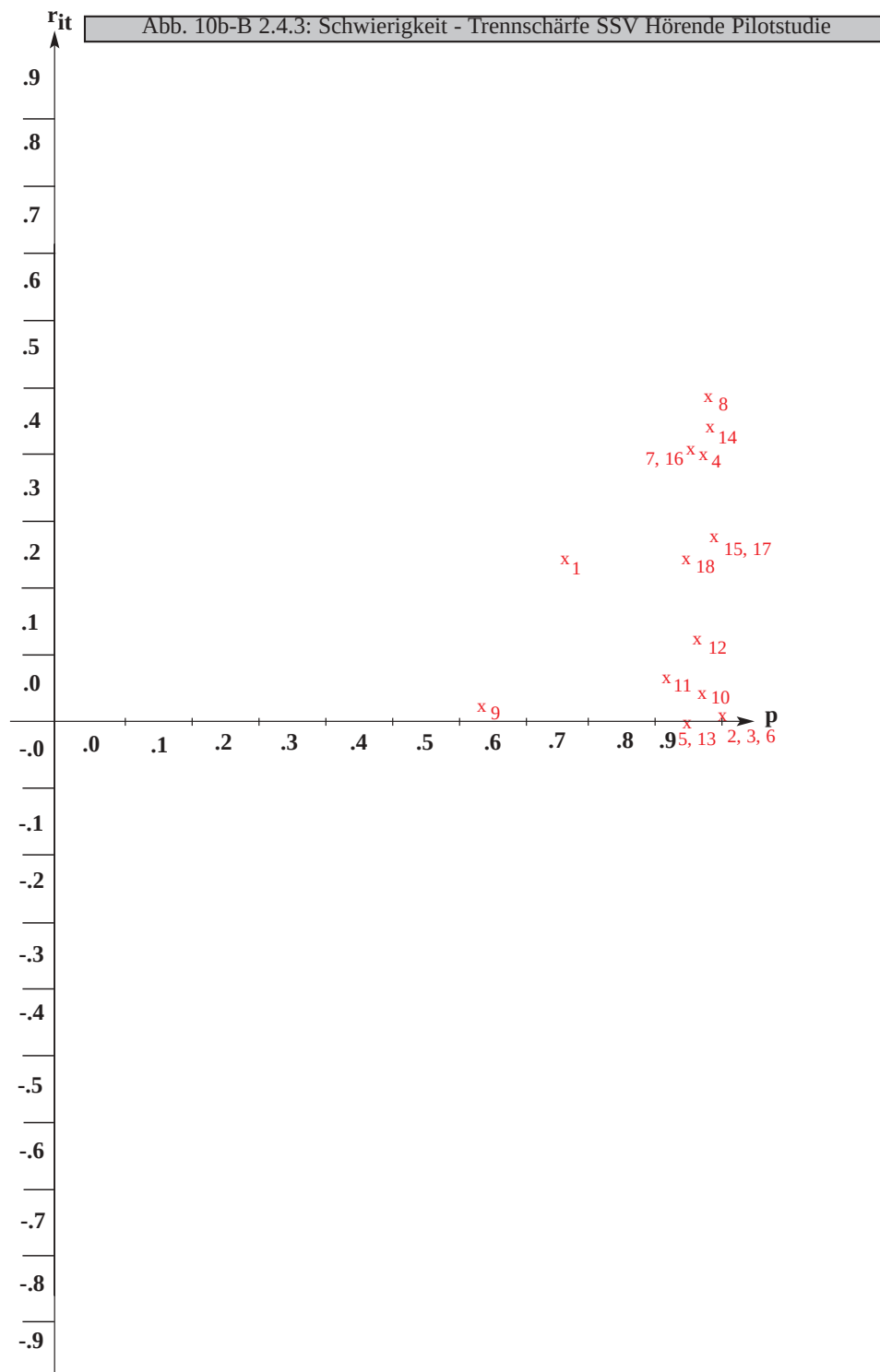


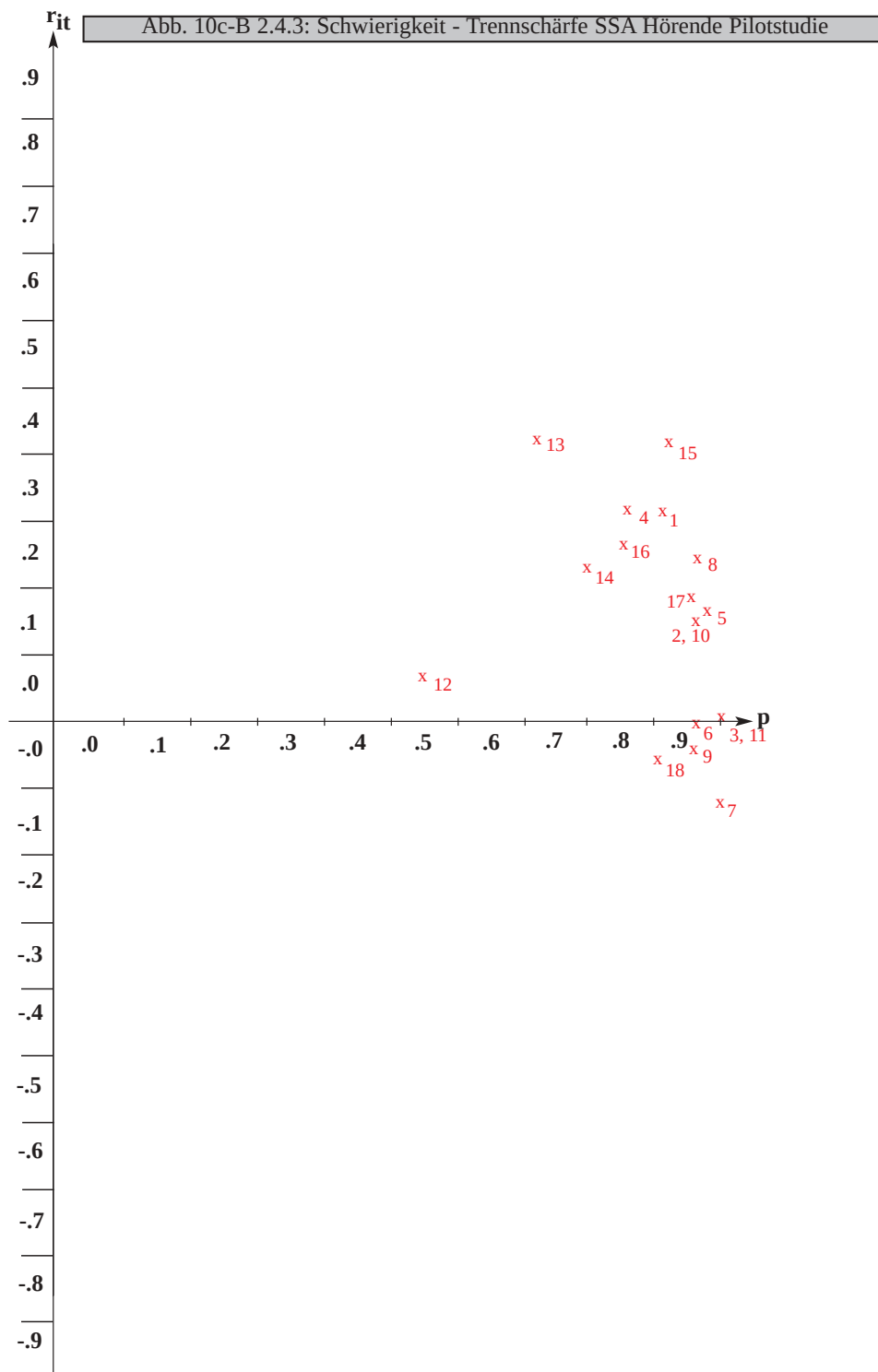


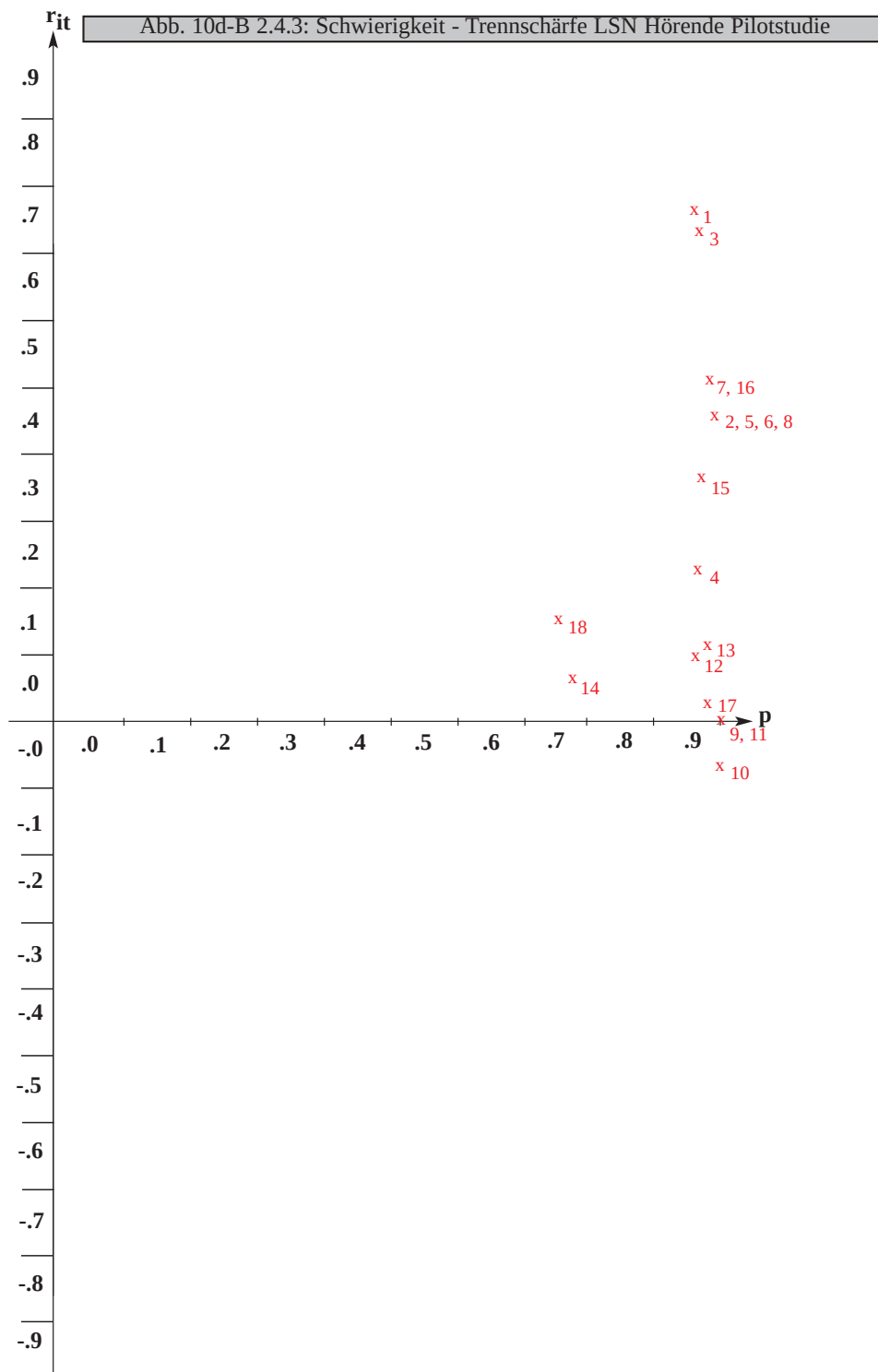
Anh. 1: B 2.4.3b:

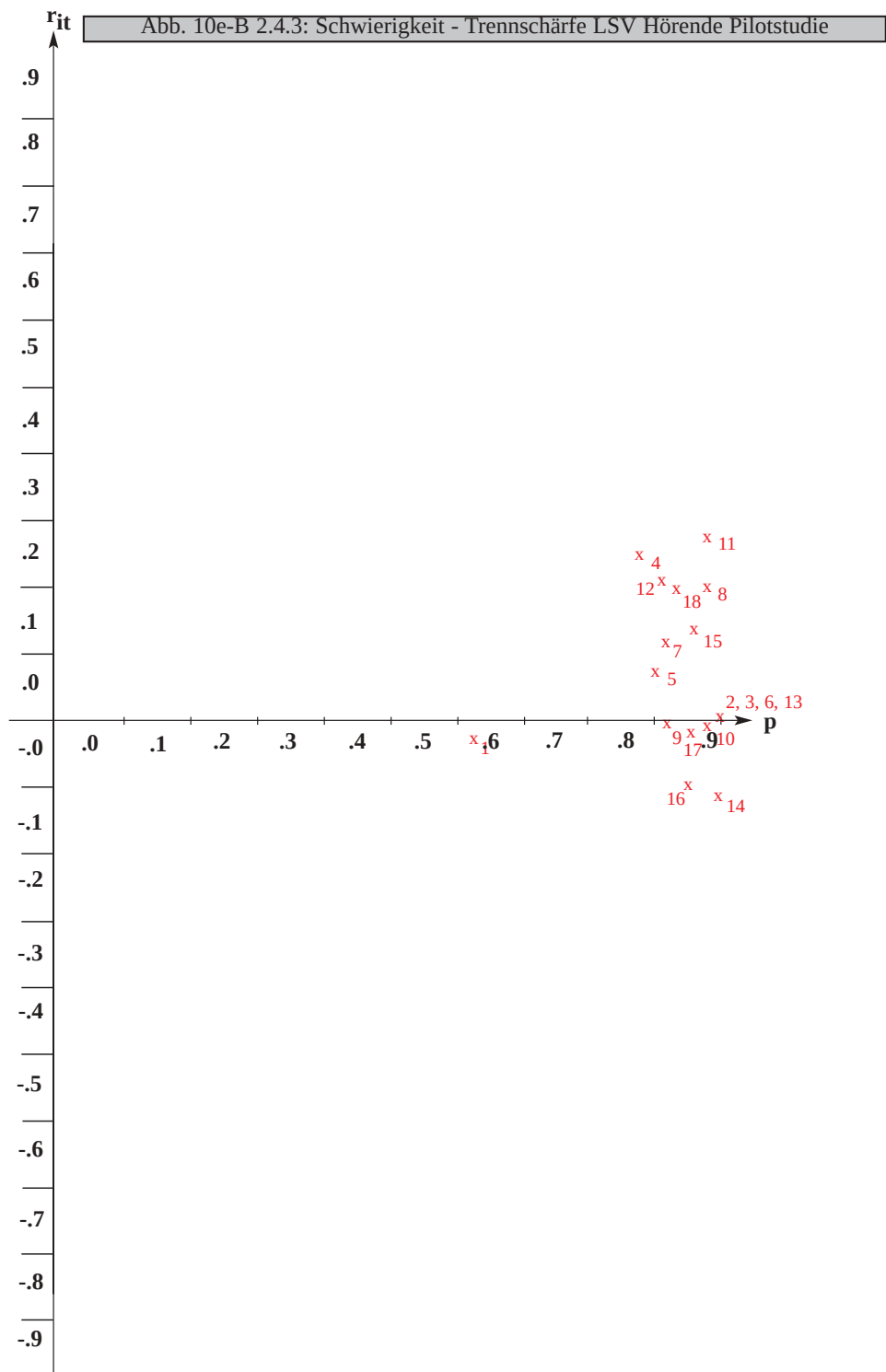
Die folgenden Koordinatensysteme verdeutlichen anschaulich die Beziehungen der Trennschärfe- und Schwierigkeitsindizes der ursprünglichen 18 Aufgaben eines jeden Subtestteils aus der Pilotversion des Tests, erhoben an einer Kontrollstichprobe von hörenden Grundschülern aus Hamburg. Die Schwierigkeit ist dabei auf der Abzisse, die Trennschärfe auf der Ordinate abgetragen.

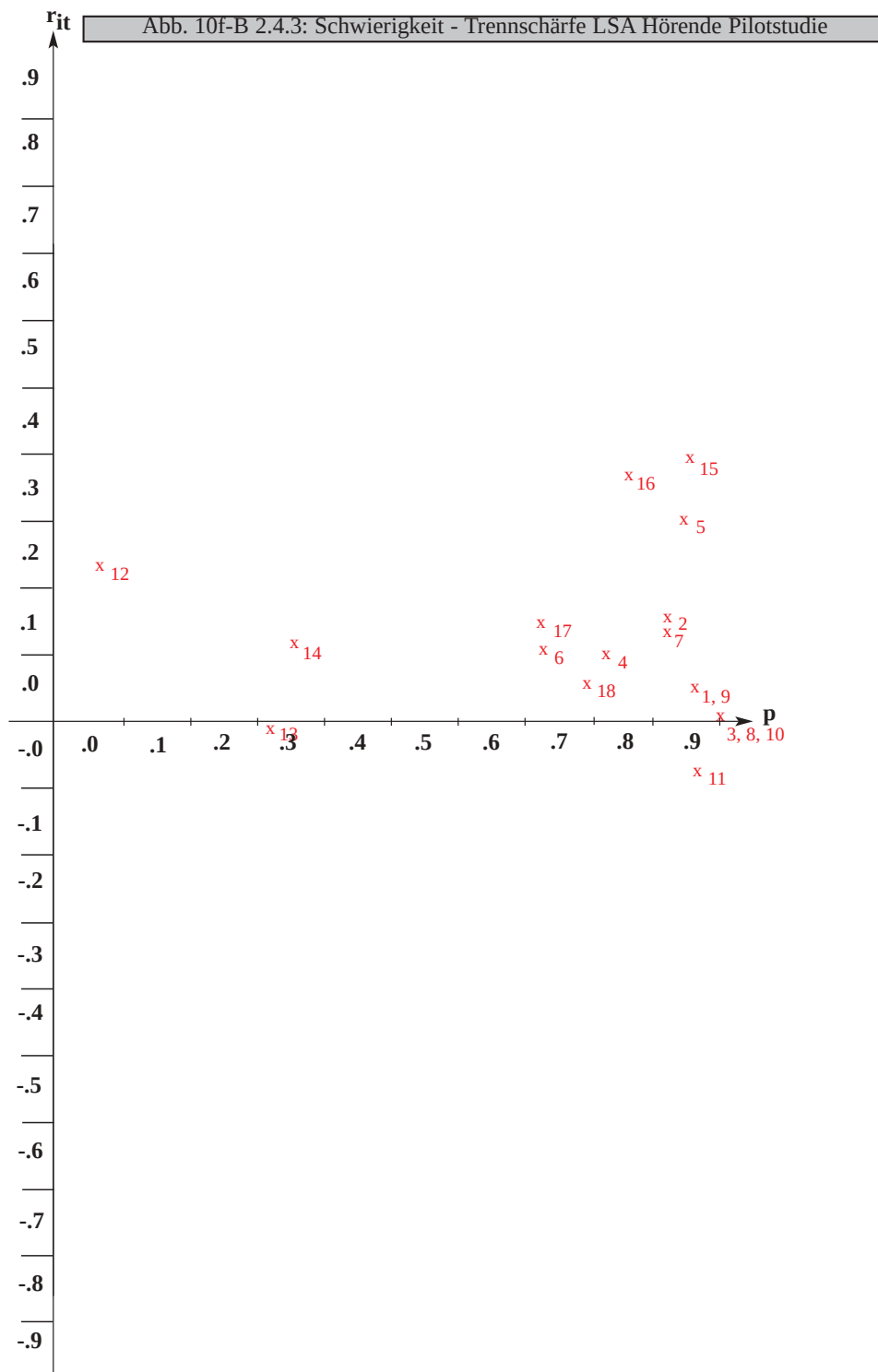












Anh. 1: B 2.4.5:

Tab. 13a-B 2.4.5: Fehleranalyse Gehörloser Nomen						
Nomen	Teil A		Teil B		Teil C	
Fehler ges.	Fehler	Distraktoren	Fehler	Distraktoren	Fehler	Distraktoren
Mutter 9,75	3	Bauer 0 Mädchen 3 Motte 0	3,75 (1 w.n.)	Bauer 0 Mädchen 1 Motte 2	3	Bauer 0 Mädchen 2 Motte 1
Auto 2	0	Bus 0 Laster 0 Zug 0	0	Bus 0 Laster 0 Zug 0	2	Bus 2 Laster 0 Zug 0
Haus 4	1	Hand 0 Kirche 1 Maus 0	1	Hand 1 Kirche 0 Maus 0	2	Hand 0 Kirche 0 Maus 2
Kopf 11	0	Junge 0 Topf 0 Zopf 0	3	Junge 1 Topf 1 Zopf 1	8	Junge 0 Topf 5 Zopf 3
Straße 9	2	Blase 0 Strauß 0 Weg 2	5	Blase 0 Strauß 1 Weg 4	2	Blase 0 Strauß 0 Weg 2
Bett 3	0	Brett 0 Decke 0 Sofa 0	1	Brett 1 Decke 0 Sofa 0	2	Brett 0 Decke 1 Sofa 1
Angst 8	0	Freude 0 Trauer 0 Wut 0	5	Freude 2 Trauer 3 Wut 0	3	Freude 0 Trauer 1 Wut 2
Stein 3	0	Kaffee 0 Stange 0 Stirn 0	2	Kaffee 1 Stange 1 Stirn 0	1	Kaffee 1 Stange 0 Stirn 0
Tier 12,75	0	Bier 0 Tor 0 Verkehr 0	4	Bier 2 Tor 0 Verkehr 2	8,75 (1 w.n.)	Bier 4 Tor 3 Verkehr 1
Bart 7	0	Bett 0 Kinn 0 Post 0	4	Bett 0 Kinn 2 Post 2	3	Bett 0 Kinn 1 Post 2
Zahn 6	0	Hahn 0 Mund 0 Zaun 0	2	Hahn 1 Mund 0 Zaun 1	4	Hahn 3 Mund 1 Zaun 0
Leute 13	0	Herde 0 Mensch 0 Paar 0	7	Herde 1 Mensch 5 Paar 1	6	Herde 1 Mensch 2 Paar 3
Zahl 7	0	Buchstabe 0 Form 0 Zeichen 0	2	Buchstabe 0 Form 1 Zeichen 1	5	Buchstabe 2 Form 0 Zeichen 3
Geld 10	2	Feld 0 Gold 1 Salz 1	4	Feld 2 Gold 1 Salz 1	4	Feld 0 Gold 3 Salz 1
Kasten 17,75	1	Kasse 1 Kissen 0 Schränk 0	7,75 (1 w.n.)	Kasse 0 Kissen 0 Schränk 7	9	Kasse 1 Kissen 6 Schränk 2
Holz 10	3	Brot 1 Gold 1 Höhle 1	2	Brot 0 Gold 1 Höhle 1	5	Brot 2 Gold 1 Höhle 2
Arbeit 8	0	Abend 0 Eimer 0 Kaffee 0	5	Abend 4 Eimer 1 Kaffee 0	3	Abend 3 Eimer 0 Kaffee 0
Schloß 13,75	0	Haus 0 Kirche 0 Schlüssel 0	5	Haus 2 Kirche 2 Schlüssel 1	8,75 (1 w.n.)	Haus 1 Kirche 0 Schlüssel 7

Tab. 13b-B 2.4.5: Fehleranalyse Gehörloser Verben						
Verben	Teil A		Teil B		Teil C	
Fehler ges.	Fehler	Distraktoren	Fehler	Distraktoren	Fehler	Distraktoren
sagen 17	6	fragen 0 schreien 6 schweigen 0	5	fragen 4 schreien 0 schweigen 1	6	fragen 1 schreien 4 schweigen 1
gehen 7	0	nagen 0 sehen 0 stehen 0	2	nagen 1 sehen 0 stehen 1	5	nagen 2 sehen 3 stehen 0
fahren 4	0	fliegen 0 frieren 0 schenken 0	2	fliegen 0 frieren 2 schenken 0	2	fliegen 1 frieren 0 schenken 1
essen 4	0	schmieren 0 trinken 0 zähneputzen 0	1	schmieren 0 trinken 0 zähneputzen 1	3	schmieren 0 trinken 1 zähneputzen 2
gucken 17,75	2	backen 0 ducken 2 küssen 0	6,75 (1 w.n.)	backen 1 ducken 3 küssen 2	9	backen 2 ducken 0 küssen 7
fliegen 5	0	fragen 0 liegen 0 winken 0	2	fragen 1 liegen 0 winken 1	3	fragen 1 liegen 2 winken 0
schlafen 3	1	liegen 1 schlagen 0 träumen 0	1	liegen 1 schlagen 0 träumen 0	1	liegen 0 schlagen 0 träumen 1
denken 13,75	0	danken 0 lenken 0 schenken 0	7	danken 2 lenken 1 schenken 4	6,75 (1 w.n.)	danken 4 lenken 2 schenken 0
hinsetzen 24,75	8	anspitzen 0 schwitzen 0 sitzen 8	11	anspitzen 0 schwitzen 1 sitzen 10	5,75 (1 w.n.)	anspitzen 2 schwitzen 1 sitzen 2
warten 12	0	braten 0 halten 0 wandern 0	5	braten 1 halten 1 wandern 3	7	braten 3 halten 0 wandern 4
verstecken 11,5	1	erschreck. 1 schlafen 0 stecken 0	5,75 (1 w.n.)	erschreck. 1 schlafen 4 stecken 0	4,75 (1 w.n.)	erschreck. 3 schlafen 0 stecken 1
weinen 8	0	regnen 0 schwitzen 0 warnen 0	3	regnen 1 schwitzen 1 warnen 1	5	regnen 1 schwitzen 3 warnen 1
schreiben 8	0	bohren 0 reiben 0 schlafen 0	4	bohren 0 reiben 1 schlafen 3	4	bohren 0 reiben 3 schlafen 1
lesen 6	1	blasen 0 fegen 0 halten 1	2	blasen 2 fegen 0 halten 0	3	blasen 2 fegen 0 halten 1
singen 9	0	geigen 0 sagen 0 schweigen 0	4	geigen 3 sagen 0 schweigen 1	5	geigen 2 sagen 2 schweigen 1
waschen 8	1	wanken 0 warten 0 wischen 1	3	wanken 0 warten 1 wischen 2	4	wanken 2 warten 1 wischen 1
brennen 16	0	brechen 0 rechnen 0 rennen 0	10	brechen 5 rechnen 3 rennen 2	6	brechen 2 rechnen 3 rennen 1
schieben 10,75	2	heben 0 schließen 1 ziehen 1	4,75 (1 w.n.)	heben 1 schließen 1 ziehen 2	4	heben 0 schließen 2 ziehen 2

Tab. 13c-B 2.4.5: Fehleranalyse Gehörloser Adjektive												
Adjek- tive	Teil A				Teil B				Teil C			
	Fehler	Distraktoren			Fehler	Distraktoren			Fehler	Distraktoren		
ganz 23	9	gerade	0		gerade	6			gerade	6		
		halb	0	6	halb	0		8	halb	0		
		rund	9		rund	0			rund	2		
klein 12	1	groß	1		groß	2			groß	0		
		klug	0	3	klug	0		8	klug	6		
		schlecht	0		schlecht	1			schlecht	2		
blau 6	0	braun	0		braun	2			braun	2		
		grau	0	3	grau	1		3	grau	0		
		schwarz	0		schwarz	0			schwarz	1		
voll 26	9	falsch	1		falsch	2			falsch	1		
		stolz	1	6	stolz	2		11	stolz	3		
		viel	7		viel	2			viel	7		
blond 15	7	braun	1		braun	2			braun	0		
		grau	3	2	grau	0		6	grau	3		
		rot	3		rot	0			rot	3		
komisch 18,75	3	gemischt	0		gemischt	4			gemischt	6		
		kurz	2	5,75 (1 w.n.)	kurz	1		10	kurz	1		
		neu	1		neu	0			neu	3		
krumm 19,5	0	kaputt	0		kaputt	6			kaputt	1		
		krank	0	9,75 (1 w.n.)	krank	1		9,75 (1 w.n.)	krank	3		
		stumm	0		stumm	2			stumm	5		
hellblau 12	2	dunkelblau	0		dunkelblau	1			dunkelblau	4		
		hellgrau	1	4	hellgrau	3		6	hellgrau	0		
		hellgrün	1		hellgrün	0			hellgrün	2		
orange 9	1	lila	0		lila	0			lila	0		
		rosa	1	3	rosa	1		5	rosa	2		
		rot	0		rot	2			rot	3		
traurig 9	0	ernst	0		ernst	3			ernst	1		
		erstaunt	0	6	erstaunt	1		3	erstaunt	2		
		fröhlich	0		fröhlich	2			fröhlich	0		
verletzt 11	1	besetzt	0		besetzt	1			besetzt	0		
		gesund	0	5	gesund	2		5	gesund	2		
		sauber	1		sauber	2			sauber	3		
warm 29	11	heiß	11		heiß	4			heiß	7		
		kalt	0	8	kalt	1		10	kalt	1		
		weich	0		weich	3			weich	2		
glatt 32,75	11	platt	10		platt	4			platt	4		
		satt	0	11	satt	1		10,75 (1 w.n.)	satt	5		
		sauber	1		sauber	6			sauber	1		
links 25,75	11	oben	2		oben	2			oben	1		
		rechts	5	6,75 (1 w.n.)	rechts	2		8	rechts	3		
		unten	4		unten	2			unten	4		
frei 12,75	1	erschrocken	0		erschrocken	2			erschrocken	2		
		frech	1	6	frech	3		5,75 (1 w.n.)	frech	3		
		klein	0		klein	1			klein	0		
schmutzig 9	1	sauber	0		sauber	3			sauber	2		
		zerknüllt	0	4	zerknüllt	1		4	zerknüllt	1		
		zerrissen	1		zerrissen	0			zerrissen	1		
weich 21,75	3	wach	2		wach	1			wach	6		
		warm	1	8,75 (1 w.n.)	warm	4		10	warm	3		
		weiß	0		weiß	3			weiß	1		
leise 11	0	heiß	0		heiß	1			heiß	1		
		leicht	0	5	leicht	2		6	leicht	3		
		links	0		links	2			links	2		

Anh. 1: B 2.4.7:

Tab. 14a-B 2.4.7: Itemrangliste Nomen				
Fehler	Rang neu	Item	Rang alt	Rangdifferenz
2	1	Auto	2	+1
3	2	Stein	8	+6
3	3	Bett	6	+3
4	4	Haus	3	-1
6	5	Zahn	11	+6
7	6	Zahl	13	+7
7	7	Bart	10	+3
8	8	Arbeit	17	+9
8	9	Angst	7	-2
9	10	Straße	5	-5
9,75	11	Mutter	1	-10
10	12	Holz	16	+4
10	13	Geld	14	+1
11	14	Kopf	4	-10
12,75	15	Tier	9	-6
13	16	Leute	12	-4
13,75	17	Schloß	18	+1
17,75	18	Kasten	15	-3

Tab. 14b-B 2.4.7: Itemrangliste Verben				
Fehler	Rang neu	Item	Rang alt	Rangdifferenz
3	1	schlafen	7	+6
4	2	essen	4	+2
4	3	fahren	3	0
5	4	fliegen	6	+2
6	5	lesen	14	+9
7	6	gehen	2	-4
8	7	waschen	16	+9

Fehler	Rang neu	Item	Rang alt	Rangdifferenz
8	8	schreiben	13	+5
8	9	weinen	12	+3
9	10	singen	15	+5
10,75	11	schieben	18	+7
11,5	12	verstecken	11	-1
12	13	warten	10	-3
13,75	14	denken	8	-6
16	15	brennen	17	+2
17	16	sagen	1	-15
17,75	17	gucken	5	-12
24,75	18	hinsetzen	9	-9

Tab. 14c-B 2.4.7: Itemrangliste Adjektive				
Fehler	Rang neu	Item	Rang alt	Rangdifferenz
6	1	blau	3	+2
9	2	schmutzig	16	+14
9	3	traurig	10	+7
9	4	orange	9	+5
11	5	leise	18	+13
11	6	verletzt	11	+5
12	7	hellblau	8	+1
12	8	klein	2	-6
12,75	9	frei	15	+6
15	10	blond	5	-5
18,75	11	komisch	6	-5
19,5	12	krumm	7	-5
21,75	13	weich	17	+4
23	14	ganz	1	-13
25,75	15	links	14	-1
26	16	voll	4	-12
29	17	warm	12	-5
32,75	18	glatt	13	-5

Anh. 1: B 3.2.1:

Folgende Tabelle führt die Ergebnisse der Normierungsstichprobe der Version I des Wortschatztests (je Subtestteil maximal 24 Punkte) auf. Dabei werden entfallende Einzelleistungen der Testpersonen, die an einzelnen Subtestteilen nicht teilnehmen konnten, durch ein Leerzeichen (-) gekennzeichnet.

Tab. 21-B 3.2.1 Testpunktwerte									
Test-person	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
1	18	17	12	7	7	6	12	10	6
2	18	19	15	11	5	4	11	15	14
3	20	20	15	13	5	4	11	8	8
4	20	17	17	13	8	10	17	12	14
5	15	21	12	11	3	5	6	9	10
6	17	21	11	7	5	9	9	5	7
7	22	21	18	12	6	9	14	12	13
8	21	17	20	24	21	22	20	21	20
9	22	24	20	22	22	22	21	21	23
10	22	19	17	20	17	12	20	19	10
11	23	22	23	22	17	22	20	18	20
12	22	22	19	20	17	19	21	18	14
13	22	24	22	24	21	22	18	22	20
14	11	9	12	10	5	6	7	7	4
15	21	19	19	22	17	22	20	20	18
16	12	14	11	11	5	7	-	-	-
17	21	19	21	15	15	21	20	14	17
18	16	18	14	9	5	8	12	13	11
19	15	10	12	11	-	-	16	16	14
20	18	15	12	11	9	11	15	14	13
21	19	22	17	15	11	13	17	15	12
22	16	22	12	16	9	7	10	5	9
23	17	21	13	19	13	4	15	16	12
24	20	18	14	17	14	15	21	14	17
25	24	21	20	22	22	23	24	23	23
26	20	19	17	15	11	9	21	15	9
27	15	18	18	11	5	6	5	10	8
28	18	20	14	12	9	8	-	-	-
29	15	20	10	18	14	13	-	-	-
30	15	19	12	9	11	7	9	6	7
31	18	16	13	8	7	9	5	8	5
32	17	22	15	12	14	14	13	14	9
33	12	16	11	8	2	7	6	4	9
34	15	14	15	15	5	8	-	-	-
35	23	20	15	21	15	14	21	21	17
36	19	20	18	14	6	11	19	18	20
37	20	16	13	11	-	-	21	14	11

Test-person	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
38	16	16	14	15	13	11	17	8	11
39	18	18	14	17	15	19	14	19	15
40	19	19	15	18	19	20	18	16	15
41	19	21	19	23	17	22	17	18	21
42	24	23	21	24	22	23	22	20	22
43	23	24	19	23	19	21	22	21	22
44	18	19	14	12	6	9	14	14	11
45	14	18	17	15	9	10	13	11	9
46	17	21	15	14	12	6	15	14	12
47	17	15	13	17	13	10	14	15	13
48	16	16	14	9	7	13	15	11	9
49	16	20	11	9	4	8	10	11	11
50	16	20	14	13	16	12	18	16	13
51	16	15	12	17	8	7	13	8	8
52	19	22	12	20	16	16	17	15	9
53	19	20	18	16	14	13	17	16	17
54	14	21	13	18	8	13	7	11	12
55	13	18	13	-	-	-	11	12	5
56	17	20	15	10	7	6	13	9	5
57	16	21	12	12	12	14	11	15	10
58	22	18	17	20	17	15	17	14	13
59	16	23	17	21	19	15	18	17	13
60	20	21	14	18	18	15	17	22	11
61	20	18	17	18	13	10	17	18	13
62	14	21	17	10	5	9	8	15	8
63	21	20	11	15	9	11	12	7	4
64	21	21	19	19	16	20	16	12	19
65	16	21	16	14	12	12	19	15	15
66	15	19	14	-	-	-	14	11	8
67	22	20	18	18	16	17	18	12	14
68	20	23	19	23	20	22	21	21	18
69	16	21	16	19	19	17	19	19	16
70	19	19	19	23	19	23	17	17	20
71	19	20	14	18	7	8	12	15	11
72	22	21	19	23	20	23	17	18	19
73	23	22	20	22	20	22	24	22	22
74	16	20	16	16	9	11	18	14	11
75	20	22	13	13	8	9	14	12	11
76	16	22	12	11	6	9	11	13	9
77	18	20	13	14	9	7	22	19	18
78	18	22	13	9	11	8	10	8	8
79	20	19	11	11	10	11	18	12	17
80	23	23	21	24	20	23	23	23	21
81	18	21	17	17	15	18	16	11	13
82	23	23	19	22	19	21	22	21	18
83	19	22	14	16	18	16	18	17	11

Test-person	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
84	21	22	19	21	15	17	18	16	16
85	21	19	19	20	15	15	18	15	14
86	20	20	20	19	17	17	21	18	17
87	17	18	10	8	8	9	-	-	-
88	16	19	16	7	8	3	15	13	11
89	21	23	19	21	17	17	20	19	20
90	20	21	20	24	16	18	17	16	16
91	16	21	18	14	8	11	14	8	8
92	11	12	14	7	6	9	10	13	6
93	18	17	14	13	7	9	8	8	10
94	19	19	14	15	15	15	19	18	12
95	23	22	18	24	19	21	21	15	20
96	17	19	14	12	4	8	19	19	18
97	16	15	11	8	7	3	12	7	5
98	22	23	23	23	19	22	20	19	22
99	14	23	15	11	11	9	13	9	11
100	17	20	19	20	16	16	20	17	20
101	-	-	-	16	13	12	-	-	-
102	19	18	18	22	17	12	20	21	15
103	18	20	16	18	16	13	21	12	18
104	-	-	-	23	22	22	24	22	24
105	22	22	22	23	24	23	24	22	23
106	19	22	21	23	20	23	23	21	19
107	21	19	20	24	20	22	23	17	22
108	18	18	20	22	15	23	23	20	21
109	20	23	19	22	16	20	22	20	22
110	18	22	19	17	12	14	17	19	21
111	19	22	17	13	12	6	19	17	18
112	18	22	15	15	15	15	19	18	17

Anh. 1: B 3.2.2:

Die folgenden Tabellen beschreiben die Summe der Antworten aller Testpersonen zu jeder Aufgabe. Zu jeder Aufgabennummer gehören jeweils drei Zeilen für die Sprachmodalitäten Gebärdensprache (GS), Schriftsprache (SS) und Lautsprache (LS). Spaltenweise hintereinander werden die Wahlhäufigkeiten des Items selbst, der drei Distraktoren und schließlich der fehlenden Antworten aufgeführt. In der ersten Tabelle befindet sich die Darstellung der Nomen, in der zweiten die der Verben und in der dritten die der Adjektive.

Tab. 22a-B 3.2.2: Distraktorenverteilung Nomen					
Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
1	Mutter	Mädchen	Bauer	Motte	fehlt
GS	69	14	12	15	2
SS	52	45	8	5	2
LS	55	16	12	23	6
2	Auto	Bus	Laster	Zug	fehlt
GS	108	1	0	1	2
SS	110	0	0	0	2
LS	102	1	1	2	6
3	Haus	Kirche	Hand	Maus	fehlt
GS	110	0	0	0	2
SS	102	2	2	4	2
LS	85	2	8	11	6
4	Kopf	Zopf	Junge	Topf	fehlt
GS	83	21	6	0	2
SS	79	16	12	3	2
LS	51	22	7	26	6
5	Straße	Weg	Strand	Strauß	fehlt
GS	55	54	1	0	2
SS	70	10	16	14	2
LS	65	33	5	3	6
6	Bett	Decke	Brett	Sofa	fehlt
GS	98	6	5	1	2
SS	94	0	8	8	2
LS	93	7	2	4	6
7	Krone	Hut	Bohne	Kranz	fehlt
GS	100	9	0	1	2
SS	64	25	9	12	2
LS	73	22	3	8	6

Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
8	Angst	Ärger	Trauer	Freude	fehlt
GS	92	10	7	1	2
SS	67	18	14	11	2
LS	81	8	7	10	6
9	Garten	Baum	Geschenk	Arbeit	fehlt
GS	52	1	29	28	2
SS	51	49	6	4	2
LS	62	3	18	23	6
10	Stein	Felsen	Kaffee	Stange	fehlt
GS	72	24	2	12	2
SS	66	17	10	17	2
LS	78	12	6	10	6
11	Schuh	Stiefel	Strumpf	Uhr	fehlt
GS	67	21	21	1	2
SS	105	0	2	3	2
LS	76	10	6	14	6
12	Bart	Haare	Kinn	Post	fehlt
GS	106	0	4	0	2
SS	55	33	9	13	2
LS	76	3	7	20	6
13	Zahn	Mund	Hahn	Zange	fehlt
GS	91	19	0	0	2
SS	83	5	3	19	2
LS	87	12	3	4	6
14	Leute	Paar	Mensch	Herde	fehlt
GS	89	6	2	13	2
SS	44	13	46	7	2
LS	59	15	14	18	6
15	Ei	Huhn	Kartoffel	Eis	fehlt
GS	109	0	1	0	2
SS	105	0	0	5	2
LS	94	1	1	10	6
16	Zahl	Regen	Buchstabe	Pfahl	fehlt
GS	104	2	2	2	2
SS	78	11	8	13	2
LS	44	22	13	27	6
17	Geld	Gold	Salz	Feld	fehlt
GS	61	47	2	0	2
SS	83	11	7	9	2
LS	72	25	6	3	6

Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
18	Fisch	Schiff	Flasche	Frosch	fehlt
GS	110	0	0	0	2
SS	102	0	3	5	2
LS	93	2	0	11	6
19	Holz	Baum	Brot	Gold	fehlt
GS	62	1	41	6	2
SS	73	28	4	5	2
LS	64	2	16	24	6
20	Arbeit	Eimer	Stein	Ampel	fehlt
GS	76	24	9	1	2
SS	73	7	14	16	2
LS	66	12	4	24	6
21	Schloss	Schlüssel	Kirche	Schaf	fehlt
GS	94	2	10	4	2
SS	40	30	33	7	2
LS	46	45	4	11	6
22	Kaffee	Kanne	Kanister	Wanne	fehlt
GS	59	5	46	0	2
SS	48	48	8	6	2
LS	55	24	13	14	6
23	Rose	Sonnenblume	Tulpe	Rasen	fehlt
GS	78	16	14	2	2
SS	71	23	7	9	2
LS	72	12	15	7	6
24	Welt	Planet	Ball	Vase	fehlt
GS	78	16	14	2	2
SS	66	12	24	8	2
LS	83	7	9	7	6

Tab. 22b-B 3.2.2: Distraktorenverteilung Verben					
Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
1	sagen	fragen	schweigen	singen	fehlt
GS	53	47	0	10	2
SS	39	35	15	19	4
LS	45	20	4	37	6
2	gehen	stehen	rennen	sehen	fehlt
GS	99	0	9	2	2
SS	75	14	9	10	4
LS	50	5	14	37	6
3	fahren	reiten	feiern	frieren	fehlt
GS	109	0	0	1	2
SS	83	8	7	10	4
LS	73	4	21	8	6
4	essen	trinken	zähneputzen	messen	fehlt
GS	107	1	2	0	2
SS	93	3	5	7	4
LS	93	2	7	4	6
5	fliegen	springen	winken	fragen	fehlt
GS	109	1	0	0	2
SS	80	11	9	8	4
LS	87	12	2	5	6
6	gucken	schminken	denken	ducken	fehlt
GS	75	35	0	0	2
SS	30	34	29	15	4
LS	44	32	12	18	6
7	springen	sitzen	klettern	schimpfen	fehlt
GS	106	0	4	0	2
SS	70	12	12	14	4
LS	83	7	6	10	6
8	schlafen	schlagen	streicheln	liegen	fehlt
GS	84	1	0	25	2
SS	80	12	4	12	4
LS	82	6	2	16	6
9	denken	träumen	warnen	schenken	fehlt
GS	50	5	54	1	2
SS	25	49	21	13	4
LS	31	18	31	26	6
10	hinsetzen	sitzen	hinlegen	schwitzen	fehlt
GS	40	59	11	0	2
SS	23	56	17	12	4
LS	42	41	8	15	6

Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
11	anziehen	ausziehen	heben	ziehen	fehlt
GS	87	18	4	1	2
SS	53	12	20	23	4
LS	64	15	17	10	6
12	warten	sitzen	wandern	braten	fehlt
GS	82	7	17	4	2
SS	62	26	10	10	4
LS	61	8	22	15	6
13	verstecken	schlafen	erschrecken	stecken	fehlt
GS	96	11	3	0	2
SS	61	14	15	18	4
LS	73	5	21	7	6
14	weinen	reiben	schwitzen	wiegen	fehlt
GS	106	1	3	0	2
SS	81	11	7	9	4
LS	82	6	13	5	6
15	werfen	fangen	strecken	wärmen	fehlt
GS	106	1	2	1	2
SS	71	26	5	6	4
LS	76	11	5	14	6
16	schreiben	malen	reiben	schlafen	fehlt
GS	106	2	2	0	2
SS	77	22	3	6	4
LS	90	5	7	4	6
17	lesen	sitzen	halten	blasen	fehlt
GS	105	1	3	1	2
SS	83	6	8	11	4
LS	82	6	6	12	6
18	singen	sagen	warnen	flöten	fehlt
GS	93	0	4	13	2
SS	45	22	14	27	4
LS	53	18	12	23	6
19	waschen	wischen	schwimmen	warten	fehlt
GS	107	2	1	0	2
SS	66	7	24	11	4
LS	76	10	18	2	6
20	brennen	löschen	rechnen	rennen	fehlt
GS	103	3	4	0	2
SS	52	30	13	13	4
LS	59	18	20	9	6

Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
21	schieben	schließen	heben	ziehen	fehlt
GS	103	3	3	1	2
SS	33	35	11	29	4
LS	69	9	13	15	6
22	stolpern	springen	stoßen	streicheln	fehlt
GS	77	29	4	0	2
SS	34	21	37	16	4
LS	45	13	29	19	6
23	stinken	riechen	staunen	winken	fehlt
GS	105	5	0	0	2
SS	48	19	26	15	4
LS	87	8	7	4	6
24	zittern	schütteln	lachen	zwicken	fehlt
GS	53	16	28	13	2
SS	15	43	27	23	4
LS	30	16	26	34	6

Tab. 22c-B 3.2.2: Distraktorenverteilung Adjektive					
Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
1	ganz	rund	gerade	halb	fehlt
GS	50	40	9	11	2
SS	39	17	20	32	4
LS	28	14	46	18	6
2	klein	klug	dünn	fein	fehlt
GS	98	2	7	3	2
SS	79	9	11	9	4
LS	55	8	7	36	6
3	blau	grau	braun	schwarz	fehlt
GS	101	1	8	0	2
SS	91	1	15	1	4
LS	103	0	3	0	6
4	voll	viel	groß	stolz	fehlt
GS	48	7	7	48	2
SS	58	13	22	15	4
LS	62	11	6	27	6
5	blond	braun	rot	schwarz	fehlt
GS	50	28	16	16	2
SS	79	16	1	12	4
LS	61	23	13	9	6

Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
6	spitz	struppig	schief	spannend	fehlt
GS	100	2	4	4	2
SS	71	8	18	11	4
LS	75	7	11	13	6
7	nass	trocken	blass	nackt	fehlt
GS	60	42	5	3	2
SS	63	15	13	17	4
LS	70	13	11	12	6
8	hellblau	dunkelblau	hellgrau	hellgrün	fehlt
GS	87	13	9	1	2
SS	71	15	7	15	4
LS	86	7	9	5	6
9	orange	rosa	gelb	rot	fehlt
GS	100	8	1	1	2
SS	65	10	3	30	4
LS	64	17	1	25	6
10	dünn	eng	dick	dumm	fehlt
GS	102	3	2	3	2
SS	68	11	13	16	4
LS	54	30	13	9	6
11	schwer	stark	leer	schwarz	fehlt
GS	19	90	1	0	2
SS	56	10	16	26	4
LS	464	57	1	4	6
12	traurig	ernst	erstaunt	fröhlich	fehlt
GS	73	28	8	1	2
SS	67	19	9	13	4
LS	75	17	7	7	6
13	verletzt	scharf	gesund	verliebt	fehlt
GS	49	59	1	1	2
SS	52	20	15	21	4
LS	63	11	9	23	6
14	warm	heiß	kalt	weich	fehlt
GS	87	22	1	0	2
SS	62	19	14	13	4
LS	70	25	4	7	6
15	schief	steil	gerade	stark	fehlt
GS	75	25	10	0	2
SS	58	16	15	19	4
LS	45	34	16	11	6

Nr.	Item	Distr. I	Distr. II	Distr. III	nicht beantw.
16	glatt	sauber	lockig	platt	fehlt
GS	105	3	0	2	2
SS	58	19	16	15	4
LS	84	6	7	9	6
17	links	rechts	oben	unten	fehlt
GS	21	44	9	36	2
SS	33	34	30	11	4
LS	32	26	32	16	6
18	frei	erschrocken	frech	gefangen	fehlt
GS	45	32	9	24	2
SS	70	5	19	14	4
LS	42	21	24	19	6
19	schmutzig	sauber	zerknittert	zerrissen	fehlt
GS	85	6	9	10	2
SS	73	19	5	11	4
LS	81	16	7	2	6
20	weich	nass	wach	warm	fehlt
GS	65	31	11	3	2
SS	56	12	18	22	4
LS	43	9	36	18	6
21	leise	süß	heiß	laut	fehlt
GS	101	0	1	8	2
SS	64	10	12	22	4
LS	89	6	0	11	6
22	gestreift	kariert	geblümt	gepunktet	fehlt
GS	79	16	7	8	2
SS	24	34	21	29	4
LS	53	20	19	14	6
23	viereckig	dreieckig	fünfeckig	sechseckig	fehlt
GS	101	5	2	2	2
SS	70	8	14	16	4
LS	57	21	14	14	6
24	nah	geschlossen	fern	nass	fehlt
GS	51	52	6	1	2
SS	31	32	34	11	4
LS	46	19	16	25	6

Anh. 1: B 3.3a:

Tab. 23-B 3.3: Punkteverteilung in Kategorien								
Subtest- teil	1-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24
GSN				4	12	40	35	19
GSV			1	2	6	20	51	30
GSA				20	34	25	27	4
SSN			13	19	20	19	14	25
SSV	2	16	21	11	18	19	16	5
SSA	2	10	25	16	17	10	9	19
LSN		4	6	14	17	27	25	13
LSV		4	13	16	17	23	19	7
LSA		8	17	21	18	17	15	10

Anh. 1: B 3.3b:

Abb. 11a-B 3.3: Kategorisierte Verteilungskurve GSN

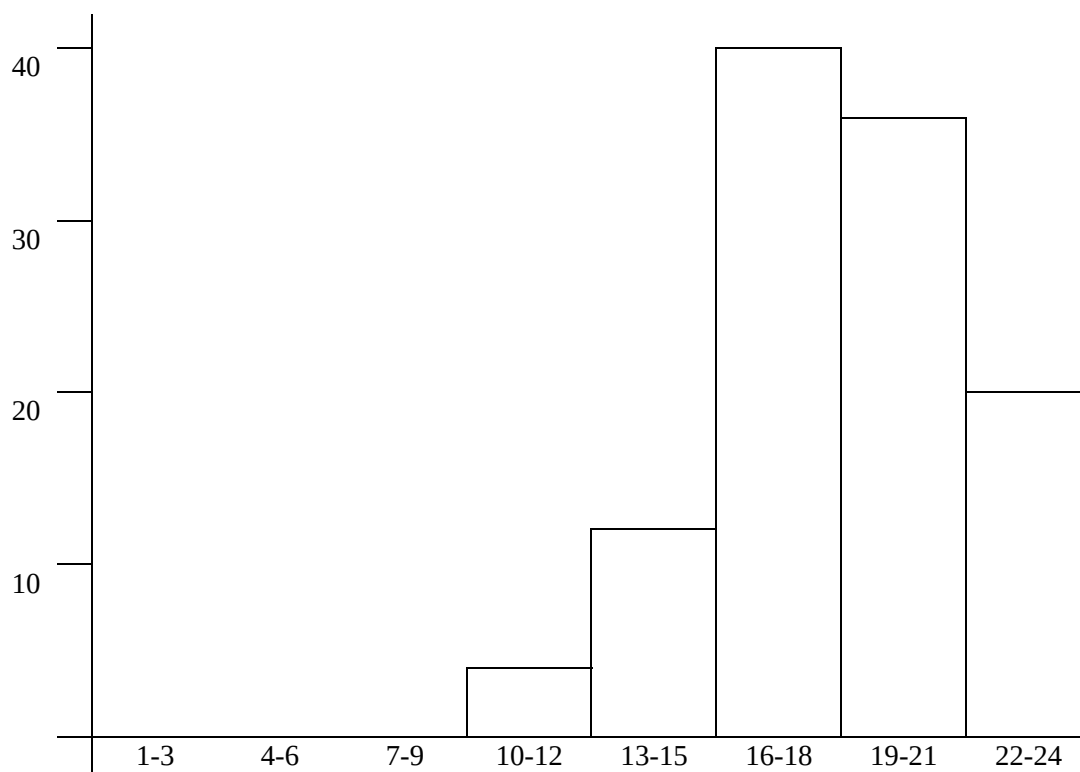


Abb. 11b-B 3.3: Kategorisierte Verteilungskurve GSV

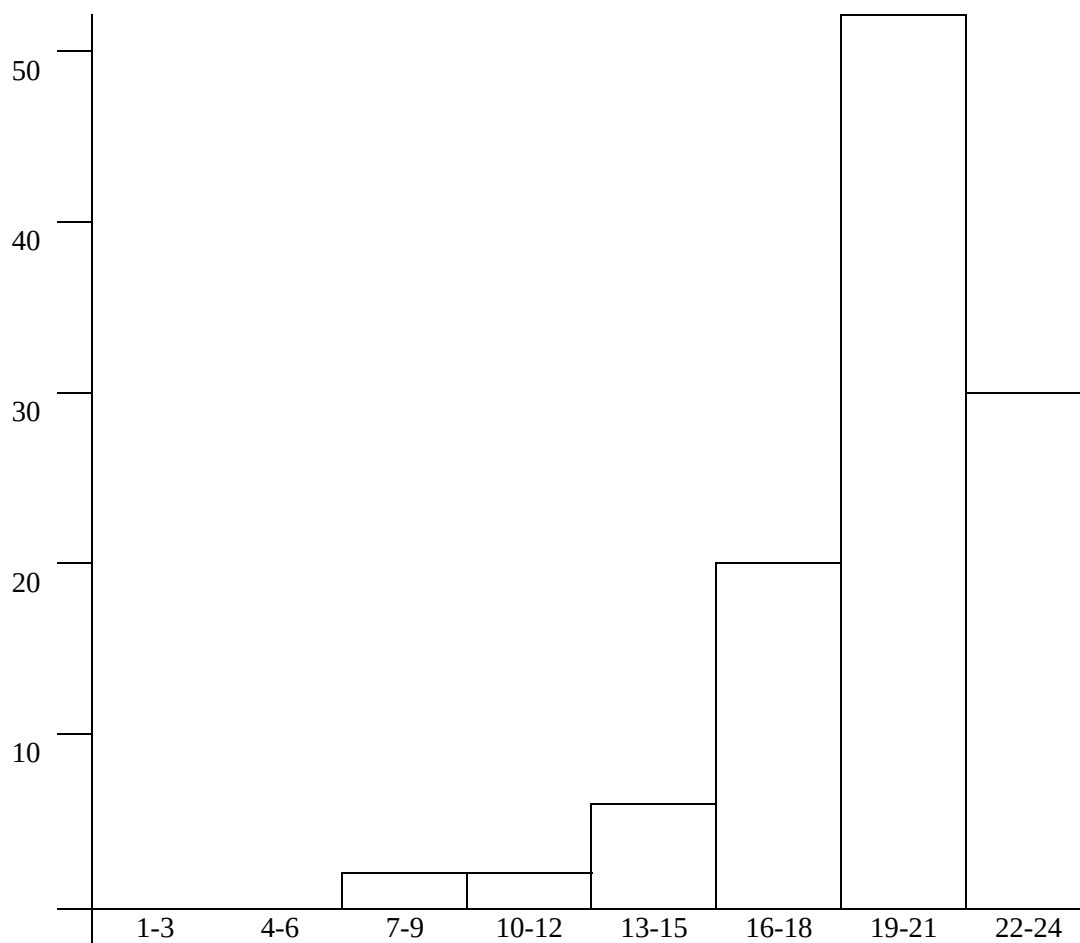


Abb. 11c-B 3.3: Kategorisierte Verteilungskurve GSA

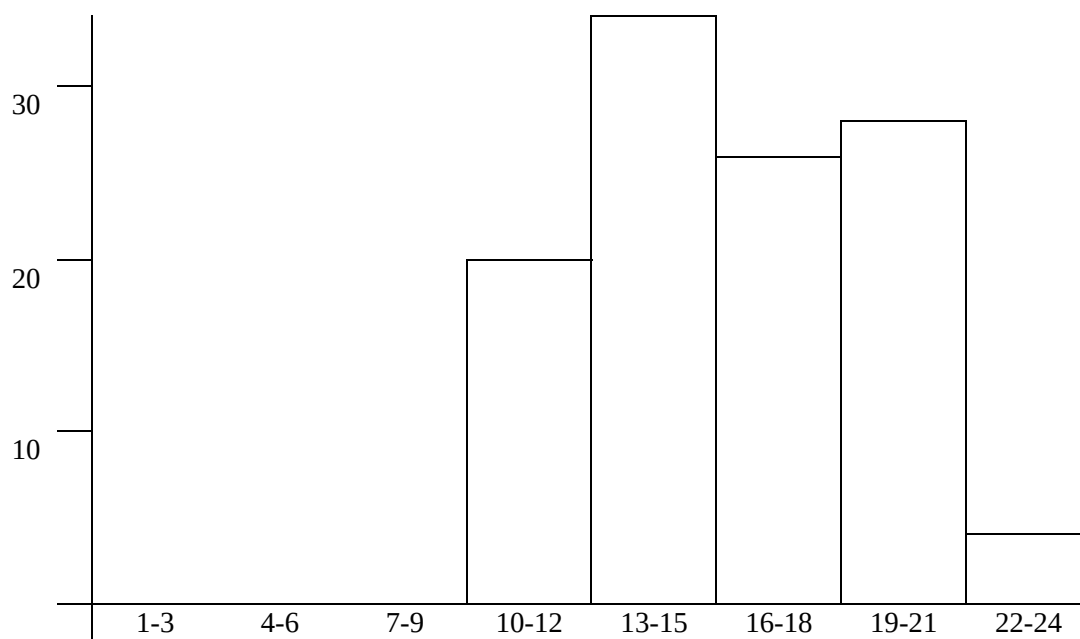


Abb. 11d-B 3.3: Kategorisierte Verteilungskurve SSN

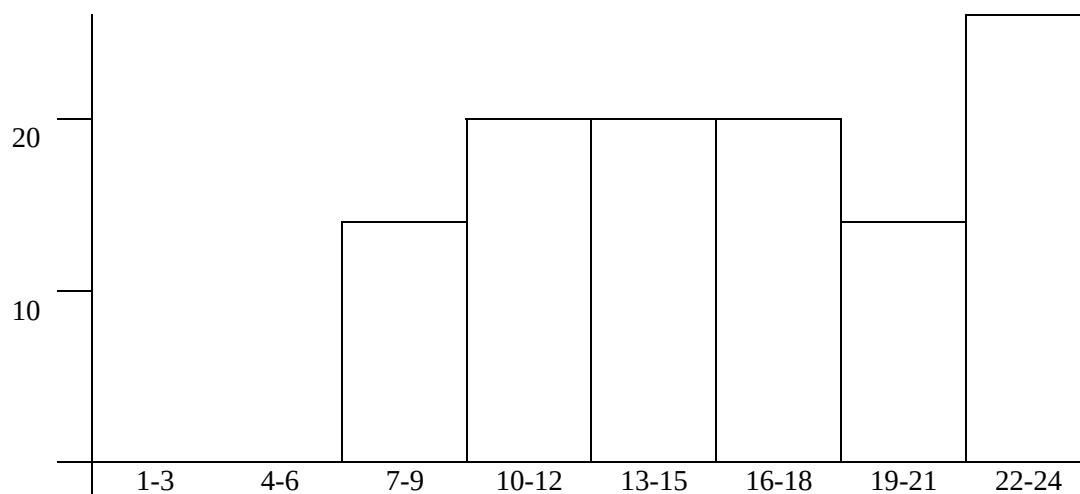


Abb. 11e-B 3.3: Kategorisierte Verteilungskurve SSV

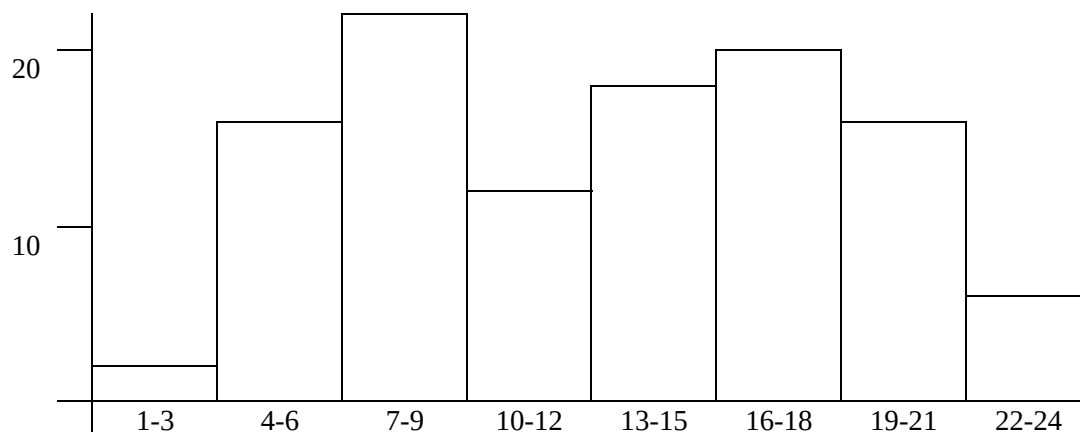


Abb. 11f-B 3.3: Kategorisierte Verteilungskurve SSA

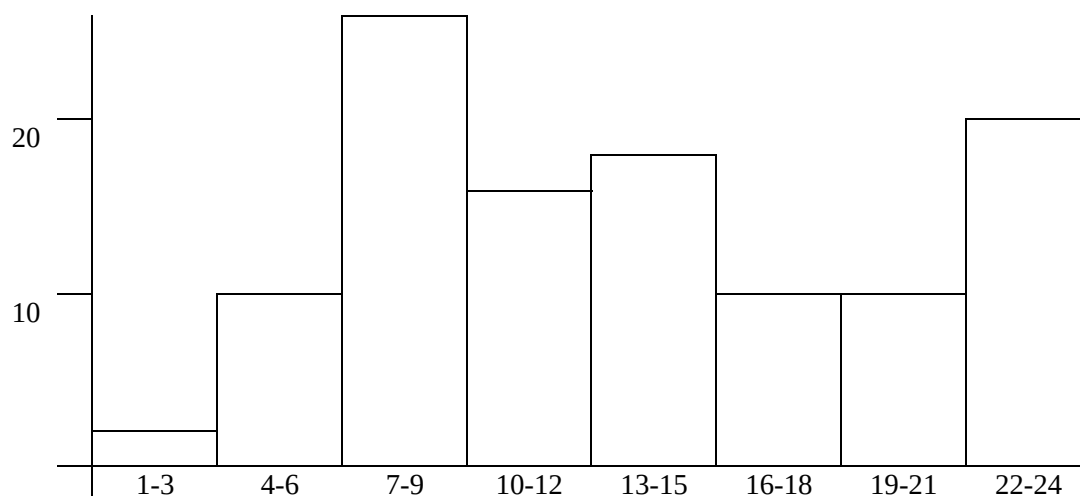


Abb. 11g-B 3.3: Kategorisierte Verteilungskurve LSN

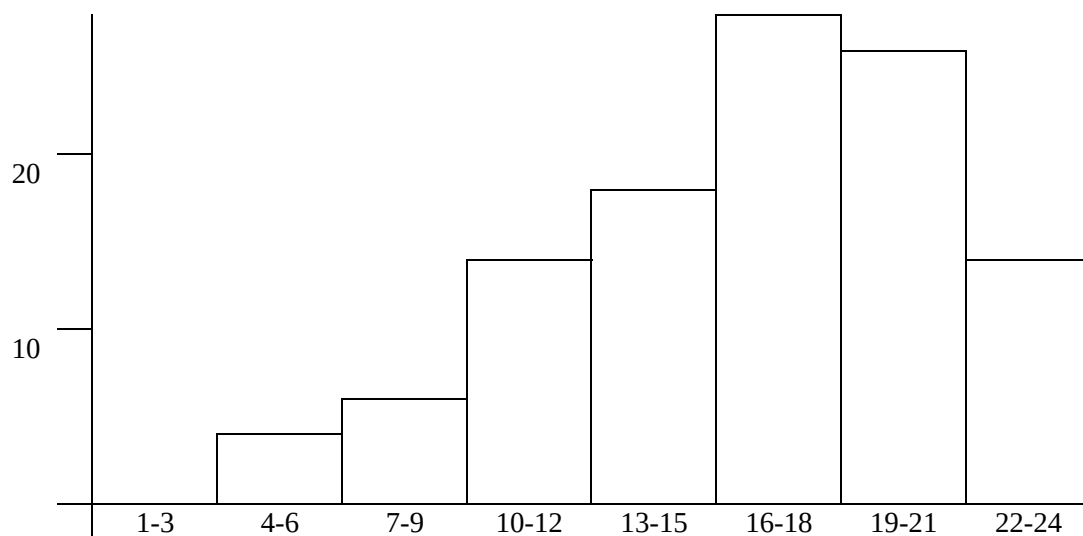


Abb. 11h-B 3.3: Kategorisierte Verteilungskurve LSV

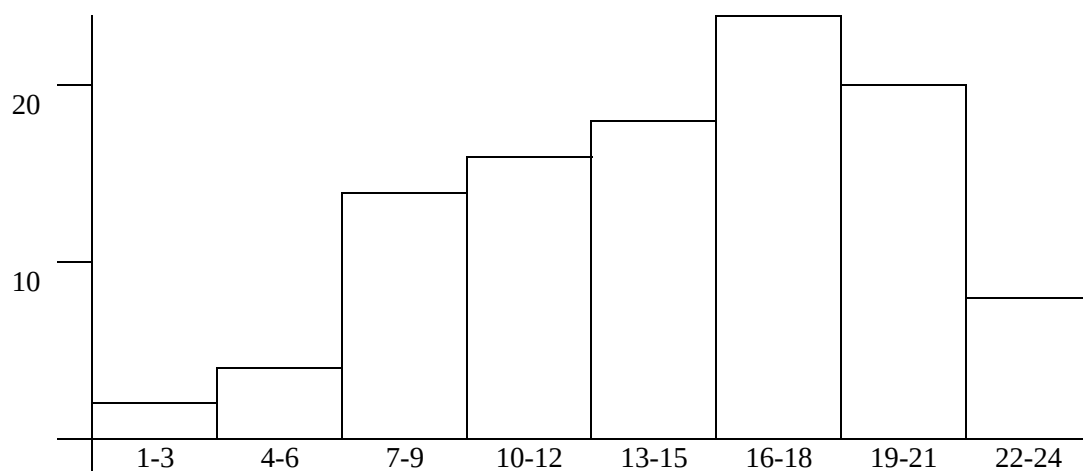
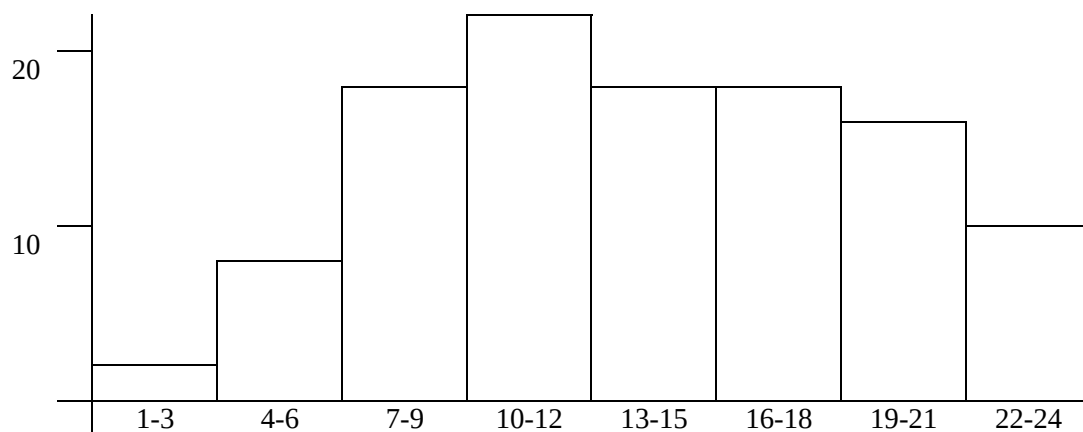
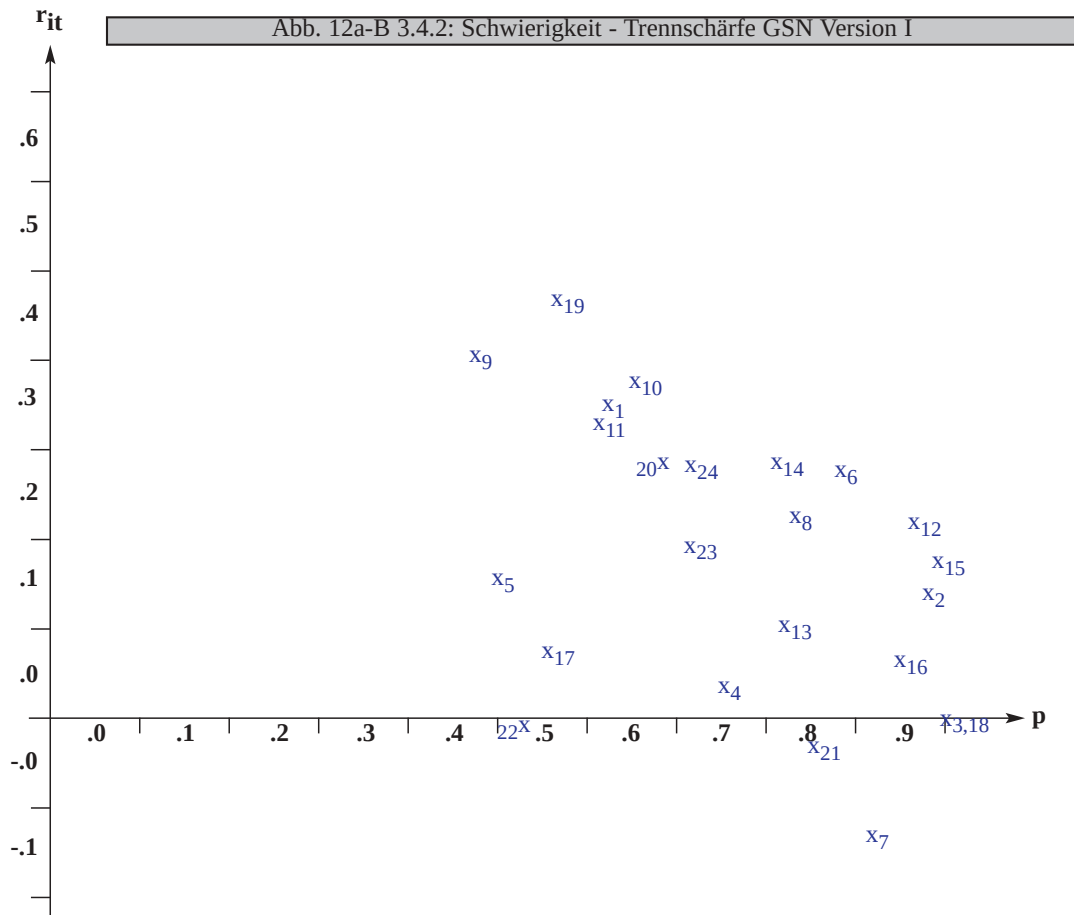


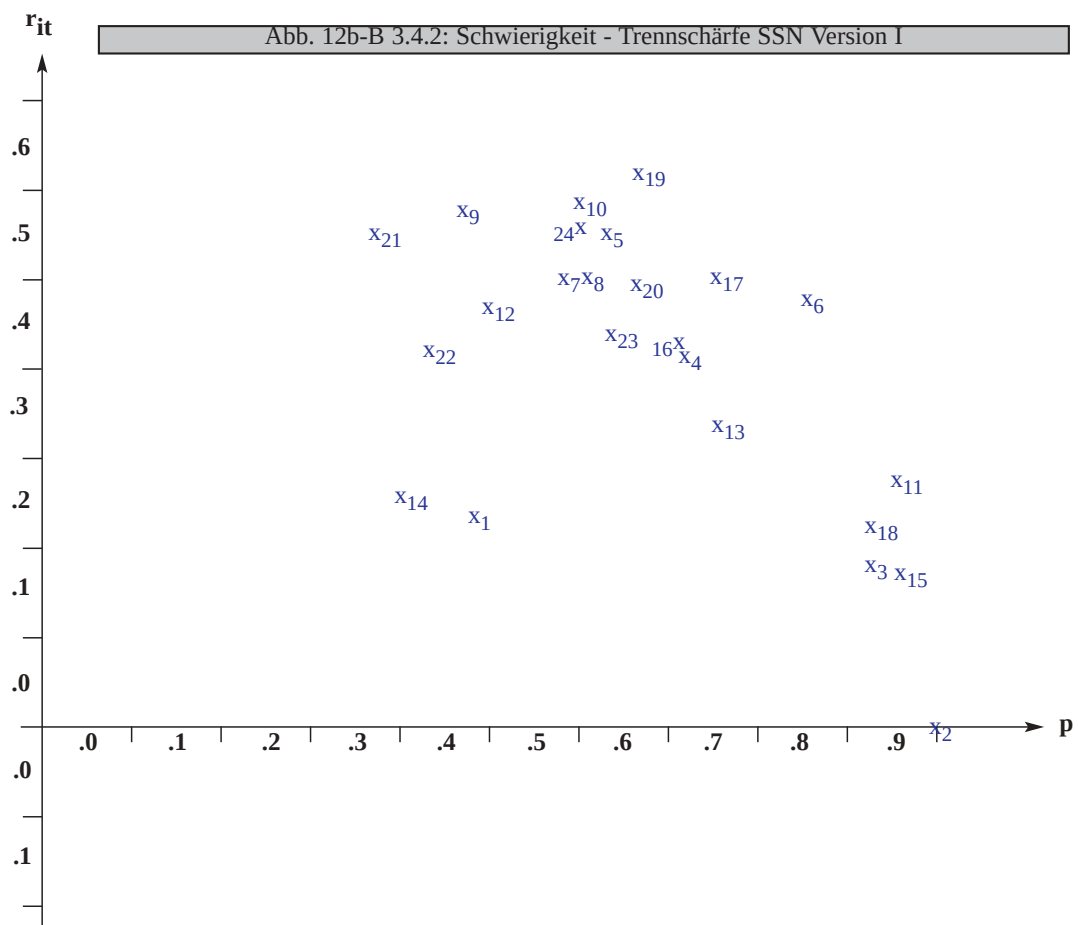
Abb. 11i-B 3.3: Kategorisierte Verteilungskurve LSA

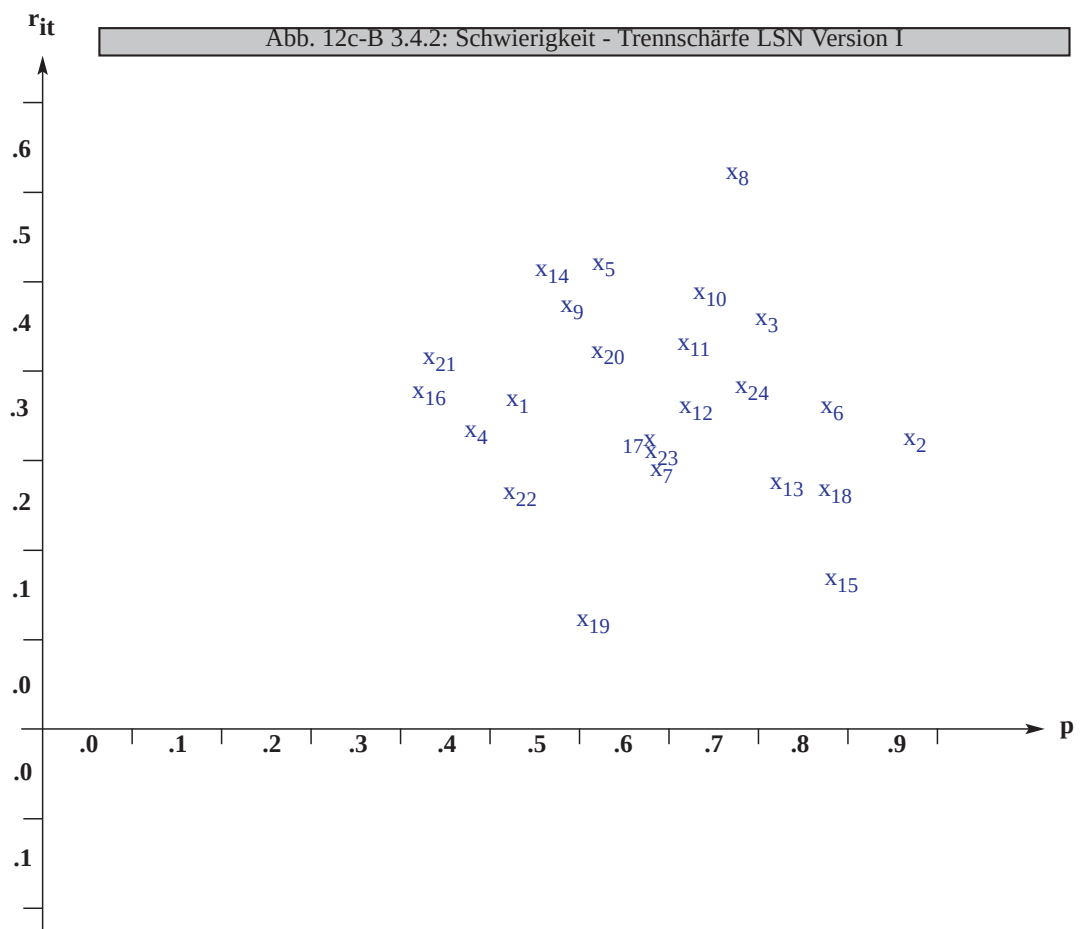


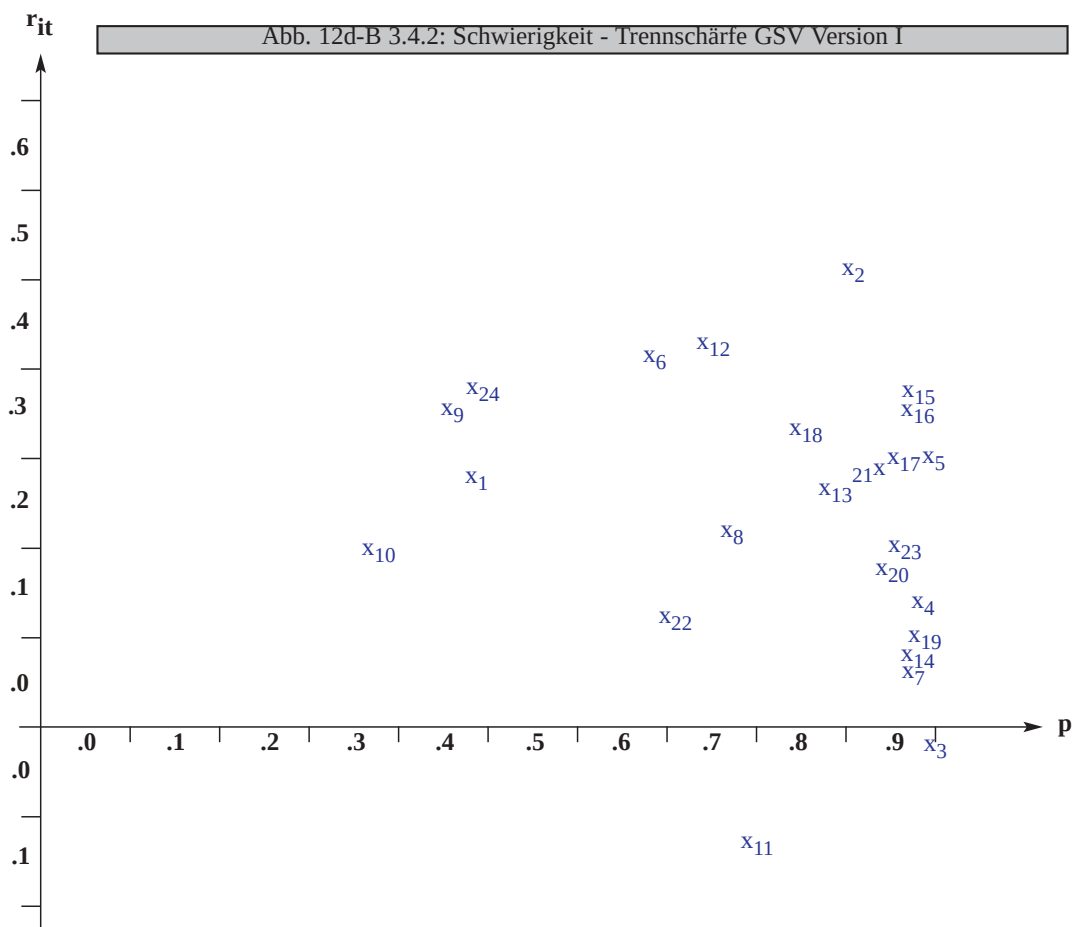
Anh. 1: B 3.4.2:

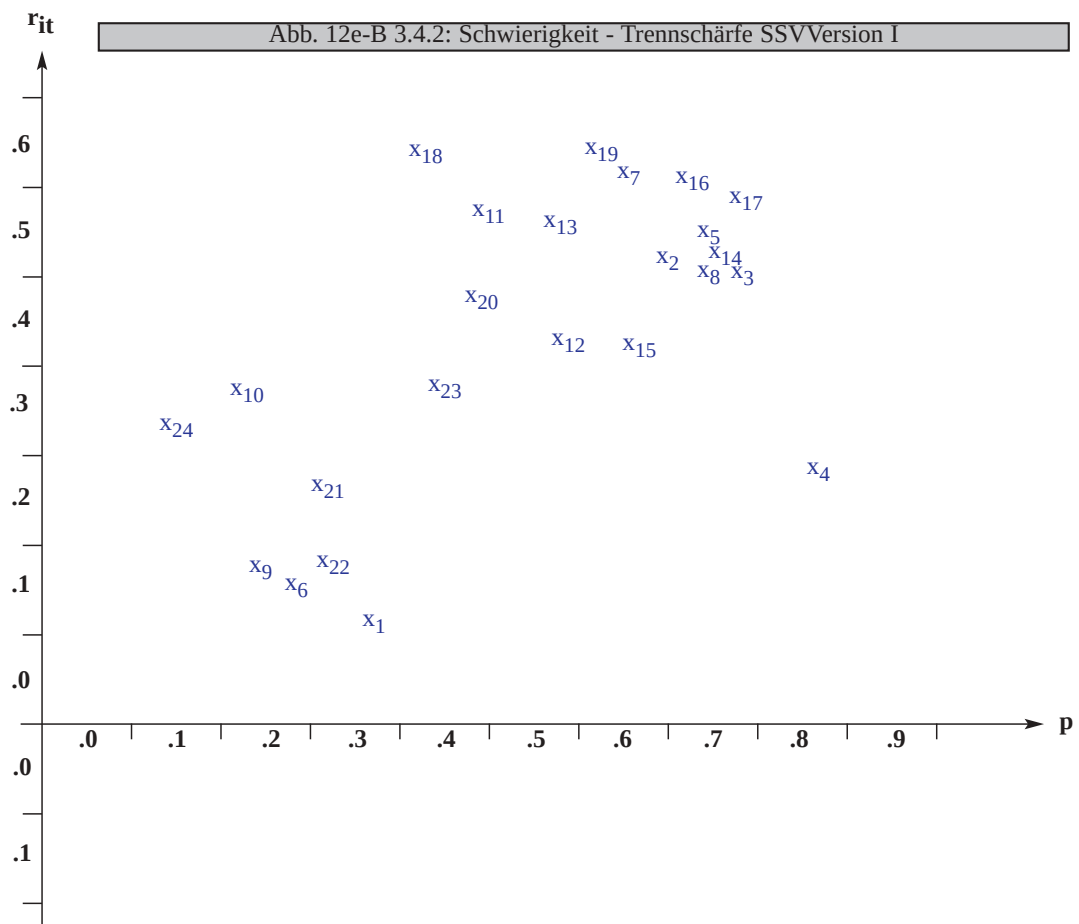
Die folgenden Koordinatensysteme verdeutlichen anschaulich die Beziehungen der Trennschärfe- und Schwierigkeitsindizes der 24 Aufgaben eines jeden Subtestteils der Testversion I. Die Schwierigkeit ist dabei auf der Abzisse, die Trennschärfe auf der Ordinate abgetragen.

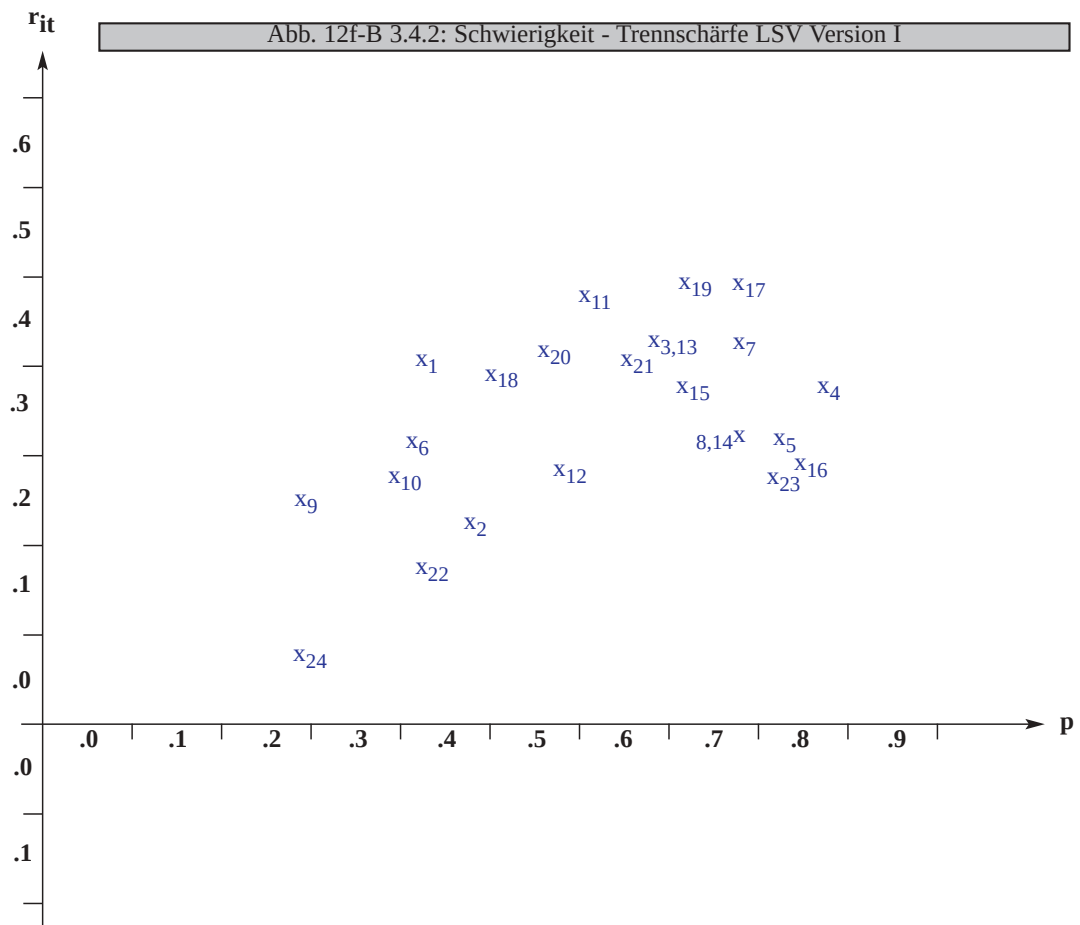


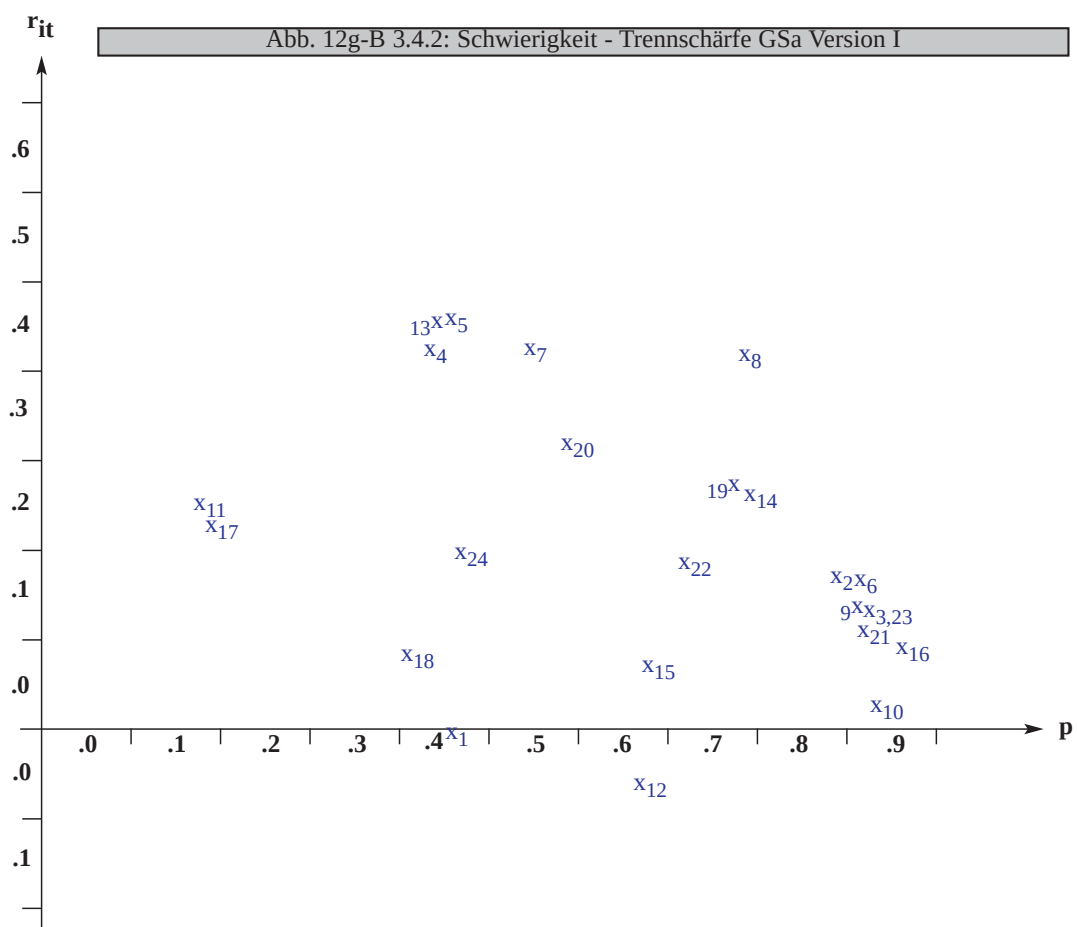


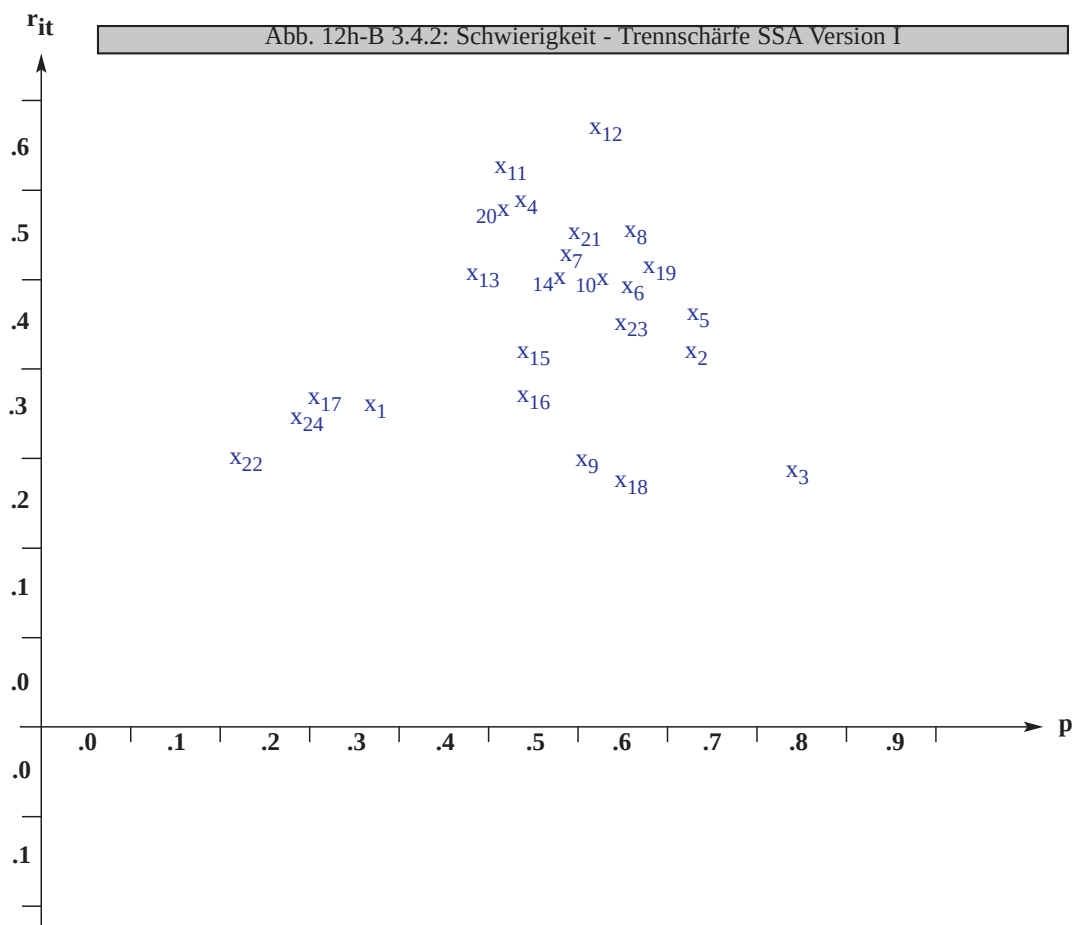


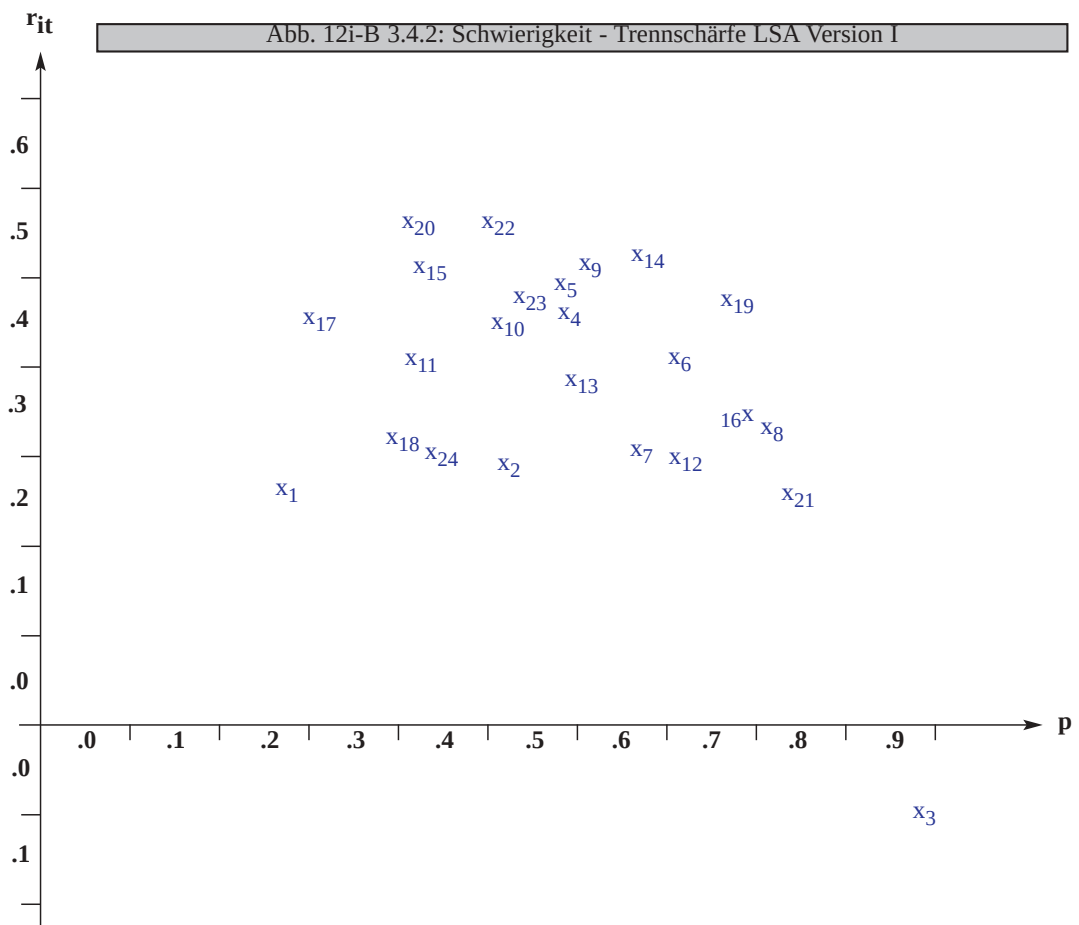












Anh. 1: B 3.4.3:

Nur wenige der folgenden Interkorrelationen sind so signifikant, dass sie einen wahren Zusammenhang abbilden. Nur diejenigen Aufgaben, die wenigstens auf dem 5%- (*) oder auf dem 1%-Niveau (**) gesichert sind, liefern einen interpretierbaren Wert.

Tab. 26-B 3.4.3 Aufgaben-Interkorrelation Subtestteil Gebärdensprache

Nomen

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1	04	-	04	21*	03	02	12	24*	19*	15	-05	05	15	-07	-10	14	-	23*	18	00	-02	25**	13
2		1	-	08	14	-05	-04	-06	-01	19*	17	34**	12	11	70**	-03	02	-	16	-09	-06	-12	-09	06
3			1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4				1	02	14	-03	09	03	-02	02	00	08	10	17	-14	-13	-	10	08	-00	-01	-09	-09
5					1	06	-06	05	22*	-04	06	00	02	-02	10	00	06	-	07	08	05	-04	04	12
6						1	-01	00	22*	11	08	-07	15	20*	-03	04	-02	-	34**	15	02	21*	-10	16
7							1	-14	-14	-10	-12	-06	02	01	-03	-08	-09	-	-15	06	-04	-04	08	01
8								1	07	14	20*	05	06	10	-04	00	05	-	21*	02	-04	09	20*	04
9									1	19*	27**	09	-05	14	09	15	01	-	36**	20*	08	02	09	25**
10										1	20*	27**	-03	28**	13	16	12	-	40**	22*	-03	-09	08	25**
11											1	24*	-07	09	12	22*	07	-	27**	-01	04	09	06	23*
12												1	04	28**	49**	17	02	-	12	08	-08	-18	20*	09
13													1	02	21	-11	12	-	08	11	-12	07	-03	13
14														1	20*	-12	03	-	23	13	-14	-03	15	20*
15															1	-02	-09	-	11	-06	-04	-09	-06	15
16																1	-05	-	11	10	02	-14	-07	29**
17																	1	-	06	11	-06	10	11	-3**
18																		1	-	-	-	-	-	-
19																			1	25**	00	-18	08	24**
20																				1	06	-10	14	09
21																					1	04	-09	02
22																						1	14	-02
23																							1	03
24																								1

Anh. 1: B 3.4.4:

Tab. 27a-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil					
Gebärdensprache Nomen					
Aufgabe	P	r _{it}	s _i	r _{it} ·s _i	Rang
1	.63	.35	.48	.17	4
2	.98	.14	.14	.02	15
3	1	.00	.00	.00	21
4	.75	.04	.43	.02	18
5	.50	.16	.50	.08	12
6	.89	.28	.31	.09	10
7	.91	- .13	.29	-.04	24
8	.84	.22	.37	.08	11
9	.47	.40	.50	.20	2
10	.66	.37	.48	.18	3
11	.61	.33	.49	.16	5
12	.96	.22	.19	.04	13
13	.83	.10	.38	.04	14
14	.81	.29	.39	.11	8
15	.99	.18	.09	.02	17
16	.95	.07	.23	.02	19
17	.56	.07	.25	.02	16
18	1	.00	.00	.00	20
19	.56	.47	.50	.23	1
20	.69	.29	.46	.13	6
21	.86	- .03	.35	-.01	23
22	.54	- .01	.50	-.004	22
23	.71	.20	.45	.09	9
24	.71	.29	.45	.13	7

Tab. 27b-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil					
Schriftsprache Nomen					
Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
1	.47	.24	.50	.12	19
2	1	.00	.00	.00	24
3	.93	.19	.26	.05	21
4	.72	.41	.45	.19	15
5	.64	.55	.48	.27	5
6	.86	.47	.35	.17	16
7	.58	.50	.49	.25	7
8	.61	.50	.49	.25	8
9	.46	.58	.50	.29	2
10	.60	.58	.49	.29	3
11	.96	.27	.15	.04	22
12	.50	.47	.50	.23	10
13	.76	.34	.43	.14	17

Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
14	.40	.26	.49	.13	18
15	.96	.17	.15	.03	23
16	.71	.44	.45	.20	14
17	.76	.50	.43	.22	11
18	.93	.23	.26	.06	20
19	.66	.62	.47	.29	1
20	.66	.50	.47	.24	9
21	.36	.55	.48	.26	6
22	.44	.42	.50	.21	13
23	.65	.44	.48	.21	12
24	.60	.56	.49	.28	4

Tab. 27c-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil Lautsprache Nomen					
Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
1	.52	.37	.50	.18	10
2	.96	.32	.19	.06	23
3	.80	.46	.40	.18	9
4	.48	.33	.50	.17	11
5	.61	.53	.49	.26	2
6	.88	.36	.33	.12	17
7	.69	.30	.46	.14	15
8	.76	.63	.43	.27	1
9	.59	.47	.49	.23	4
10	.74	.49	.44	.22	5
11	.72	.44	.20	.09	19
12	.72	.36	.20	.07	21
13	.82	.27	.38	.10	18
14	.56	.51	.50	.25	3
15	.89	.16	.32	.05	24
16	.42	.38	.49	.19	8
17	.68	.32	.47	.15	13
18	.88	.27	.33	.09	20
19	.60	.13	.49	.06	22
20	.62	.43	.49	.21	6
21	.43	.41	.50	.20	7
22	.52	.26	.50	.13	16
23	.68	.31	.47	.15	14
24	.78	.38	.41	.16	12

Tab. 27d-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil Gebärdensprache Verben					
Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
1	.48	.29	.50	.14	6
2	.90	.52	.30	.16	5
3	.99	-.01	.30	-.004	23
4	.97	.14	.16	.02	19
5	.99	.30	.300	.09	10
6	.68	.41	.47	.19	1
7	.96	.07	.18	.01	22
8	.76	.22	.43	.09	9
9	.46	.36	.50	.18	4
10	.36	.20	.48	.10	8
11	.79	-.13	.41	-.05	24
12	.75	.44	.44	.19	2
13	.87	.27	.33	.09	11
14	.96	.08	.19	.02	21
15	.96	.38	.19	.07	13
16	.96	.36	.19	.07	14
17	.96	.30	.21	.06	15
18	.85	.33	.36	.12	7
19	.97	.10	.16	.02	20
20	.94	.18	.25	.04	17
21	.94	.29	.25	.07	12
22	.70	.13	.46	.06	16
23	.96	.20	.21	.04	18
24	.48	.38	.50	.19	3

Tab. 27e-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil Schriftsprache Verben					
Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
1	.36	.11	.48	.05	24
2	.69	.53	.46	.24	8
3	.77	.51	.42	.21	14
4	.86	.28	.35	.10	20
5	.74	.55	.44	.24	10
6	.28	.15	.45	.07	23
7	.65	.62	.48	.29	3
8	.74	.51	.44	.22	12
9	.23	.18	.42	.08	22
10	.21	.38	.41	.15	17
11	.49	.57	.50	.29	4
12	.57	.44	.49	.22	13
13	.57	.56	.50	.28	5
14	.75	.53	.43	.23	11
15	.66	.43	.48	.20	15

Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
16	.71	.61	.45	.28	6
17	.77	.59	.42	.25	7
18	.42	.64	.49	.32	1
19	.61	.64	.49	.31	2
20	.48	.48	.50	.24	9
21	.31	.27	.46	.12	18
22	.32	.18	.47	.08	21
23	.44	.38	.50	.19	16
24	.14	.33	.35	.12	19

Tab. 27f-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil Lautsprache Verben					
Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
1	.43	.40	.49	.20	7
2	.47	.23	.50	.12	19
3	.69	.43	.46	.20	6
4	.88	.37	.33	.12	17
5	.82	.32	.38	.12	18
6	.42	.31	.49	.16	12
7	.78	.43	.41	.18	10
8	.77	.32	.42	.13	15
9	.29	.25	.46	.11	20
10	.40	.28	.49	.14	14
11	.60	.48	.49	.23	1
12	.58	.28	.49	.14	13
13	.69	.44	.46	.20	5
14	.77	.32	.42	.13	16
15	.72	.37	.45	.17	11
16	.85	.29	.36	.10	22
17	.77	.50	.42	.21	3
18	.50	.39	.50	.19	9
19	.72	.50	.45	.22	2
20	.56	.42	.50	.21	4
21	.65	.41	.48	.19	8
22	.43	.17	.49	.09	23
23	.82	.27	.38	.11	21
24	.28	.07	.45	.03	24

Tab. 27g-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil Gebärdensprache Adjektive					
Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
1	.46	- .001	.50	-.005	23
2	.89	.17	.31	.05	13
3	.92	.14	.27	.04	18
4	.44	.43	.50	.21	4
5	.46	.46	.50	.23	1
6	.91	.16	.29	.05	14
7	.55	.43	.50	.22	3
8	.79	.42	.41	.17	5
9	.91	.14	.29	.04	16
10	.93	.03	.26	.01	22
11	.17	.25	.38	.10	10
12	.66	- .06	.47	-.03	24
13	.45	.45	.50	.23	2
14	.79	.27	.41	.11	8
15	.68	.07	.47	.03	19
16	.96	.09	.21	.02	21
17	.19	.24	.39	.09	11
18	.41	.08	.49	.04	15
19	.77	.27	.42	.11	7
20	.59	.31	.49	.15	6
21	.92	.11	.27	.03	20
22	.72	.19	.45	.09	12
23	.92	.14	.27	.04	17
24	.46	.20	.50	.10	9

Tab. 27h-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil Schriftsprache Adjektive					
Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
1	.36	.36	.48	.17	18
2	.73	.42	.44	.19	16
3	.84	.29	.36	.11	24
4	.54	.58	.50	.29	3
5	.73	.46	.44	.21	15
6	.66	.49	.48	.23	12
7	.58	.53	.49	.26	7
8	.66	.56	.48	.27	6
9	.60	.30	.49	.14	21
10	.63	.50	.48	.24	10
11	.52	.64	.50	.32	2
12	.62	.67	.49	.33	1
13	.48	.51	.50	.24	9
14	.57	.51	.49	.25	8
15	.54	.42	.50	.21	14

Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
16	.54	.36	.50	.18	17
17	.31	.36	.46	.17	19
18	.65	.27	.48	.13	22
19	.68	.51	.47	.24	11
20	.52	.57	.50	.29	4
21	.59	.55	.49	.27	5
22	.22	.30	.42	.12	23
23	.65	.45	.48	.21	13
24	.29	.34	.45	.16	20

Tab. 27i-B 3.4.4: Reliabilitätsindizes der Testaufgaben im Subtestteil Lautsprache Adjektive					
Aufgabe	p	r _{it}	s _i	r _{it} ·s _i	Rang
1	.26	.26	.44	.12	22
2	.52	.29	.50	.15	18
3	.97	-.10	.17	-.02	24
4	.59	.46	.49	.23	9
5	.58	.50	.49	.25	6
6	.71	.41	.46	.19	14
7	.66	.31	.47	.15	17
8	.81	.33	.39	.13	21
9	.60	.51	.49	.25	5
10	.51	.45	.50	.23	8
11	.42	.41	.49	.20	12
12	.71	.30	.46	.14	20
13	.60	.39	.49	.19	13
14	.66	.53	.47	.25	4
15	.43	.51	.49	.25	3
16	.79	.35	.41	.14	19
17	.30	.45	.46	.21	10
18	.40	.32	.49	.16	15
19	.76	.48	.43	.20	11
20	.41	.56	.49	.28	2
21	.84	.26	.37	.10	23
22	.50	.56	.50	.28	1
23	.54	.47	.50	.24	7
24	.43	.30	.50	.15	16

Anh. 1: B 3.4.5a:

Die Trennschärfewerte der Distraktoren sollen im negativen Bereich oder um Null liegen. Als kritisch wurden Distraktoren bewertet, wenn sie in einem Subtestteil einen Wert von über .40 oder in mindestens zwei Subtestteilen Werte über .30 aufweisen. Diese Distraktoren sind durch Schraffierung gekennzeichnet.

Tab. 29a-B 3.4.5: Distraktorentrennschärfe HRP											
NR	GSN	SSN	LSN	NR	GSV	SSV	LSV	NR	GSA	SSA	LSA
1	.17	.04	.08	1	-.06	.02	.18	1	.0	.0	.0
2	.04	.0	.13	2	.0	.0	.0	2	.0	.0	.0
3	.0	.02	.05	3	.0	.0	.0	3	.0	.0	.0
4	-.06	.0	.14	4	.0	.0	.0	4	-.03	.20	.10
5	.03	.10	.05	5	.29	-.06	.15	5	.0	.0	.0
6	.04	.0	-.02	6	.13	.0	.13	6	.0	.30	.0
7	.0	.0	.0	7	.0	.0	.0	7	.0	.0	.0
8	.0	.0	.0	8	.0	-.06	.0	8	.0	.0	.0
9	.05	.36	.16	9	-.01	-.16	.05	9	.0	.0	.0
10	.14	.0	.21	10	.12	.02	-.08	10	-.03	.0	.10
11	.12	.0	.18	11	-.21	.20	.11	11	.11	.0	.33
12	.10	.22	.23	12	-.01	.0	-.04	12	-.01	.07	.04
13	.03	.0	.05	13	.0	.0	.0	13	.17	.0	.02
14	.10	.0	.07	14	.19	.0	.01	14	.17	.25	.20
15	.0	.0	.16	15	.0	.0	.0	15	.10	.0	.14
16	.0	.0	.0	16	.0	.0	.0	16	.0	.0	.0
17	.0	.25	.06	17	.19	.13	.16	17	.0	.0	.0
18	.0	.0	.0	18	.0	.0	.0	18	.0	.0	.0
19	.05	.32	.13	19	.13	-.04	.0	19	.0	.0	.0
20	-.02	.0	.03	20	.29	-.08	.25	20	.0	.0	.0
21	.10	.0	.38	21	.35	.0	.0	21	.0	.0	.0
22	.07	.25	.09	22	.01	.0	.11	22	.0	.0	.0
23	.0	.0	.0	23	-.01	-.08	.02	23	.0	.0	.0
24	.0	.20	-.02	24	.03	-.07	-.01	24	.0	.0	.0
alpha	.15	.40	.29		.28	-.15	.17		.19	.29	.29

Tab. 29b-B 3.4.5: Distraktorentrennschärfe I											
NR	GSN	SSN	LSN	NR	GSV	SSV	LSV	NR	GSA	SSA	LSA
1	.12	.07	.09	1	.15	.10	.31	1	.01	.33	.22
2	.08	.0	.27	2	.37	.34	.12	2	.14	.17	.16
3	.0	-.03	.12	3	.0	.19	.23	3	.13	.34	-.12
4	-.01	.11	.11	4	.16	.31	.39	4	.03	.28	.14
5	.0	.21	.12	5	.34	.31	.32	5	.36	.38	.20
6	.13	.14	.29	6	.35	.15	.15	6	.26	.31	.11
7	-.10	.22	.15	7	.03	.35	.26	7	.22	.21	.04
8	.20	.50	.61	8	.07	.28	.18	8	.30	.56	.31
9	-.01	.46	.16	9	-.06	-.25	.03	9	.17	.26	.45
10	.19	.39	.25	10	.17	.08	.01	10	-.03	.37	.29
11	.17	.06	.23	11	-.16	.14	.16	11	.15	.09	.36
12	.24	.26	.45	12	-.03	.14	-.01	12	-.04	.61	.27
13	.06	.10	.26	13	.26	.19	.30	13	.46	.21	.21
14	.29	.28	.43	14	.06	.38	.19	14	.21	.39	.48
15	.15	.0	.10	15	.22	.46	.12	15	.11	.20	.30
16	-.09	.08	.22	16	.16	.39	.20	16	.02	.17	.18
17	.03	.26	.15	17	.21	.40	.18	17	.27	.30	.44
18	.0	.0	.28	18	.22	.37	.30	18	-.16	.07	.13
19	-.01	.38	.16	19	.14	.39	.43	19	.33	.42	.39
20	.02	.02	.06	20	.26	.18	.27	20	-.05	.10	.29
21	-.11	.37	.14	21	.23	.19	.32	21	.04	.22	.09
22	.11	.32	.08	22	.04	.12	.04	22	.16	.20	.52
23	.19	.38	.34	23	.12	-.20	.15	23	.13	.38	.45
24	.07	.18	.07	24	.03	-.06	.06	24	.03	.13	-.07
alpha	.34	.63	.62		.48	.60	.56		.54	.69	.70

Tab. 29c-B 3.4.5: Distraktorentrennschärfe G											
NR	GSN	SSN	LSN	NR	GSV	SSV	LSV	NR	GSA	SSA	LSA
1	.18	.19	.22	1	.09	.08	-.001	1	.12	.15	.17
2	.0	.0	-.08	2	.0	.22	-.04	2	.03	.19	.06
3	.0	.01	.10	3	.0	.19	.11	3	.14	-.001	.0
4	.11	.29	.15	4	-.01	.19	.21	4	.40	.41	.33
5	.13	.32	.24	5	.0	.03	-.06	5	.0	.0	.0
6	.20	.0	.23	6	.31	.13	.11	6	.09	.12	.19
7	-.11	.17	.16	7	.0	.35	.12	7	.30	.16	.12
8	.26	.11	.14	8	.14	.24	.30	8	.33	.45	.16
9	.14	.06	.19	9	.29	-.03	.08	9	.13	.19	.24
10	.15	.09	.07	10	.08	.03	-.04	10	.02	.26	.17
11	.25	.13	.33	11	-.20	.57	.21	11	.23	.18	.48
12	.15	.27	.27	12	.34	-.04	.03	12	.09	.12	.13
13	.11	.12	.16	13	.11	.22	.14	13	.31	.23	.11
14	.08	-.14	.15	14	.14	.21	.12	14	.19	.09	.27
15	.11	.0	-.07	15	.21	.25	.14	15	.10	.17	.15
16	.01	.14	.18	16	.16	.26	.07	16	.07	.30	.14
17	.19	.27	.21	17	.13	.19	.22	17	-.11	.01	-.04
18	.0	.0	.13	18	.23	.29	.15	18	.24	.23	.26
19	.44	.22	.34	19	.06	.12	.002	19	.0	.0	.0
20	.24	.26	.28	20	.05	.11	.12	20	.46	.17	.13
21	-.06	.13	-.01	21	.21	.04	.18	21	.0	.12	.16
22	.05	.01	-.03	22	.16	.21	.06	22	.14	.31	.16
23	.07	.23	.14	23	.15	-.17	.01	23	.06	.35	.30
24	.27	.43	.38	24	.25	.34	-.02	24	.20	-.04	.15
alpha	.52	.51	.59		.47	.42	.31		.57	.59	.56

Tab. 29d-B 3.4.5: Distraktorentrennschärfe M											
NR	GSN	SSN	LSN	NR	GSV	SSV	LSV	NR	GSA	SSA	LSA
1	.28	.21	.15	1	.24	-.05	.36	1	-.19	-.04	.13
2	.0	.0	.0	2	.37	.41	.18	2	.02	.28	.13
3	.0	.20	.40	3	.01	.37	.33	3	.08	.28	-.12
4	-.10	.13	.23	4	.0	.08	.002	4	.24	.54	.36
5	.18	.42	.18	5	.0	.19	.09	5	.16	.37	.33
6	.16	.18	.12	6	.0	-.06	-.01	6	-.04	.39	.22
7	-.07	.37	.15	7	.0	.24	.11	7	.16	.30	.30
8	.14	.09	.19	8	.0	.15	.04	8	.18	.48	.35
9	.24	.16	.35	9	.200	.26	.26	9	.04	.25	.14
10	.24	.12	.31	10	.17	.34	.27	10	.02	.21	.20
11	.18	.24	.33	11	.03	.20	.22	11	.04	.54	.03
12	.0	.25	.09	12	.35	.27	.32	12	.18	.25	.25
13	.0	.30	.13	13	.24	.25	.27	13	-.12	.29	.22
14	.33	.16	.41	14	.0	.26	.09	14	.0	.20	.04
15	.0	.25	.00	15	.41	.06	.15	15	.0	.29	.23
16	-.01	.22	-.1	16	.22	.17	.03	16	.17	.01	.17
17	.03	.41	.06	17	-.10	.30	.40	17	.0	.0	.0
18	.0	.33	.20	18	.0	.23	.25	18	.05	.19	.16
19	.37	.27	.10	19	.14	.26	.14	19	.0	.0	.0
20	.18	.25	.32	20	.06	.17	.20	20	.01	.30	.43
21	.08	.42	.36	21	.37	.10	.25	21	.01	.30	.0
22	.03	.21	.21	22	.21	.14	.12	22	.0	.0	.0
23	.04	.31	.15	23	.0	.46	.23	23	.06	.48	.34
24	-.05	.24	.11	24	.16	-.19	-.03	24	.12	.26	.24
alpha	.43	.63	.51		.45	.62	.55		.28	.73	.62

Tab. 29e-B 3.4.5: Distraktorentrennschärfe MOD											
NR	GSN	SSN	LSN	NR	GSV	SSV	LSV	NR	GSA	SSA	LSA
1	.18	.14	.15	1	.09	-.05	.36	1	.16	.0	.13
2	.0	.0	.0	2	.0	.32	.18	2	.03	.20	.13
3	.0	.17	.40	3	.0	.29	.33	3	.14	.27	-.12
4	.11	.12	.23	4	-.01	.10	.002	4	.40	.14	.36
5	.13	.30	.18	5	.0	.20	.09	5	.0	.0	.33
6	.20	.20	.12	6	.31	-.04	-.01	6	.09	.29	.22
7	-.11	.34	.15	7	.0	.0	.11	7	.30	.14	.30
8	.26	.0	.19	8	.14	.13	.04	8	.33	.37	.35
9	.14	.0	.35	9	.29	.27	.26	9	.13	.0	.14
10	.15	.07	.31	10	.08	.17	.27	10	.02	.18	.20
11	.25	.0	.33	11	-.20	.24	.22	11	.23	.36	.03
12	.15	.0	.09	12	.34	.27	.32	12	.09	.0	.25
13	.11	.28	.13	13	.11	.28	.27	13	.31	.25	.22
14	.08	.0	.41	14	.14	.21	.09	14	.19	.20	.04
15	.11	.22	.004	15	.21	.09	.15	15	.10	.0	.23
16	.01	.25	-.10	16	.16	.16	.03	16	.07	-.02	.17
17	.19	.39	.06	17	.13	.24	.40	17	-.11	.0	.0
18	.0	.36	.20	18	.23	.23	.25	18	.24	.26	.16
19	.44	.0	.10	19	.06	.28	.14	19	.0	.0	.0
20	.24	.0	.32	20	.05	.17	.20	20	.46	.19	.43
21	-.06	.30	.36	21	.21	.11	.25	21	.0	.19	.0
22	.05	.18	.21	22	.16	.16	.12	22	.14	.0	.0
23	.07	.26	.15	23	.15	.43	.23	23	.06	.46	.34
24	.27	.0	.11	24	.25	-.17	-.03	24	.20	.31	.24
alpha	.52	.49	.51		.47	.58	.55		.57	.57	.62

Anh. 1: B 3.4.5b:

Distraktoren, die eine signifikante, positive Korrelation mit dem Ravenergebnis aufzeigen, sind besonders für die Testpersonen attraktiv, die laut Raven als intelligent einzustufen sind. Damit wird die Wahl dieses Distraktors entgegen der Konzeption eines Grundwortschatztests vom kognitiven Leistungsniveau der Testperson bestimmt. Distraktoren mit diesbezüglich kritischen Werten sind durch Schraffierung gekennzeichnet.

Tab. 30a-B 3.4.5: Korrelation Distraktor-Raven HRP											
	GSN	SSN	LSN		GSV	SSV	LSV		GSA	SSA	LSA
1	-.05	-.11	.02	1	-.40*	-.02	-.17	1	-	-	-
2	-.01	-	-.12	2	-	-	-	2	-	-	-
3	-	-.03	-.19	3	-	-	-	3	-	-	-
4	-.11	-	.05	4	-	-	-	4	-.14	-.18*	-.02
5	.00	-.20*	-.04	5	-.11	-.14	-.13	5	-	-	-
6	-.22*	-	-.01	6	-.35*	-	-.08	6	-	-.14	-
7	-	-	-	7	-	-	-	7	-	-	-
8	-	-	-	8	-	-.26*	-	8	-	-	-
9	-.08	-.14	-.19*	9	.07	.12	-.10	9	-	-	-
10	-.15	-	-.23*	10	-.10	-.18*	-.05	10	.02	-	-.03
11	-.14	-	-.22*	11	-.03	-.03	.05	11	-.12	-	-.25*
12	-.16*	-.21*	-.05	12	-.16	-	-.09	12	.23*	-.07	.13
13	-.05	-	.02	13	-	-	-	13	-.22*	-	-.19*
14	-.16*	-	-.34*	14	-.17*	-	.10	14	-.07	-.14	-.08
15	-	-	-.19*	15	-	-	-	15	-.16*	-	-.25*
16	-	-	-	16	-	-	-	16	-	-	-
17	-.09*	-.01	-.20*	17	-.18*	.06	-.01	17	-	-	-
18	-	-	-	18	-	-	-	18	-	-	-
19	.04	-.12	-.02	19	-.10	.04	.07	19	-	-	-
20	.00	-	.04	20	.10	.14	-.03	20	-	-	-
21	.14	-	-.18	21	-.22*	-	-	21	-	-	-
22	-.04	-.10	-.04	22	-.07	-	.08	22	-	-	-
23	-	-	-	23	.05	-.08	.01	23	-	-	-
24	.02	-.04	-.12	24	-.15	-.11	-.03	24	-	-	-

Tab. 30b-B 3.4.5: Korrelation Distraktor-Raven I											
	GSN	SSN	LSN		GSV	SSV	LSV		GSA	SSA	LSA
1	-.05	-.11	.02	1	-.41*	-.04	-.41*	1	-.07	-.28*	-.25*
2	-.05	-	-.14	2	-.14	-.22*	-.15	2	.11	.02	-.05
3	-	-.03	-.19	3	-	-.12	-.16	3	-.01	-.18	.19*
4	-.12	-.11	.05	4	-.20*	-.28*	-.15	4	-.04	-.26*	-.09
5	.00	-.20*	-.04	5	-.11	-.20*	-.13	5	-.08	-.22*	-.08
6	-.22*	-.02	-.06	6	-.35*	-.11	-.08	6	.01	-.19*	-.14
7	-.04	-.15	-.01	7	-.20*	.00	-.20*	7	-.20*	-.09	.03
8	-.21*	-.12	-.16	8	-.06	-.26*	-.06	8	-.23*	-.33*	-.25*
9	-.08	-.14	-.19*	9	.07	.12	-.100	9	-.14	-.07	-.17*
10	-.15	-.18*	-.23*	10	-.02	-.12	-.09	10	.06	-.25*	-.06
11	-.18*	-.08	-.20*	11	-.03	-.03	.05	11	-.12	-.05	-.25*
12	-.16*	-.21*	-.05	12	-.16	-.01	-.09	12	.14	-.22*	.04
13	-.05	-.03	-.11	13	-.08	-.13	.05	13	-.24*	-.04	.02
14	-.16*	-.24*	-.35*	14	-.12	-.14	.18*	14	-.08	-.26*	-.11
15	-.06	-	-.28*	15	-.12	-.24*	-.08	15	-.26*	-.06	-.26*
16	.09	-.16	-.19*	16	-.10	-.18*	-.03	16	-.04	-.03	.05
17	-.09	-.01	-.20*	17	-.18*	-.12	-.01	17	.02	-.22*	-.23*
18	-	-	-.13	18	.04	-.19*	-.15	18	.22*	.03	-.04
19	.04	-.11	-.02	19	-.09	-.19*	.02	19	-.17*	-.32*	-.31*
20	.00	-.03	.04	20	.10	-.14	-.03	20	.10	-.13	.01
21	.05	-.24*	-.15	21	-.26*	-.02	-.20*	21	-.08	-.15	-.17
22	.04	-.06	-.10	22	-.07	.06	.08	22	-.32*	-.24*	-.35*
23	-.21*	-.22*	-.18*	23	.05	-.08	.01	23	-.22*	-.23*	-.20*
24	.02	-.04	-.12	24	-.15	-.11	-.03	24	-.06	-.05	-.02

Tab. 30c-B 3.4.5: Korrelation Distraktor-Raven G											
	GSN	SSN	LSN		GSV	SSV	LSV		GSA	SSA	LSA
1	-.17*	-.13	-.02	1	-.40*	-.03	-.17	1	-.10	.08	-.21*
2	-	-	-.09	2	-	-.13	-.15	2	-.03	-.15	.05
3	-	-.03	-.19	3	-	-.12	-.16	3	-.01	-.11*	-
4	.02	-.08	.06	4	-.20*	-.20*	-.09	4	-.25*	-.42*	-.24*
5	.00	-.20*	-.04	5	-	-.07	-.09	5	-	-	-
6	-.22*	-	-.01	6	-.35*	-.12	-.11	6	.09	-.09	-.01
7	-.04	-.15	-.01	7	-	-.14	-.08	7	-.20*	-.09	.03
8	-.23*	-.07	-.05	8	-.10	-.09	-.03	8	-.23*	-.33*	-.25*
9	-.11	-.03	-.30*	9	-.09	.10	-.09	9	-.01	-.28*	-.16*
10	.14	-.02	.11	10	-.10	-.18*	-.05	10	.02	-.07	-.03
11	-.18*	-.08	-.20*	11	-.08	-.15	-.05	11	-.12	-.05	-.25*
12	-.16*	-.20*	-.06	12	-.08	-.03	-.13	12	-.08	-.05	-.07
13	-.05	-.15	.02	13	-.08	-.23*	.00	13	-.22*	-.13	-.19*
14	-.11	-.11	-.16	14	-.17*	-.09	.10	14	-.07	-.14	-.08
15	-.06	-	-.20*	15	-.06	-.13	-.11	15	-.16*	.01	-.25*
16	.03	-.04	-.24*	16	-.19*	-.25*	-.03	16	-.04	-.22*	.01
17	-.04	-.07	.05	17	-.12	-.06	.15	17	.09	-.01	.01
18	-	-	-.08	18	-.05	-.20*	.01	18	-.32*	-.05	-.11
19	-.22*	-.07	-.16	19	-.10	-.03	.07	19	-	-	-
20	-.23*	-.01	-.11	20	-.08	-.16	.02	20	-.14	-.17*	-.13
21	-.01	-.07	.08	21	-.22*	.17*	-.11	21	-	-.13	-.17
22	.06	.08	-.09	22	-.07	.06	.08	22	-.18*	-.14	-.18*
23	-.23*	-.27*	-.21*	23	.05	-.08	.01	23	-.22*	-.23*	-.20*
24	-.16	-.18*	-.20*	24	-.12	-.08	-.13	24	-.07	-.02	-.18*

Tab. 30d-B 3.4.5: Korrelation Distraktor-Raven M											
	GSN	SSN	LSN		GSV	SSV	LSV		GSA	SSA	LSA
1	.01	-.11	-.07	1	-.41*	-.03	-.40*	1	.02	-.04	-.03
2	-	-	-	2	-.13	-.16	-.11	2	.01	-.22*	-.16*
3	-	-.29*	-.17	3	.09	-.25*	-.01	3	-.004	-.20*	.19*
4	-.12	-.09	.05	4	-	-.17	-.12	4	-.25*	-.42*	-.24*
5	-.06	-.07	-.22*	5	-	-.05	-.05	5	.07	-.16*	-.06
6	.15*	-.19*	.10	6	-	.14	-.03	6	-.05	-.15	-.11
7	-.07	-.28*	-.19*	7	-	-.03	.07	7	-.13	-.23*	-.03
8	-.22*	-.07	-.05	8	-.10	-.01	.00	8	-.23*	-.33*	-.25*
9	.03	-.05	-.05	9	-.11	-.24*	-.08	9	-.01	-.28*	-.16*
10	-.26*	-.06	-.04	10	-.02	-.24*	-.11	10	-.14	-.09	-.18*
11	.09	-.18*	-.05	11	-.07	-.05	-.07	11	-.10	-.27*	.04
12	-	-.19*	.01	12	-.13	-.20*	-.09	12	-.08	-.05	-.07
13	-	.00	-.17*	13	-.08	-.16	-.06	13	.00	-.02	-.26*
14	-.04	-.14	-.01	14	-	-.16	-.03	14	-	-.08	-.03
15	-	-.18*	-.10	15	-.18*	-.04	-.20*	15	-	-.09	.03
16	-.05	.08	.09	16	-.17*	-.20*	-.06	16	-.06	.11	-.16*
17	-.09	-.08	-.20*	17	-.14	-.11	-.22*	17	-	-	-
18	-	-.15	-.11	18	-	-.13	-.05	18	-.10	.01	-.17
19	-.20*	-.17*	.00	19	-.10	-.15	.12	19	-	-	-
20	-.12	-.09	-.10	20	-.08	-.13	-.06	20	.01	-.29*	-.15
21	.09	-.19*	-.19*	21	-.22*	.17*	-.11	21	.08	-.18	-
22	-.04	-.19	-.15	22	-.25	.02	.12	22	-	-	-
23	.04	.04	.03	23	-	-.12	-.08	23	-.22*	-.23*	-.20*
24	.05	.03	.03	24	-.10	-.05	.00	24	-.04	-.24*	-.17*

Tab. 30e-B 3.4.5: Korrelation Distraktor-Raven MOD											
	GSN	SSN	LSN		GSV	SSV	LSV		GSA	SSA	LSA
1	-.17*	-.11	-.07	1	-.40*	-.03	-.40*	1	-.10	-	-.03
2	-	-	-	2	-	-.16	-.11	2	-.03	-.11	-.16*
3	-	-.29*	-.17	3	-	-.20*	-.01	3	-.01	-.20*	.19*
4	.02	-.09	.05	4	-.20*	-.17	-.12	4	-.25*	-.18*	-.24*
5	.00	-.07	-.22*	5	-	-.05	-.05	5	-	-	-.06
6	-.22*	-.19*	.10	6	-.35*	.14	-.03	6	.09	-.19*	-.11
7	-.04	-.28*	-.19*	7	-	-	.07	7	-.20*	-.23*	-.03
8	-.23*	-	-.05	8	-.10	-.01	.00	8	-.23*	-.33*	-.25*
9	-.11	-	-.05	9	-.09	-.24*	-.08	9	-.01	-	-.16*
10	.14	-.06	-.04	10	-.10	-.12	-.11	10	.02	-.09	-.18*
11	-.18*	-	-.05	11	-.08	-.05	-.07	11	-.12	-.06	.04
12	-.16*	-	.01	12	-.08	-.20*	-.09	12	-.08	-	-.07
13	-.05	.00	-.17*	13	-.08	-.16	-.06	13	-.22*	-.02	-.26*
14	-.11	-	.00	14	-.17*	-.16	-.03	14	-.07	.08	-.03
15	-.06	-.18*	-.10	15	-.06	-.04	-.20*	15	-.16*	-	.03
16	.03	.08	.09	16	-.19*	-.20*	-.06	16	-.04	.11	-.16*
17	-.04	-.08	-.20*	17	-.12	-.11	-.22*	17	.09	-	-
18	-	-.15	-.11	18	-.05	-.13	-.05	18	-.32*	.01	-.17
19	-.22*	-	.00	19	-.10	-.15	.12	19	-	-	-
20	-.23*	-	-.10	20	-.08	-.13	-.06	20	-.14	-.13	-.15
21	-.01	-.19*	-.19*	21	-.22*	.17*	-.11	21	-	-.18	-
22	.06	-.19	-.15	22	-.07	.02	.12	22	-.18*	-	-
23	-.23*	.04	.03	23	.05	-.12	-.08	23	-.22*	-.23*	-.20*
24	-.16	-	.03	24	-.12	-.05	.00	24	-.07	-.24*	-.17*

Anh. 1: B 3.5.3:

Da für die Erstellung der Testversion II keine neuen Aufgaben in den Itemkorpus aufgenommen wurden, sondern lediglich kleine Veränderungen einzelner Distraktoren vorgenommen oder Items vollständig eliminiert wurden, wurde keine neue Kontrollstichprobe erhoben. Stattdessen wurde für die ursprüngliche Analysestichprobe eine neue Punktbewertung der Version I ohne die eliminierten Aufgaben vorgenommen. Diese Ergebnisse sind im Folgenden für jede Testperson und für jeden Subtestteil aufgeführt:

Tab. 34-B 3.5.3: Testergebnisse Version I-6									
NR	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
1	13	14	8	5	4	5	9	7	4
2	13	15	12	8	2	3	7	13	11
3	14	16	13	9	2	3	7	8	6
4	14	13	12	10	5	9	12	10	8
5	10	16	9	8	2	2	4	6	8
6	11	17	8	4	3	5	7	4	4
7	16	17	13	8	5	7	10	10	9
8	16	13	16	18	16	18	16	15	15
9	17	18	15	16	16	17	17	15	17
10	17	14	14	17	13	10	14	15	8
11	17	17	17	17	14	17	16	15	16
12	17	17	15	16	14	15	15	13	11
13	17	18	18	18	16	17	13	16	15
14	6	7	10	9	4	4	3	4	0
15	16	15	14	16	13	17	14	15	13
16	6	11	7	7	3	4			
17	15	15	16	11	11	16	15	11	12
18	11	13	10	5	4	6	9	10	7
19	9	8	9	8			13	13	10
20	12	11	8	8	7	8	10	8	8
21	14	18	12	12	9	11	13	11	10
22	11	17	9	13	6	6	7	3	7
23	13	16	10	14	11	4	13	11	9
24	17	14	12	14	11	11	16	11	14
25	18	15	16	16	16	17	18	17	17
26	14	15	14	11	10	6	15	12	7
27	10	15	14	8	3	5	2	8	6
28	12	16	10	7	8	6			
29	11	16	7	12	10	8			
30	9	15	10	6	9	5	7	6	4
31	12	13	7	4	5	4	3	6	3
32	11	17	11	8	11	8	10	12	6
33	7	12	8	4	1	4	4	1	5
34	9	10	11	12	2	7			

NR	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
35	17	15	12	15	12	12	15	14	15
36	14	15	14	10	5	8	15	14	16
37	15	12	10	7			15	10	9
38	11	12	10	11	10	8	13	7	8
39	13	15	12	14	11	15	12	15	13
40	14	14	12	15	14	16	14	13	13
41	14	17	15	18	14	17	13	14	16
42	18	17	17	18	17	17	16	15	17
43	18	18	16	18	15	15	18	16	17
44	13	15	11	9	4	6	12	12	8
45	9	13	13	11	7	6	8	8	6
46	12	17	11	10	8	5	11	12	10
47	11	11	10	13	11	8	10	11	11
48	11	12	10	6	6	10	11	8	6
49	10	16	8	6	0	7	7	7	9
50	10	16	11	9	11	10	13	11	10
51	11	11	9	12	7	4	10	7	6
52	15	16	9	15	12	12	14	11	6
53	14	15	14	11	11	11	12	12	13
54	9	16	10	13	6	10	5	9	10
55	9	14	10				7	8	4
56	11	15	12	6	6	6	8	7	3
57	12	17	7	9	9	9	10	12	6
58	17	15	14	14	13	12	14	11	10
59	11	17	12	16	15	11	14	12	10
60	15	16	11	15	14	11	13	16	8
61	15	14	13	14	11	8	13	14	8
62	9	17	13	7	4	7	4	9	5
63	15	15	8	11	7	7	10	7	2
64	15	16	14	15	13	15	12	9	14
65	10	16	13	10	8	9	14	11	12
66	10	16	12				11	9	5
67	17	15	15	15	13	14	14	11	10
68	16	18	16	18	16	17	15	16	16
69	10	16	13	14	16	14	14	16	12
70	14	15	14	17	15	17	12	16	16
71	13	17	11	12	5	3	10	12	7
72	16	16	15	18	17	17	12	14	15
73	18	17	16	17	16	16	18	16	16
74	12	16	13	11	6	7	12	9	8
75	15	17	11	9	5	7	11	9	8
76	11	18	9	8	5	7	7	10	6
77	12	17	11	11	7	5	16	16	13
78	12	18	11	7	9	6	6	7	5
79	14	16	9	7	8	8	13	9	14
80	17	17	16	18	16	17	17	18	16

NR	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
81	13	17	15	13	12	13	11	10	10
82	17	18	14	17	15	15	16	16	15
83	13	16	11	11	13	13	14	13	7
84	15	17	16	16	12	12	13	14	13
85	15	16	16	16	13	11	13	11	10
86	14	16	15	14	14	12	16	15	13
87	12	14	7	6	6	6			
88	11	14	12	5	6	3	12	11	8
89	17	18	16	15	14	12	15	15	15
90	14	17	16	18	13	14	14	12	13
91	11	16	14	10	7	10	12	7	5
92	6	9	12	4	4	6	6	10	3
93	13	14	10	9	6	8	5	6	8
94	14	14	11	12	12	11	15	14	11
95	17	16	15	18	15	18	15	12	16
96	11	15	11	9	2	5	14	14	13
97	10	11	7	6	6	2	8	6	3
98	16	17	18	17	15	16	14	16	17
99	9	18	12	8	9	6	8	8	9
100	12	15	16	15	12	12	14	15	16
101				11	10	9			
102	14	14	13	16	13	9	14	16	11
103	12	15	14	12	12	10	16	10	14
104				17	16	17	18	17	18
105	16	17	17	17	18	17	18	17	18
106	15	17	17	17	16	18	17	18	14
107	16	14	15	18	15	16	17	14	17
108	13	15	17	17	12	18	17	15	17
109	15	17	16	17	13	15	16	16	17
110	13	17	15	13	9	12	12	15	16
111	14	16	13	9	10	5	14	14	14
112	14	17	12	11	12	11	14	14	14

Anh. 1: B 3.5.4a:

Jede Aufgabe erhält spaltenweise drei neue Plazierungen für ihren Schwierigkeitsgrad in Gebärdens-, Schrift- und Lautsprache. Die drei Wortarten werden dabei getrennt in jeweils allen Sprachmodalitäten einander gegenüber gestellt. Bei Aufgaben, die gemäß der Untersuchung innerhalb eines Subtestteils gleiche Schwierigkeitswerte aufweisen, werden zur Differenzierung ihre ursprünglichen Rangplätze in der Wortliste herangezogen und zur Unterscheidung verwendet:

Tab. 35a-B 3.5.4: neue Platzverteilung aufgrund der Schwierigkeit Nomen						
alte Nr.	GSN		SSN		LSN	
	p	neue Nr.	p	neue Nr.	p	neue Nr.
1	.62	13	.46	16	.51	16
4	.74	8	.71	6	.47	17
5	.49	17	.63	11	.60	11
6	.88	4	.84	3	.85	2
8	.82	5	.60	12	.72	5
9	.46	18	.46	17	.55	14
10	.64	12	.59	13	.71	6
11	.60	14	.94	1	.70	7
12	.95	2	.49	15	.70	8
13	.81	6	.74	4	.80	3
14	.80	7	.39	18	.54	15
15	.97	1	.94	2	.86	1
16	.93	3	.70	7	.41	18
17	.55	16	.74	5	.66	9
19	.55	15	.65	8	.58	13
20	.68	11	.65	9	.60	12
23	.70	9	.63	10	.66	10
24	.70	10	.59	14	.76	4

Tab. 35b-B 3.5.4: neue Platzverteilung aufgrund der Schwierigkeit Verben						
alte Nr.	GSV		SSV		LSV	
	p	neue Nr.	p	neue Nr.	p	neue Nr.
2	.88	11	.67	6	.46	15
5	.97	1	.71	3	.79	3
6	.67	16	.27	16	.41	16
7	.95	3	.63	8	.76	4
8	.75	14	.71	4	.75	5
9	.45	17	.22	17	.29	18
10	.36	18	.21	18	.38	17
12	.73	15	.55	10	.55	12
13	.86	12	.55	11	.67	10
14	.95	4	.72	2	.74	6
15	.95	5	.63	7	.70	8
16	.95	6	.69	5	.82	1
17	.94	7	.74	1	.74	7
18	.83	13	.40	14	.49	14
19	.96	2	.59	9	.70	9
20	.92	9	.46	12	.55	13
21	.92	10	.30	15	.63	11
23	.94	8	.430	13	.80	2

Tab. 35c-B 3.5.4: neue Platzverteilung aufgrund der Schwierigkeit Adjektive						
alte Nr.	GSA		SSA		LSA	
	p	neue Nr.	p	neue Nr.	p	neue Nr.
2	.88	6	.71	1	.51	12
4	.43	17	.52	11	.57	9
5	.45	15	.71	2	.56	10
6	.89	4	.63	4	.69	4
7	.54	13	.56	9	.64	5
8	.78	7	.63	5	.78	1
9	.89	5	.58	8	.58	7
10	.91	2	.61	7	.49	13
11	.17	18	.50	14	.41	16
13	.44	16	.46	16	.58	8
14	.78	8	.55	10	.63	6
15	.67	11	.52	12	.41	17
16	.94	1	.52	13	.76	2
19	.76	9	.65	3	.73	3
20	.58	12	.50	15	.40	18
22	.71	10	.21	18	.48	14
23	.90	3	.63	6	.53	11
24	.46	14	.28	17	.43	15

Anh. 1: B 3.5.4b:

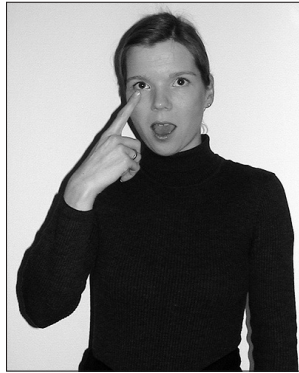
Tab. 36-B 3.5.4: Mittelwert der Schwierigkeitsrangfolge für Testheft I								
Nomen			Verben			Adjektive		
alte Nr.	Mittelwert PGS, PLS	neue Nr.	alte Nr.	Mittelwert PGS, PLS	neue Nr.	alte Nr.	Mittelwert PGS, PLS	neue Nr.
1	.56	16	2	.67	13	2	.69	9
4	.61	13	5	.88	2	4	.50	15
5	.55	17	6	.54	16	5	.51	14
6	.86	2	7	.85	4	6	.79	2
8	.77	5	8	.75	11	7	.59	11
9	.51	18	9	.37	18	8	.78	3
10	.68	7	10	.37	17	9	.74	5
11	.65	11	12	.64	15	10	.70	8
12	.82	3	13	.76	10	11	.29	18
13	.80	4	14	.84	5	13	.51	13
14	.67	10	15	.82	8	14	.71	7
15	.92	1	16	.88	1	15	.54	12
16	.67	9	17	.84	6	16	.85	1
17	.60	14	18	.66	14	19	.75	4
19	.57	15	19	.83	7	20	.49	16
20	.64	12	20	.73	12	22	.59	10
23	.68	8	21	.78	9	23	.72	6
24	.73	6	23	.87	3	24	.44	17

Anh. B 3.5.4:

Abb. 13a-B 3.5.4: Gebärdenfotos und Ausführungsanweisungen Nomen



Übungsbsp. 1: Hund
Ht schlägt 2x gegen
Oberschenkel: „bei Fuß“



Übungsbsp. 2: Auge
Z zeigt auf Auge



1. Ei
L-H deutet „Eiform“ an



2. Bett
<C-He im Bogen an Brust
„ziehen am Federbett“ bzw.
„ahmen Form nach“



3. Bart
5-H (Ht zum Gesicht) seit-
lich vom Ohr zum Kinn:
„Form des Barts“



4. Zahn
Z zeigt auf Zahn



5. Angst

A-H (Hr v) schlägt 2x an Brust: „Herzklopfen“



6. Welt

V-He (Z zueinander) beschreiben Kreis nach u (M aneinander): „Weltkugel“



7. Stein

r F „wie ein Stein“ im Bogen über l F



8. Rose

1. Fa „R“
2. O-H (Hr v) öffnet sich nach o zur 5-H (= „Blume“)



9. Zahl

F (hr v) öffnet Fi nacheinander von D bis K zu 5-H



10. Leute

X-H (Ht v) 3x versetzt nach u

11. Schuh

Fe (Hr o) mit halber Drehung zur Brust führen (Hr u): „Schuh über den Fuß streifen“



12. Arbeit

r F 2x kreisförmig auf l F



13. Kopf
Fi tippen 2x an Kopft



14. Geld
D und Z reiben aneinander



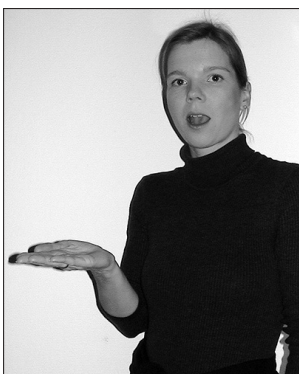
15. Holz
r Hk „hackt“ auf Z der l *B-H



16. Mutter
F (Hr außen) streicht zweimal
an Wange entlang



17. Straße
*B-He (Hr. außen) im Bogen
nach v: „Verlauf der Straße“



18. Garten
B-H (Ht o) dreht 2x im H-gelenk nach v (Hr o): „umgraben“



Abb. 13b-B 3.5.4: Gebärdenfotos und Ausführungsanweisungen Verben



1. schreiben
F-H in „Wellenlinienform“
nach r



2. fliegen
B-He ahmen 2x „Fliegen des
Vogels“ nach



3. stinken
Dund Z der Q-H „halten Nase
zu“



4. springen
Zund M der V-H (Hr o)
„springen“ vom H-ballen der l
B-H (Ht o) über Ht nach v



5. weinen
Z (Hr v) 2x an Wange nach u
deuten „weinen“ an



6. lesen
B-He 2x hin und her: „Zeilen
lesen“



7. waschen

5-He deuten mit kreisförmiger Bewegung „Gesicht waschen“ an



8. werfen

F (Ht v) im Bogen nach o v zur 5-H öffnen: „etwas werfen“



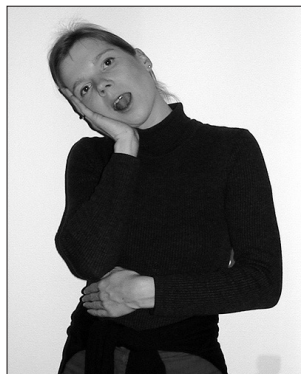
9. schieben

B-He (Ht v) „schieben“ nach v



10. verstecken

„Kopf versteckt sich“ hinter B-He (r Hr r, l Ht r)



11. schlafen

B-H an Wange



12. brennen

5-He (Hr außen) mit Fi-spiel nach o: „Lodern der Flammen“



13. gehen

Z und M der V-H (Hr o) ahmen „Gehbewegung“ nach



14. singen

Z in Bögen hin und her: „dirigieren“



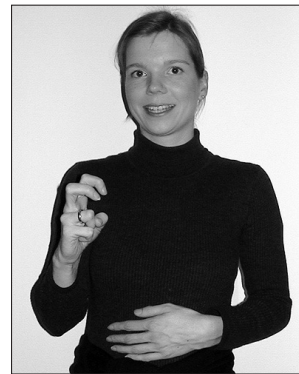
15. warten

5-He (Hr o) 2x nach u



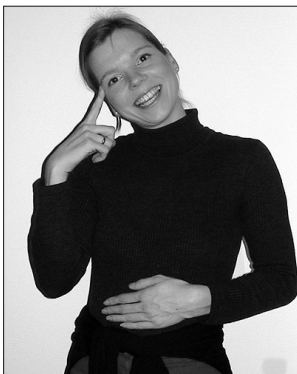
16. gucken

V-H (Hr v) von Augen nach v:
„gucken mit den Augen“



17. hinsetzen

V-H (Hr außen) im Bogen von l o nach r u, H wippt unten nach



18. denken

Z an Schläfe

Abb. 13c-B 3.5.4: Gebärdenfotos und Ausführungsanweisungen Adjektive



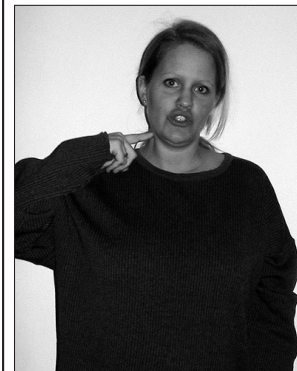
1. glatt
Ht der r B-H „rutscht schnell“
über Ht der l B-H nach v



2. spitz
r Q-H an l Z entlang nach o
zur <Q-H: Z bekommt
„Spitze“



3. hellblau
1. 5-H (Hr v) im Viertelkreis nach o: „es wird hell“
2. U-H (Ht v) Schlangenlinie von o nach u



4. schmutzig
G-H mit Z-Spitze r am Hals,
mit H-gelenkdrehung Hr von
h nach v



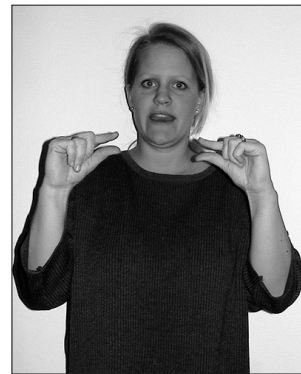
4. orange

5-H (Ht o) dreht im H-gelenk hin und her „Orange in der H“



5. viereckig

L-He(Ht v) bewegen sich von D an D und Z an Z nach r und l auseinander, schließen sich dann zu <Q-He: deuten Form des „Vierecks“ an



6. warm

5-H (Hr v) schließt am Mund vorbei nach r zur F



7. dünn

B-He (Hr außen, geringer Abstand) bewegen sich von o nach u am Körper entlang



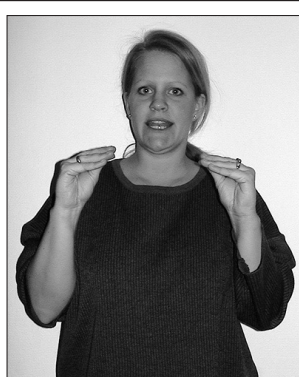
9. klein

L-H „verkleinert sich“ zur Q-H (Hk v)



10. gestreift

5-H (Ht o) bewegt sich von o nach u: „deutet Streifen an“



11. nass

<*B-He (Hk außen) schließen 2x zu <O-He



12. schief

*B-He (Ht außen) schräg nach r o: „schiefe Ebene“



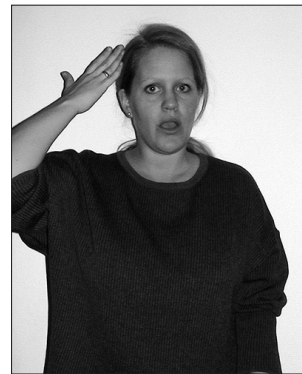
13. verletzt

M der r U-H (Hr v) „schneidet“ Z der l U-H (Hr l) nach r



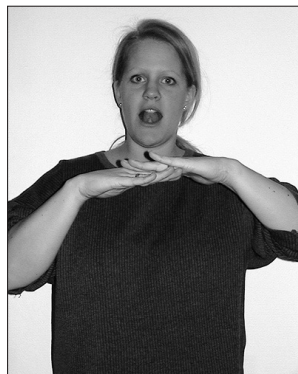
14. blond

5-H (Ht v) mit H-gelenkdrehung schnell nach o gegen Haare (Hr v)



15. voll

B-He (Hr o) deuten Höhe eines Glases an, u Hr schnellst gegen o Ht



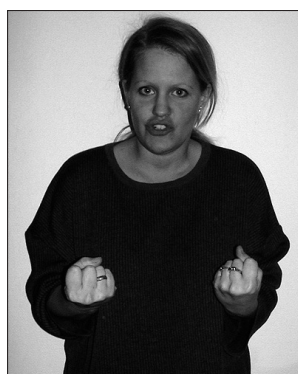
16. weich

Fi der C-He nähern sich 2x den D: „etwas Weiches zusammendrücken“



17. nah

r *B-H (Hr v) nähert sich l
*B-H



18. schwer

Fe (Ht o) gleichzeitig langsam nach o

Anh. 1: B 3.5.6a:

Im Folgenden werden die neuen Schwierigkeits- und Trennschärfewerte des Itemkorpus von Version II dargestellt, wobei die Items durchgehend nach der Reihenfolge aus Testheft I (eigentlich für Gebärden- und Lautsprache, in diesem Fall auch für Schriftsprache) sortiert sind:

Tab. 40a-B 3.5.6: Schwierigkeit und Trennschärfe Nomen (Testversion II)						
Nr.	GSN		SSN		LSN	
	p	r_{it}	p	r_{it}	p	r_{it}
1. Ei	.99	.17	.96	.20	.89	.10
2. Bett	.89	.25	.86	.45	.88	.39
3. Bart	.96	.25	.50	.45	.72	.33
4. Zahn	.83	.10	.76	.33	.82	.28
5. Angst	.84	.23	.61	.51	.76	.64
6. Welt	.71	.29	.60	.55	.78	.39
7. Stein	.66	.40	.60	.59	.74	.47
8. Rose	.71	.18	.65	.42	.68	.30
9. Zahl	.95	.10	.71	.42	.42	.34
10. Leute	.81	.31	.40	.26	.56	.49
11. Schuh	.61	.32	.96	.27	.72	.44
12. Arbeit	.69	.30	.66	.50	.62	.46
13. Kopf	.75	.04	.72	.39	.48	.34
14. Geld	.56	.07	.76	.50	.68	.31
15. Holz	.56	.52	.66	.58	.60	.14
16. Mutter	.63	.35	.47	.23	.52	.33
17. Straße	.50	.16	.64	.56	.61	.52
18. Garten	.47	.41	.46	.57	.59	.46

Tab. 40b-B 3.5.6: Schwierigkeit und Trennschärfe Verben (Testversion II)

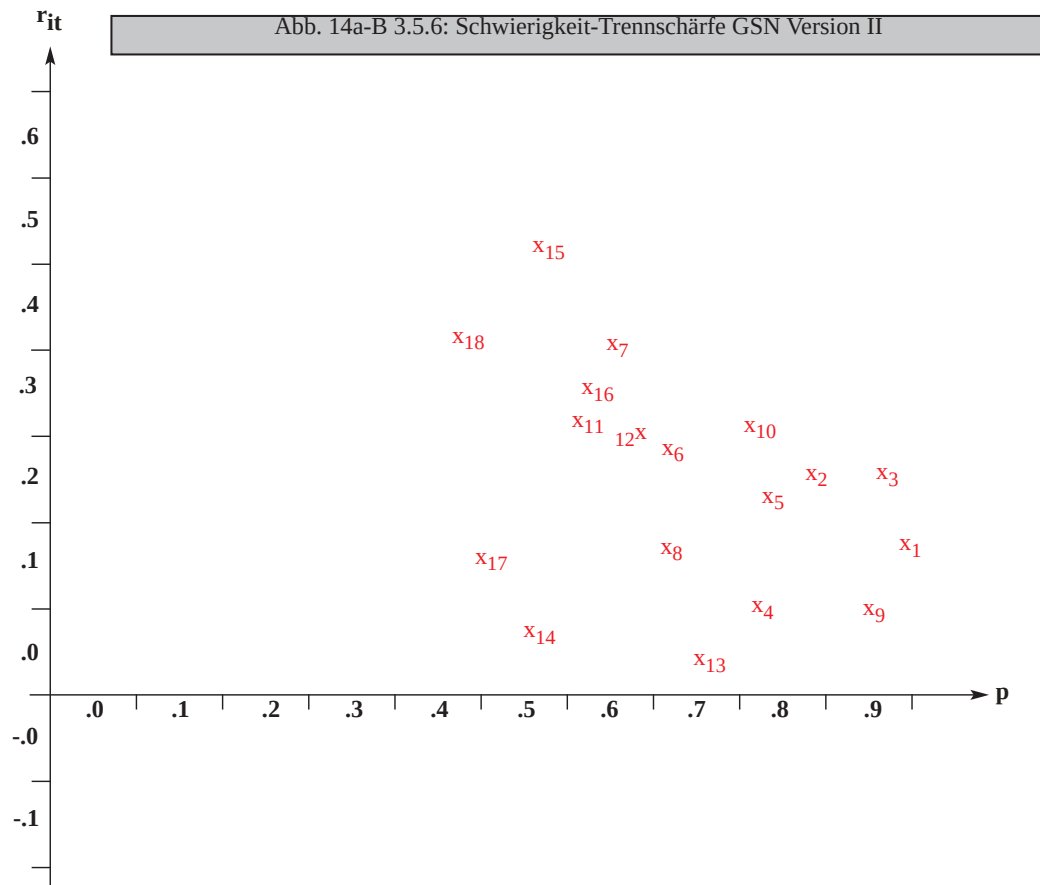
Nr.	GSV		SSV		LSV	
	p	r _{it}	p	r _{it}	p	r _{it}
1. schreiben	.96	.30	.71	.61	.85	.32
2. fliegen	.99	.28	.74	.57	.82	.35
3. stinken	.96	.19	.44	.40	.82	.33
4. springen	.96	.07	.65	.62	.78	.37
5. weinen	.96	.09	.75	.52	.77	.30
6. lesen	.96	.23	.77	.61	.77	.50
7. waschen	.97	.10	.61	.63	.72	.54
8. werfen	.96	.37	.66	.40	.72	.29
9. schieben	.94	.26	.31	.25	.65	.38
10. verstecken	.87	.23	.57	.56	.69	.41
11. schlafen	.76	.21	.74	.52	.77	.31
12. brennen	.94	.23	.48	.47	.56	.41
13. gehen	.90	.53	.69	.53	.47	.21
14. singen	.85	.36	.42	.62	.50	.42
15. warten	.75	.45	.57	.43	.58	.32
16. gucken	.68	.40	.28	.15	.42	.27
17. hinsetzen	.36	.17	.21	.36	.40	.28
18. denken	.46	.37	.23	.15	.29	.24

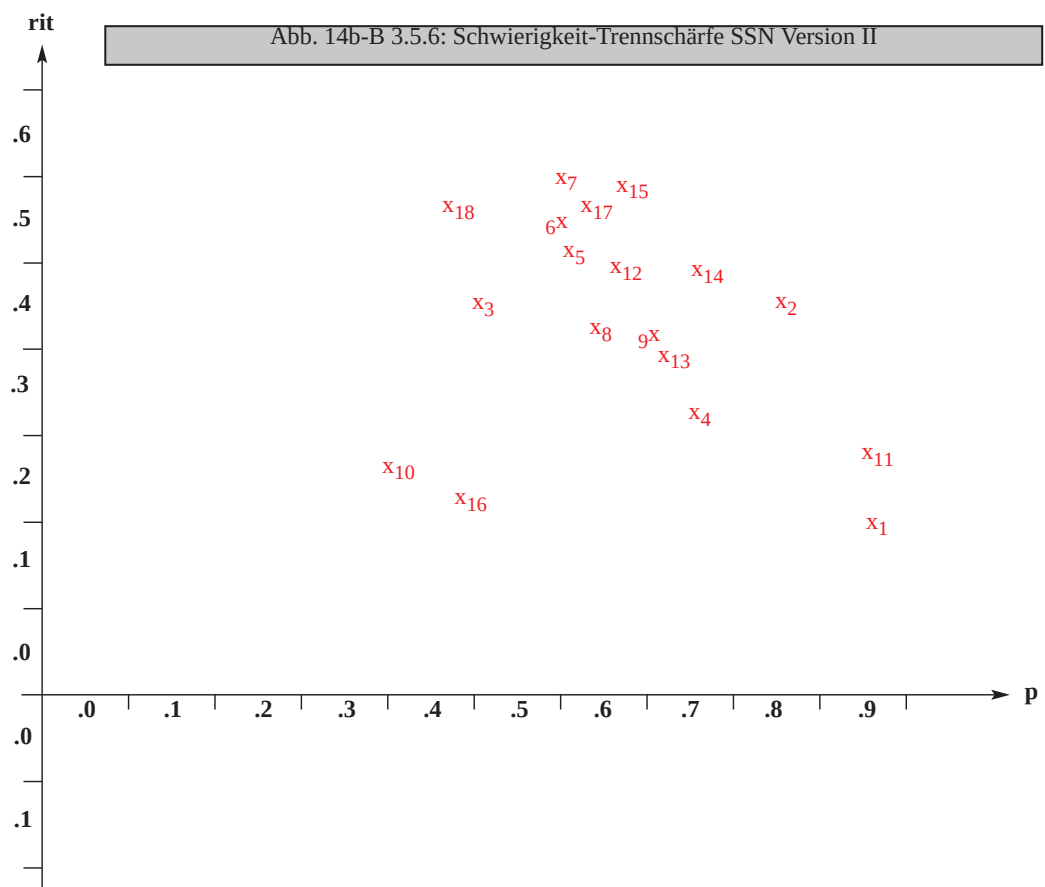
Tab. 40c-B 3.5.6: Schwierigkeit und Trennschärfe Adjektive (Testversion II)

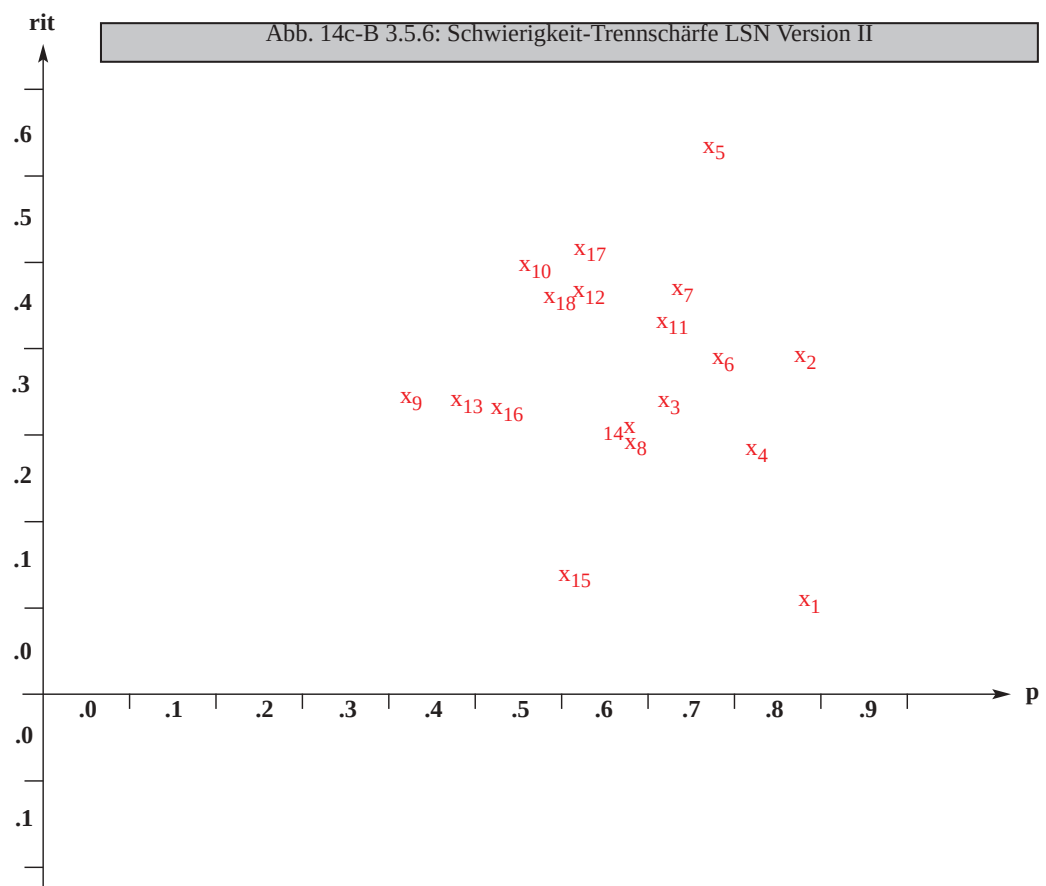
Nr.	GSA		SSA		LSA	
	p	r _{it}	p	r _{it}	p	r _{it}
1. glatt	.96	.17	.54	.37	.79	.35
2. spitz	.91	.18	.66	.47	.71	.39
3. hellblau	.79	.34	.66	.55	.81	.34
4. schmutzig	.77	.32	.68	.43	.76	.49
5. orange	.91	.14	.60	.28	.60	.53
6. viereckig	.92	.10	.65	.42	.54	.44
7. warm	.79	.29	.57	.54	.66	.53
8. dünn	.93	.02	.63	.48	.51	.46
9. klein	.89	.18	.73	.39	.52	.30
10. gestreift	.72	.19	.22	.29	.50	.54
11. nass	.55	.38	.58	.54	.66	.31
12. schief	.68	.15	.54	.42	.43	.52
13. verletzt	.45	.40	.48	.47	.59	.37
14. blond	.46	.46	.73	.50	.58	.47
15. voll	.44	.43	.54	.59	.59	.42
16. weich	.59	.36	.52	.56	.41	.58
17. nah	.46	.21	.29	.36	.43	.31
18. schwer	.17	.25	.52	.63	.42	.44

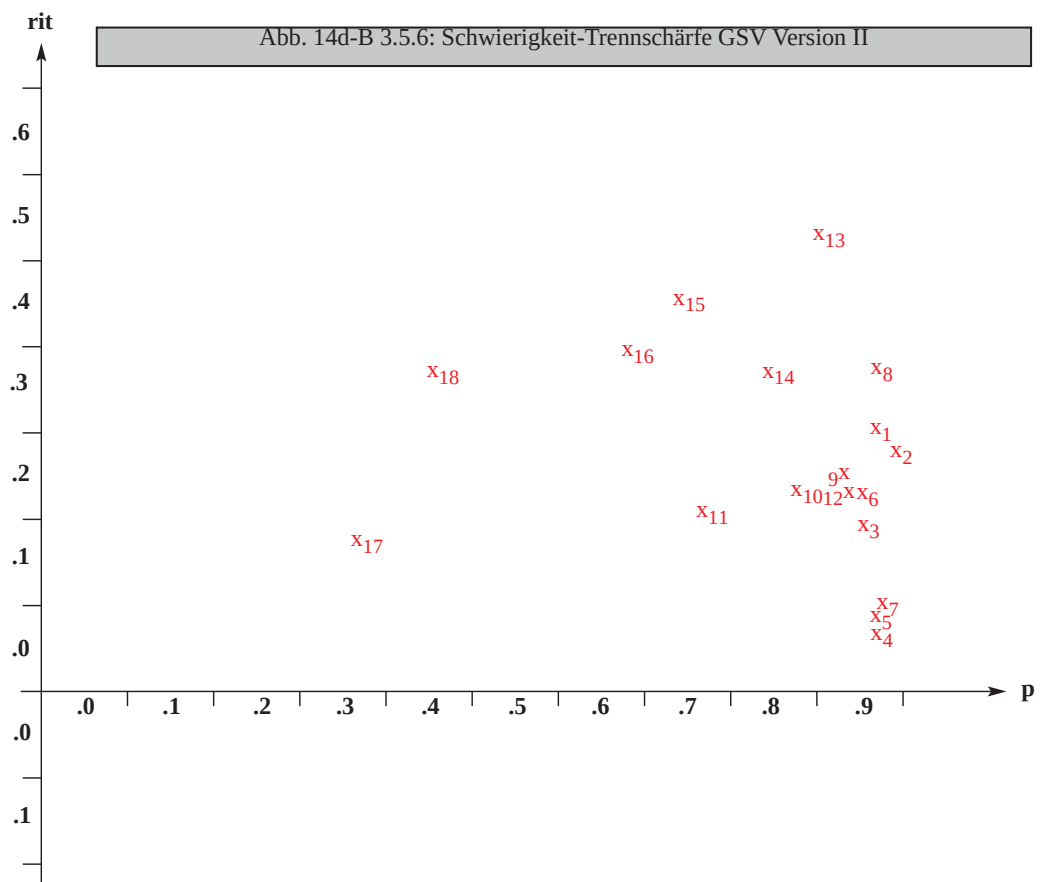
Anh. 1: B 3.5.6b:

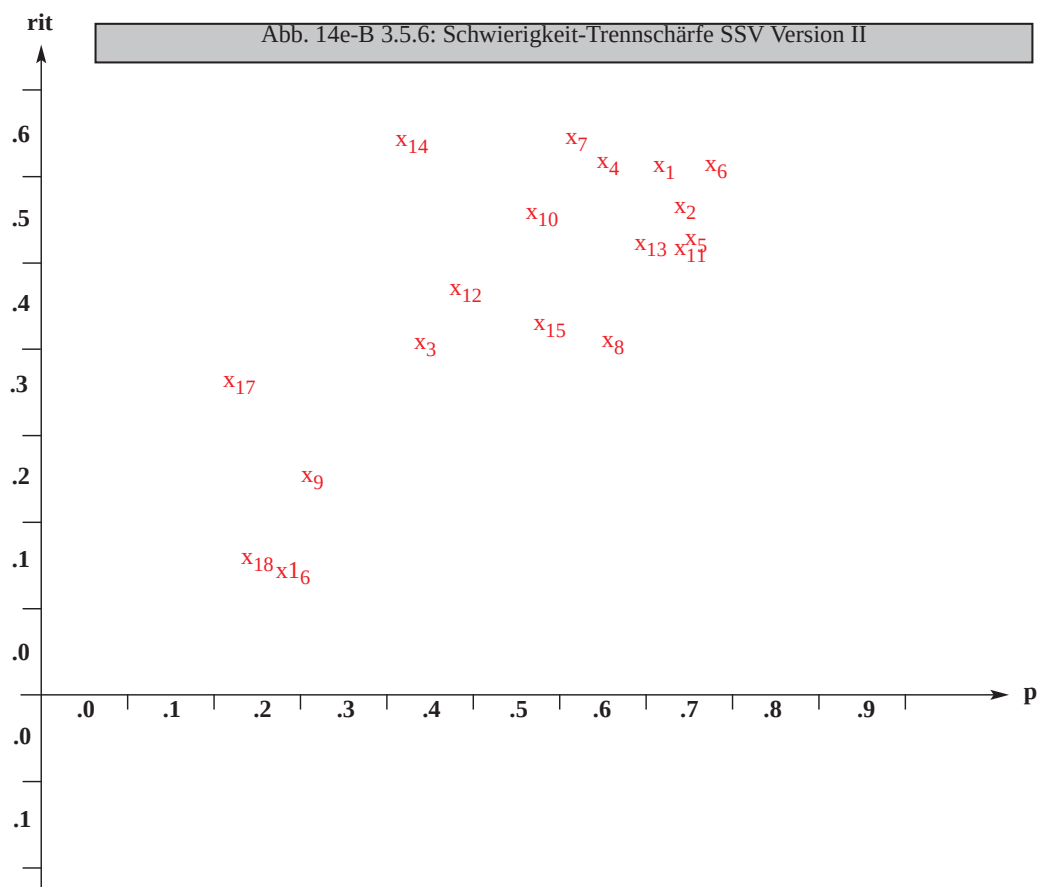
Die folgenden Koordinatensysteme verdeutlichen anschaulich die Beziehungen der Trennschärfe- und Schwierigkeitsindizes des neuen Itemkorpus Version II mit 18 Items in jedem Subtestteil. Die Items sind durchgehend nach der Reihenfolge aus Testheft I (eigentlich für Gebärden- und Lautsprache, in diesem Fall auch für Schriftsprache) sortiert. Die Schwierigkeit ist auf der Abzisse, die Trennschärfe auf der Ordinate abgetragen:

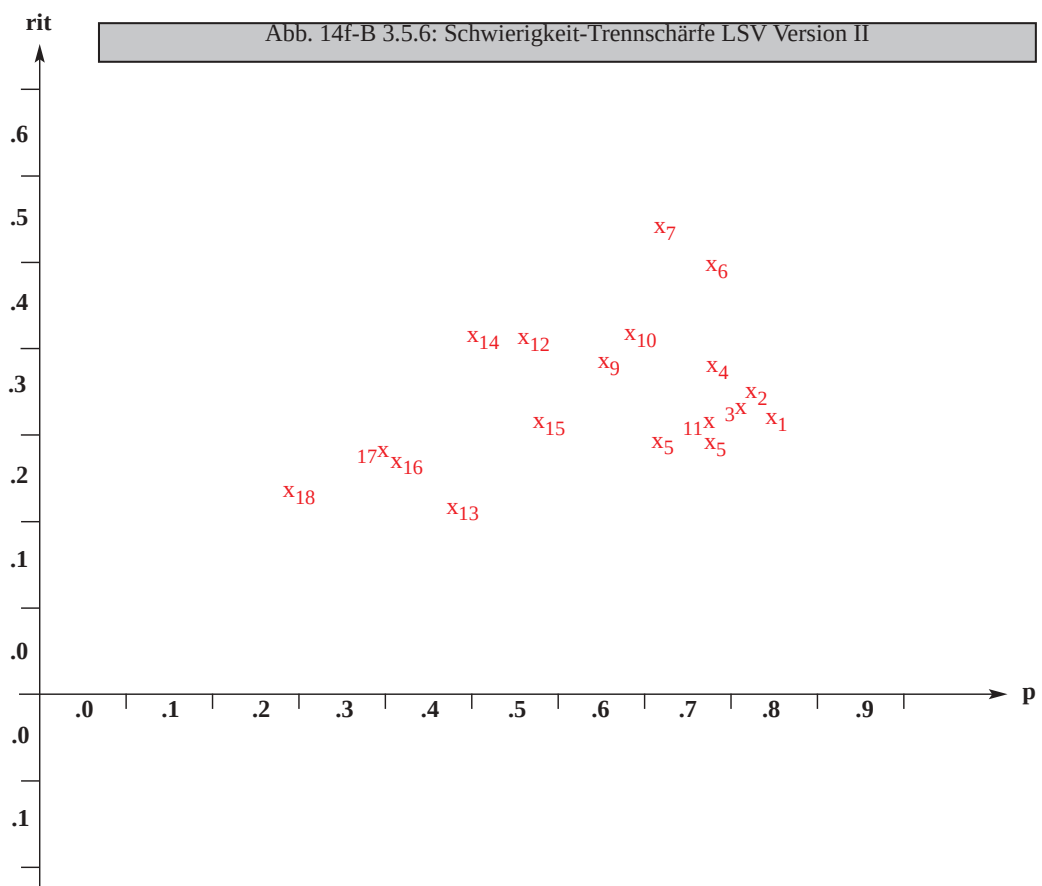


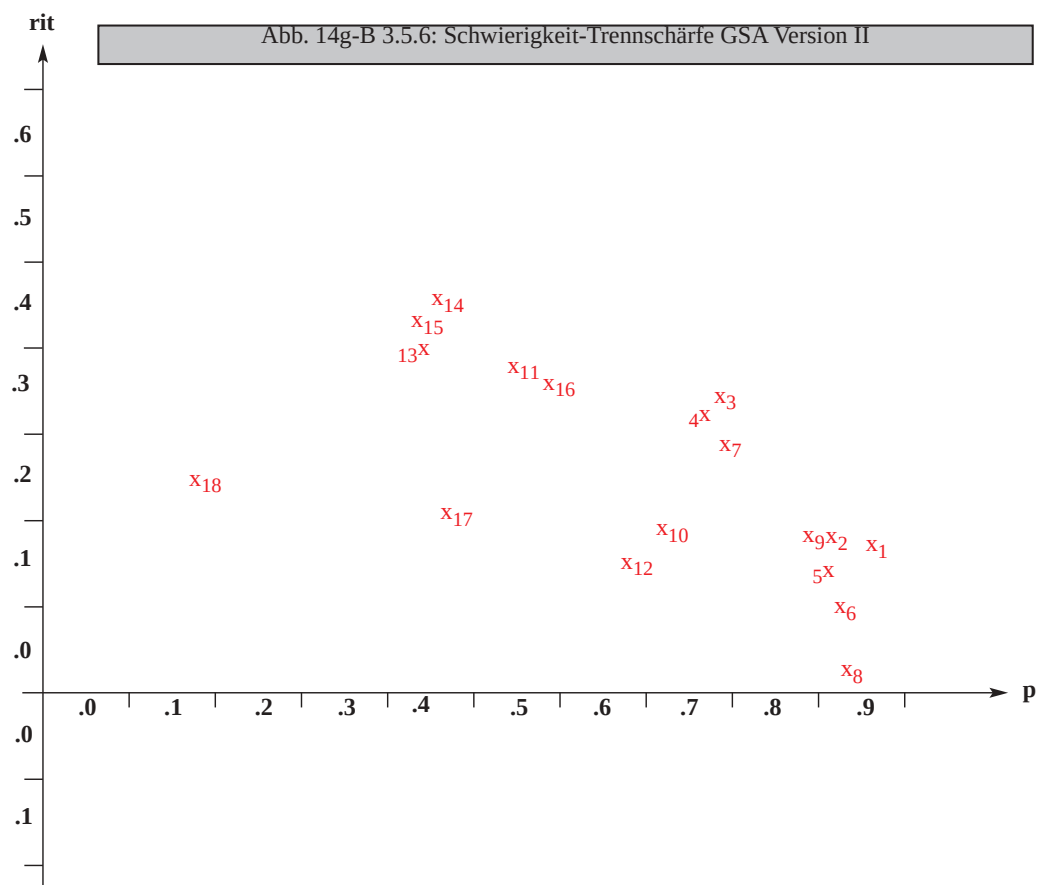


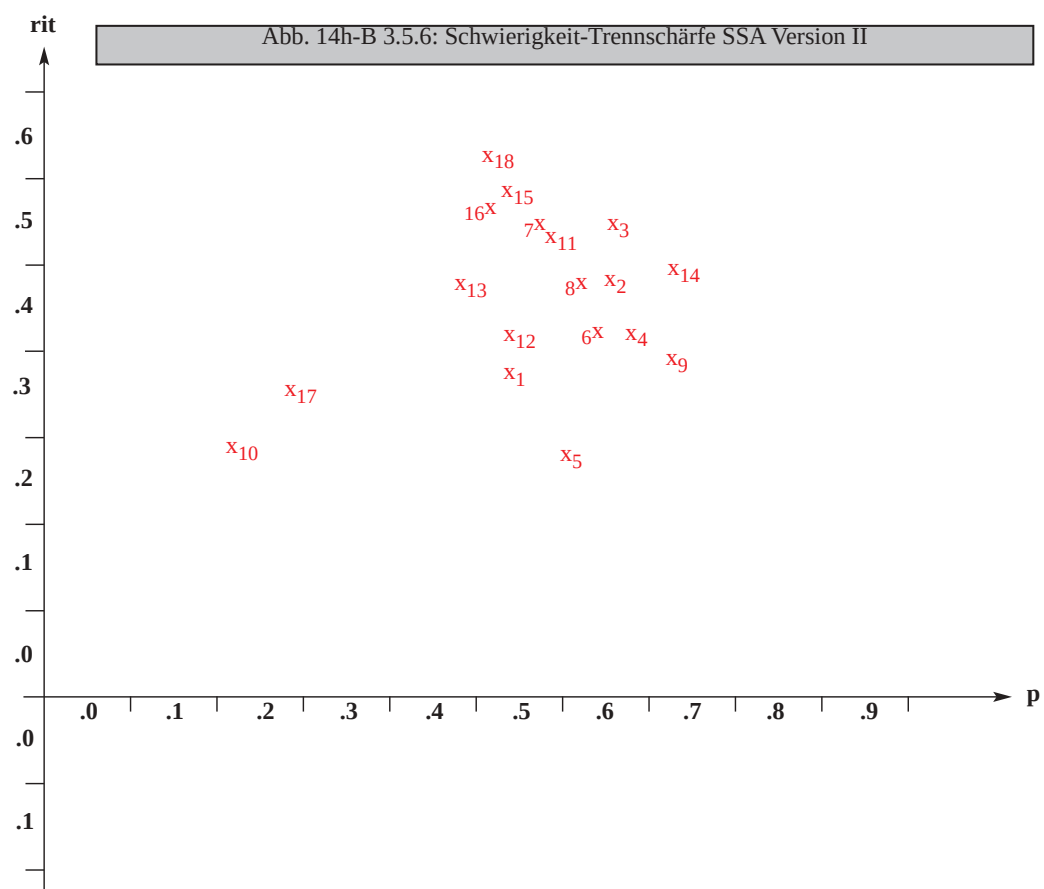


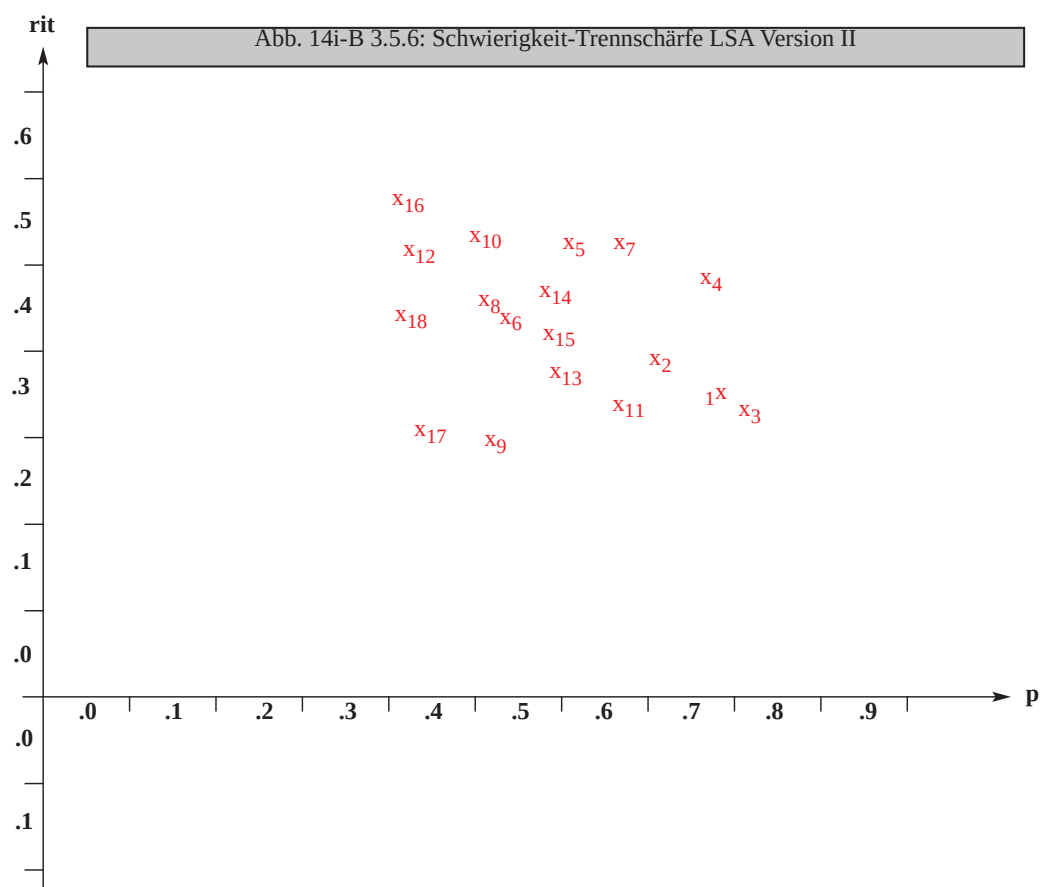












Tab. 42a-B 3.5.7: Modifiziertes Testprotokoll Deckblatt

PERLESKO

Prüfverfahren zur Erfassung lexikalisch-semantischer Kompetenz

Angaben zur Testperson

Name:

Geburtstag:

Alter:

Klasse:

Subtest Gebärdensprache

Datum:

Uhrzeit:

Subtest Schriftsprache

Datum:

Uhrzeit:

Subtest Lautsprache

Datum:

Uhrzeit:

Tab. 42b-B 3.5.7: Modifiziertes Testprotokoll GSN

PERLESKO - SUBTESTEIL GEBARDENSPRACHE NOMEN																	
Item	Distraktor						Bemerkungen	gewählter Distraktor									
	I	II	III	IV	V	VI		HRP	G	M	I						
U1 Hund	Katze I	Pferd G/I	Hahn M/I														
U2 Auge	Nase G/I	Stirn HRP/G/I	Auto M														
1 Ei	Eis M	Huhn HRP/I	Kartoffel G/I														
2 Bett	Sofa I	Decke HRP/G/I	Brett M														
3 Bart	Haare HRP/I	Post G/M	Kinn HRP/G/I														
4 Zahn	Zange M/I	Hahn M	Mund HRP/G/I														
5 Angst	Freude I	Arger G/M/I	Trauer G/I														
6 Welt	Planet HRP/G/I	Ball G	Vase G/M														
7 Stein	Kaffee G	Stange M	Felsen HRP/I														
8 Rose	Sonnenblume G/I	Tulpe G/I	Rasen M/I														
9 Zahl	Regen G	Pfahl M	Buchstabe I														
10 Leute	Herde M/I	Mensch G/I	Paar HRP/I														
11 Schuh	Uhr M	Strumpf G/I	Stiefel HRP/G/I														
12 Arbeit	Eimer HRP/M/I	Äpfel M	Stein G														
13 Kopf	Topf M	Junge G	Zopf HRP/M/I														
14 Geld	Salz G	Gold HRP/M/I	Feld M														
15 Holz	Baum HRP/I	Brot G/M	Gold M														
16 Mutter	Motte M	Bauer G	Mädchen HRP/I														
17 Straße	Strauß M	Weg HRP/G/I	Strand M														
18 Garten	Arbeit M	Baum HRP/I	Geschenk G														
Σ																	

Tab. 42c-B 3.5.7: Modifiziertes Testprotokoll GSV

PERLESKO - SUBTEST GERÄDENSPRACHE VERBEN																	
Item	Distraktor						Bemerkungen	gewählter Distraktor									
	I	II	III	IV	V	VI		HRP	G	M	I	HRP	G	M	I	HRP	G
1 schreiben	reiben G/M	schlafen M	malen G/I														
2 fliegen	fragen M	springen HRP/I	winken G														
3 stinken	staunen M	riechen HRP/G/I	winken M														
4 springen	klettern I	schimpfen M	sitzen G														
5 weinen	schwitzen I	reiben HRP/G/I	wiegen M														
6 lesen	sitzen HRP/I	halten G	blasen M														
7 waschen	wischen HRP/G/M/I	warten M	schwimmen G														
8 werfen	fangen G/I	wärmen M	strecken I														
9 schieben	ziehen I	schließen HRP/G/M/I	heben I														
10 verstecken	erschrecken M/I	schlafen G	stecken M														
11 schlafen	liegen I	schlagen G/M	streicheln G														
12 brennen	rennen M	löschen HRP/I	rechnen G/M														
13 gehen	stehen G/M/I	sehen M	rennen I														
14 singen	warnen G	sagen G/M/I	flöten I														
15 warten	braten M	wandern G/M	sitzen HRP/I														
16 gucken	schminken HRP/G/I	ducken M	denken G														
17 hinsetzen	sitzen HRP/G/M/I	hinlegen M/I	schwitzen M														
18 denken	warnen G	träumen HRP/G/I	schenken M														
Σ																	

Tab. 42d-B 3.5.7: Modifiziertes Testprotokoll GSA

PERLESKO - SUBTESTTEIL GEBÄRDENSPRACHE ADJEKTIVE																	
Item	Distraktor						Bemerkungen	gewählter Distraktor									
	I	II	III	IV	V	VI		HRP	G	M	I						
1 glatt	lockig 	platt M	sauber G														
2 spitz	schief 	struppig G	spannend M														
3 hellblau	hellgrau G/M/I	hellgrün G/M/I	dunkelblau G/M/I														
4 schmutzig	zerrissen 	sauber 	zerknittert 														
5 orange	rosa G/M/I	rot 	gelb 														
6 viereckig	sechseckig G/M/I	dreieckig G/M/I	fünfeckig G/M/I														
7 warm	heiß HRP/G/I	kalt 	weich M														
8 dünn	dumm M	eng HRP/G/M/I	dick 														
9 klein	fein M	dünn 	klug G/M														
10 gestreift	geblümt 	kariert G/I	gepunktet 														
11 nass	blass M	trocken G/I	nackt M														
12 schief	gerade 	steil HRP/G/I	stark M														
13 verletzt	gesund 	scharf HRP/G/I	verliebt M/S														
14 blond	braun M/I	rot 	schwarz 														
15 voll	viel HRP/G/M/I	groß G/M/I	stolz G/M														
16 weich	nass G	wach M/I	warm M														
17 nah	nass M	geschlossen G	fern 														
18 schwer	stark HRP/G/I	leer M	schwarz M														
Σ																	

Tab. 42e-B 3.5.7: Modifiziertes Testprotokoll SSN

PERLESKO - SUBTESTEIL SCHRIFTSPRACHE NOMEN																	
Item	Distraktor						Bemerkungen	gewählter Distraktor									
	I	II	III	IV	V	VI		HRP	G	M	I	S					
U1 Hund	Katze I	Pferd G/I	Hahn M/I/S														
U2 Auge	Nase G/I	Auto M/S	Stirn HRP/G/I														
1 Schuh	Stiefel G/I	Strumpf G/I	Uhr M														
2 Ei	Eis M/S	Kartoffel G/I	Huhn I														
3 Bett	Brett M/S	Sofa I	Decke HRP/G/I														
4 Zahn	Mund G/I	Hahn M/S	Zange M/I/S														
5 Geld	Gold M/I/S	Salz G	Feld M/S														
6 Kopf	Zopf M/I/S	Junge G	Topf M/S														
7 Zahl	Buchstabe I	Regen G	Pfahl M/S														
8 Holz	Baum HRP/I	Brot G/M	Gold M														
9 Arbeit	Stein G	Eimer M/I	Ampel M														
10 Rose	Sonnenblume G/I	Tulpe G/I	Rasen M/I/S														
11 Straße	Strauß M/S	Strand M/S	Weg HRP/G/I														
12 Angst	Trauer G/I	Freude I	Arger G/M/I														
13 Stein	Felsen I	Stange M/S	Kaffee G														
14 Welt	Planet HRP/G/I	Ball G	Vase G/M														
15 Bart	Haare HRP/I	Kinn HRP/G/I	Post G/M														
16 Mutter	Bauer G	Motte M/S	Mädchen HRP/I														
17 Garten	Baum HRP/I	Geschenk G	Arbeit M														
18 Leute	Herde M/I	Mensch G/I	Paar I														
Σ																	

Tab. 42f-B 3.5.7: Modifiziertes Testprotokoll SSV

PERLESKO - SUBTESTEIL SCHRIFTSPRACHE VERBEN

Item	Distraktor							Bemerkungen	gewählter Distraktor											
	I		II		III				HRP	G	M	I	S							
1	lesen	sitzen I		halten HRP/G/I			blasen M/S													
2	weinen	reiben G/I/S		schwitzen I			wiegen M/S													
3	fliegen	springen HRP/I		fragen M/S			winken G													
4	schlafen	schlagen G/M/S		streicheln G			liegen HRP/I													
5	schreiben	reiben G/M/S		schlafen M/S			malen G/I													
6	gehen	rennen I		stehen G/M/I/S			sehen M/S													
7	werfen	fangen G/I		wärmen M/S			strecken I													
8	springen	sitzen G		klettern I			schimpfen M													
9	waschen	schwimmen G		wischen HRP/G/M/I/S			warten M/S													
10	warten	wandern G/M/S		braten M/S			sitzen I													
11	verstecken	stecken M/S		erschrecken M/I/S			schlafen G													
12	brennen	löschen HRP/I		rennen M/S			rechnen G/M/S													
13	stinken	riechen HRP/G/I		winken M/S			staunen M/S													
14	singen	flöten I		sagen G/M/I/S			warnen G													
15	schieben	ziehen I		schließen G/M/I/S			heben I													
16	gucken	denken G		ducken M/S			schminken G/I													
17	denken	träumen HRP/G/I		warnen G			schchenken M/S													
18	hinsetzen	sitzen HRP/G/M/I/S		schwitzen M			hinlegen M/I/S													
Σ																				

Tab. 42g-B 3.5.7: Modifiziertes Testprotokoll SSA

PERLESKO - SUBTESTTEIL SCHRIFTSPRACHE ADJEKTIVE															
Item		Distraktor					gewählter Distraktor								
		I	II		III	Bemerkungen									
1	klein	klug G/M	dünn I		fein M/S										
2	blond	braun M/I	schwarz I		rot I										
3	schmutzig	zerknittert I	zerrissen I		sauber I										
4	spitz	struppig G	spannend M		schief HRP/I/S										
5	hellblau	hellgrün G/M/I/S	hellgrau G/M/I/S		dunkelblau G/M/I/S										
6	viereckig	dreieckig G/M/I/S	sechseckig G/M/I/S		fünfeckig G/M/I/S										
7	dünn	dumm M/S	dick I		eng G/M/I										
8	orange	rot I	rosa G/M/I		gelb I										
9	nass	nackt M/S	blass M/S		trocken G/I										
10	warm	kalt I	weich M/S		heiß HRP/G/I										
11	voll	stolz G/M	viel HRP/G/M/I/S		groß G/M/I										
12	schief	gerade I	stark M		steil G/I										
13	glatt	lockig I	platt M/S		sauber G										
14	schwer	leer M	schwarz M/S		stark G/I										
15	weich	warm M	nass G		wach M/I/S										
16	verletzt	gesund I	scharf G/I		verleibt M/S										
17	nah	fern I	geschlossen G		nass M/S										
18	gestreift	gepunktet I	kariert G/I		geblümt I										
Σ															

Tab. 42h-B 3.5.7: Modifiziertes Testprotokoll LSN

PERLESKO - SUBTESTTEIL LAUTSPRACHE NOMEN																	
Item	Distraktor						Bemerkungen	gewählter Distraktor									
	I	II	III	IV	V	VI		HRP	G	M	I	HRP	G	M	I	HRP	G
Ü1 Hund	Katze I	Pferd G/I	Hahn M/I														
Ü2 Auge	Nase G/I	Stirn HRP/G/I	Auto M														
1 Ei	Eis M	Huhn HRP/I	Kartoffel G/I														
2 Bett	Sofa I	Decke HRP/G/I	Brett M														
3 Bart	Haare HRP/I	Post G/M	Kinn HRP/G/I														
4 Zahn	Zange M/I	Hahn M	Mund HRP/G/I														
5 Angst	Freude I	Ärger G/M/I	Trauer G/I														
6 Welt	Planet HRP/G/I	Ball G	Vase G/M														
7 Stein	Kaffee G	Stange M	Felsen HRP/I														
8 Rose	Sonnenblume G/I	Tulpe G/I	Rasen M/I														
9 Zahl	Regen G	Pfahl M	Buchstabe I														
10 Leute	Herde M/I	Mensch G/I	Paar HRP/I														
11 Schuh	Uhr M	Strumpf G/I	Stiefel HRP/G/I														
12 Arbeit	Eimer HRP/M/I	Ampel M	Stein G														
13 Kopf	Topf M	Junge G	Zopf HRP/M/I														
14 Geld	Salz G	Gold HRP/M/I	Feld M														
15 Holz	Baum HRP/I	Brot G/M	Gold M														
16 Mutter	Motte M	Bauer G	Mädchen HRP/I														
17 Straße	Strauß M	Weg HRP/G/I	Strand M														
18 Garten	Arbeit M	Baum HRP/I	Geschenk G														
Σ																	

Tab. 42i-B 3.5.7: Modifiziertes Testprotokoll LSV

PERLESKO - SUBTEST LAUTSPRACHE VERBEN																	
Item	Distraktor						Bemerkungen	gewählter Distraktor									
	I	II	III	IV	V	VI		HRP	G	M	I	HRP	G	M	I	HRP	G
1 schreiben	reiben G/M	schlafen M	malen G/I														
2 fliegen	fragen M	springen HRP/I	winken G														
3 stinken	staunen M	riechen HRP/G/I	winken M														
4 springen	klettern I	schimpfen M	sitzen G														
5 weinen	schwitzen I	reiben HRP/G/I	wiegen M														
6 lesen	sitzen HRP/I	halten G	blasen M														
7 waschen	wischen HRP/G/M/I	warten M	schwimmen G														
8 werfen	fangen G/I	wärmen M	strecken I														
9 schieben	ziehen I	schließen HRP/G/M/I	heben I														
10 verstecken	erschrecken M/I	schlafen G	stecken M														
11 schlafen	liegen I	schlagen G/M	streicheln G														
12 brennen	rennen M	löschen HRP/I	rechnen G/M														
13 gehen	stehen G/M/I	sehen M	rennen I														
14 singen	warnen G	sagen G/M/I	flöten I														
15 warten	braten M	wandern G/M	sitzen HRP/I														
16 gucken	schminken HRP/G/I	ducken M	denken G														
17 hinsetzen	sitzen HRP/G/M/I	hinlegen M/I	schwitzen M														
18 denken	warnen G	träumen HRP/G/I	schenken M														
Σ																	

Tab. 42j-B 3.5.7: Modifiziertes Testprotokoll LSA

PERLESKO - SUBTESTEIL LAUTSPRACHE ADJEKTIVE																	
Item	Distraktor						Bemerkungen	gewählter Distraktor									
	I	II	III	HRP	G	M		I									
1	glatt	lockig 		platt M	sauber G												
2	spitz	schief 		struppig G	spannend M												
3	hellblau	hellgrau G/M/I		hellgrün G/M/I	dunkelblau G/M/I												
4	schmutzig	zerrissen 		sauber 	zerknittert 												
5	orange	rosa G/M/I		rot 	gelb 												
6	viereckig	sechseckig G/M/I		dreieckig G/M/I	fünfeckig G/M/I												
7	warm	heiß HRP/G/I		kalt 	weich M												
8	dünn	dumm M		eng HRP/G/M/I	dick 												
9	klein	fein M		dünn 	klug G/M												
10	gestreift	geblümt 		kariert G/I	gepunktet 												
11	nass	blass M		trocken G/I	nackt M												
12	schief	gerade 		steil HRP/G/I	stark M												
13	verletzt	gesund 		scharf HRP/G/I	verliebt M/S												
14	blond	braun M/I		rot 	schwarz 												
15	voll	viel HRP/G/M/I		groß G/M/I	stolz G/M												
16	weich	nass G		wach M/I	warm M												
17	nah	nass M		geschlossen G	fern 												
18	schwer	stark HRP/G/I		leer M	schwarz M												
Σ																	

Anh. 1: B 3.7.2a:

Zur Kontrolle der Qualität der Testhalbierung nach der üblichen odd-even-Methode wurde zusätzlich eine Halbierung aufgrund statistischer Kennwerte vorgenommen. In der folgenden Tabelle ist die Zusammenstellung der Aufgabenpaare nach ähnlicher Schwierigkeit und Trennschärfe aufgeführt. Zusätzlich lässt sich die Zuordnung der Aufgaben zu den jeweiligen Testhälften A und B jedes Subtestteils ablesen.

Tab. 46-B 3.7.2: Aufteilung der Items in Testhälften A und B									
Subtest- teil	Aufgabenpaare								
GSN A	1	4	5	8	12	14	16	19	20
GSN B	11	23	17	6	15	10	13	9	24
GSV A	2	5	6	10	13	14	15	19	20
GSV B	12	17	18	9	8	7	16	23	21
GSA A	4	6	7	9	11	14	19	22	23
GSA B	5	2	13	16	24	8	20	15	10
SSN A	1	4	6	9	10	15	20	23	24
SSN B	14	13	17	12	19	11	8	16	5
SSV A	2	8	9	12	13	16	19	20	21
SSV B	5	14	6	15	18	17	7	23	10
SSA A	2	4	7	8	10	13	15	19	22
SSA B	9	20	14	5	6	11	16	23	24
LSN A	1	8	10	12	15	16	20	23	24
LSN B	9	5	11	17	13	4	14	19	6
LSV A	2	5	9	13	14	15	18	19	23
LSV B	6	7	10	21	8	12	20	17	16
LSA A	6	9	11	13	14	16	22	23	24
LSA B	7	5	15	4	19	8	20	10	2

Anh. 1: B 3.7.2b:

Um die Gleichwertigkeit der so konstruierten Testhälften A und B zu prüfen, werden in der folgenden Tabelle die Gesamtwerte (Summenmittelwerte) für Schwierigkeit und Trennschärfe beider Testhälften in jedem Subtestteil einander gegenübergestellt. Weiterhin werden die jeweiligen Rohwertmittelwerte und Varianzen der Testhälften aufgelistet, die für die Berechnung und Auswahl der verschiedenen Halbierungskonsistenz-Formeln zur Berechnung der Reliabilität notwendig sein werden:

Tab. 47-B 3.7.2: Gesamtwerte der Testhälften A und B nach Halbierung über statistische Kennwerte				
Subtest- teil	Mittlere Schwierigkeit	Mittlere Trennschärfe	Rohwert- Mittelwert	Varianz
GSN A	.74	.25	6,6	1,82
GSN B	.71	.21	6,3	1,83
GSV A	.86	.27	7,5	1,58
GSV B	.84	.27	7,4	1,70
GSA A	.69	.25	6,1	1,76
GSA B	.69	.25	6,1	1,84
SSN A	.66	.43	5,8	2,38
SSN B	.67	.43	5,9	2,36
SSV A	.55	.46	4,8	2,54
SSV B	.55	.47	4,7	2,50
SSA A	.56	.46	4,9	2,67
SSA B	.56	.46	4,9	2,66
LSN A	.68	.37	5,8	2,49
LSN B	.66	.37	5,6	2,50
LSV A	.64	.34	5,5	2,36
LSV B	.64	.35	5,5	2,40
LSA A	.58	.43	5,0	2,63
LSA B	.59	.43	5,0	2,60

Es wird deutlich, daß die Aufteilung der Aufgaben nach statistischen Kennwerten zu vergleichbaren Testhälften geführt hat, da die Mittelwerte der Aufgabenschwierigkeiten bzw. -trennschärfen in den Hälften A und B aller Subtestteile weitestgehend übereinstimmen. Die Schwierigkeitswerte weichen durchschnittlich 1%, davon nie stärker als 3% (in Subtestteil

GSN), meistens nur 1% (in 2 Subtestteilen) oder überhaupt nicht (in vier Subtestteilen) voneinander ab. Die Trennschärfewerte weichen durchschnittlich 0,7%, davon nur in Subtestteil GSN 4%, sonst 1% (in zwei Subtestteilen) oder überhaupt nicht (in sechs Subtestteilen) voneinander ab. Trotz der Begünstigung der Werte gegenüber einer vollkommen zufälligen Testhalbierung, sind die Ergebnisse als zufriedenstellend zu bewerten.

Anh. 1: B 3.7.2c:

Nach Anwendung der in B 3.7.2 beschriebenen Halbierungskonsistenzformeln auf die so entstandenen Testhälften ergeben sich annähernd gleiche Reliabilitätskoeffizienten wie für die Testhalbierung nach der odd-even-Methode.

Tab. 48-B 3.7.2: Reliabilitätskoeffizienten der Split-half-Methode nach statistischen Kennwerten		
	r_{tt} Spearman-Brown	r_{tt} Kristof
GSN	.80	.80
GSV	.80	.80
GSA	.81	.81
SSN	.87	.87
SSV	.90	.91
SSA	.86	.86
LSN	.85	.85
LSV	.84	.84
LSA	.88	.88
gesamt über z'	.85	.85

Anh. 1: B 3.7.3:

Tab. 53-B 3.7.3: Lehrer-Rating der passiven Wortschatzkenntnis			
Testperson	GS	SS	LS
8	7	7	6
9	7	7	7
10	4	3	4
11	5	5	5
12	5	5	5
13	7	7	7
14	1	1	1
15	7	6	6
16	3	2	2
17	7	6	6
18	4	4	4
19	4	3	6
20	5	4	4
21	5	4	5
22	4	4	3
23	6	5	4
24	6	4	4
25	5	6	6
26	6	3	3
27	6	3	3
28	5	4	3
29	6	6	5
30	4	3	3
31	2	2	2
32	6	5	5
33	4	3	4
34	3	2	5
35	6	5	6
36	6	4	6
37	4	2	5
38	4	5	3
39	6	5	3
40	3	4	4
41	7	7	7
42	6	7	6
43	6	6	6
44	2	1	2
45	4	4	4
46	3	2	3
47	6	4	5
48	4	4	5
49	4	2	4
50	5	3	5
51	1	1	1
52	6	5	5

Testperson	GS	SS	LS
53	7	5	5
54	4	4	4
55	2	2	2
56	4	3	3
57	3	5	3
58	6	4	4
59	4	5	4
60	3	6	6
61	5	6	6
62	4	2	2
68	6	5	4
69	4	5	5
70	5	5	5
71	4	1	1
72	5	6	5
73	6	6	6
74	4	2	1

Anh. 1: B 3.8.4:

Normierung der Testversion II bei noch nicht bewiesener Normalverteilung über Flächen-
transformation, also über kumulierte Häufigkeiten:

Tab. 55a-B 3.8.4: Normierung über Flächentransformation Subtestteil Gebärdensprache Nomen				
RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	3	3	1,4	78
7	1	4	3,2	81
8	0	4	3,6	82
9	8	12	7,3	85
10	8	20	14,5	89
11	15	35	25	93
12	11	46	33,5	95
13	11	57	46,8	99
14	16	73	59,1	102
15	12	85	71,8	105
16	8	93	80,9	108
17	13	106	90,5	113
18	4	110	98,2	120

**Tab. 55b-B 3.8.4: Normierung über Flächentransformation Subtestteil
Gebärdensprache Verben**

RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	0	0	0	-
7	1	1	0,5	74
8	1	2	1,4	78
9	1	3	2,3	80
10	1	4	3,2	81
11	5	9	5,9	84
12	4	13	10	87
13	5	18	14,1	89
14	12	30	21,8	92
15	21	51	36,8	96
16	23	74	56,8	101
17	26	100	79,1	108
18	10	110	95,5	116

**Tab. 55c-B 3.8.4: Normierung über Flächentransformation Subtestteil
Gebärdensprache Adjektive**

RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	0	0	0	-
5	0	0	0	-
6	0	0	0	-
7	6	6	2,7	80
8	6	12	8,2	86
9	7	19	17,3	90
10	12	31	22,7	92
11	13	44	34,1	95
12	14	58	46,4	99
13	10	68	57,3	101
14	12	80	67,3	104
15	10	90	77,3	107
16	13	103	87,7	111
17	5	108	95,9	117
18	2	110	99,1	123

**Tab. 55d-B 3.8.4: Normierung über Flächentransformation Subtestteil
Schriftsprache Nomen**

RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	0	0	0	-
3	0	0	0	-
4	4	4	1,8	79
5	3	7	5	83
6	6	13	9,1	86
7	6	19	14,5	89
8	9	28	21,4	92
9	9	37	29,5	94
10	5	42	35,9	96
11	11	53	43,2	98
12	7	60	51,4	100
13	5	65	56,8	101
14	7	72	62,3	103
15	8	80	69,1	105
16	8	88	76,4	107
17	11	99	85	110
18	11	110	95	116

**Tab. 55e-B 3.8.4: Normierung über Flächentransformation Subtestteil
Schriftsprache Verben**

RW	f	f_{cum}	PR	Z
0	1	1	0,5	74
1	1	2	1,4	78
2	5	7	4,2	82
3	3	10	7,9	85
4	6	16	12	88
5	7	23	18,1	90
6	9	32	25,5	93
7	6	38	32,4	95
8	4	42	37	96
9	6	48	41,7	97
10	5	53	46,8	99
11	9	62	53,2	100
12	9	71	61,6	102
13	10	81	70,4	105
14	7	88	78,2	107
15	7	95	84,7	110
16	10	105	92,6	114
17	2	107	98,1	120
18	1	108	99,5	126

Tab. 55f-B 3.8.4: Normierung über Flächentransformation Subtestteil
Schriftsprache Adjektive

RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	2	2	0,9	76
3	4	6	3,7	82
4	6	12	8,3	86
5	8	20	14,8	89
6	11	31	23,6	92
7	8	39	32,4	95
8	9	48	40,3	97
9	5	53	46,8	99
10	6	59	51,9	100
11	8	67	58,3	102
12	8	75	65,7	104
13	2	77	70,4	105
14	3	80	72,7	106
15	6	86	76,9	107
16	5	91	81,9	109
17	13	104	90,3	113
18	4	108	98,1	120

Tab. 55g-B 3.8.4: Normierung über Flächentransformation Subtestteil
Lautsprache Nomen

RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	0	0	0	-
2	1	1	0,5	74
3	2	3	1,9	79
4	3	6	4,2	82
5	2	8	6,6	84
6	2	10	8,5	86
7	8	18	13,2	88
8	4	22	18,9	91
9	2	24	21,7	92
10	8	32	26,4	93
11	5	37	32,5	95
12	11	48	40,1	97
13	12	60	50,9	100
14	16	76	64,2	103
15	11	87	76,9	107
16	9	96	86,3	110
17	5	101	92,9	114
18	5	106	97,6	119

**Tab. 55h-B 3.8.4: Normierung über Flächentransformation Subtestteil
Lautsprache Verben**

RW	f	f_{cum}	PR	Z
0	0	0	0	-
1	1	1	0,5	74
2	0	1	0,5	74
3	1	2	1,4	78
4	2	4	2,8	80
5	0	4	3,8	82
6	5	9	6,1	84
7	8	17	12,3	88
8	7	24	19,3	91
9	7	31	25,9	93
10	8	39	33	95
11	12	51	42,5	98
12	10	61	52,8	100
13	5	66	59,9	102
14	11	77	67,5	104
15	12	89	78,3	107
16	12	101	89,6	112
17	3	104	96,7	118
18	2	106	99,1	123

**Tab. 55i-B 3.8.4: Normierung über Flächentransformation Subtestteil
Lautsprache Adjektive**

RW	f	f_{cum}	PR	Z
0	1	1	0,5	74
1	0	1	0,5	74
2	1	2	1,4	78
3	4	6	3,8	82
4	4	10	7,5	85
5	5	15	11,8	88
6	9	24	18,4	91
7	5	29	25	93
8	12	41	33	95
9	5	46	41	97
10	10	56	48,1	99
11	5	61	55,2	101
12	3	64	59	102
13	9	73	64,6	103
14	7	80	72,2	105
15	6	86	78,3	107
16	10	96	85,8	110
17	8	104	94,3	115
18	2	106	99,1	123

Anh. 1: B 3.9.2a:

Die folgende Tabelle führt die mittlere Standardabweichung je zweier Subtestteile, berechnet aus der Wurzel der mittleren Varianzen, je zweier Subtestteile auf:

Tab. 59-B 3.9.2: Standardabweichung je zweier Subtestteile für die Beurteilung intraindividuelle Unterschiede									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	2,87	2,56	2,88	3,60	3,79	3,89	3,40	3,29	3,75
GSV		2,21	2,57	3,36	3,57	3,67	3,15	3,02	3,52
GSA			2,89	3,61	3,80	3,90	3,41	3,30	3,75
SSN				4,21	4,37	4,46	4,04	3,94	4,33
SSV					4,54	4,61	4,21	4,12	4,50
SSA						4,69	4,30	4,21	4,58
LSN							3,86	3,76	4,17
LSV								3,66	4,08
LSA									4,46

Anh. 1: B 3.9.2b:

Tab. 60-B 3.9.2: Reliabilität je zweier Subtestteile für die Beurteilung intraindividueller Unterschiede									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1	.47	.62	.69	.69	.68	.71	.63	.65
GSV		1	.41	.37	.38	.35	.28	.36	.40
GSA			1	.74	.68	.74	.62	.69	.73
SSN				1	.85	.83	.70	.72	.74
SSV					1	.86	.77	.74	.70
SSA						1	.71	.69	.77
LSN							1	.79	.77
LSV								1	.79
LSA									1

Anh. 1: B 3.10.3a:

Standardabweichung je zweier Subtestteile zu allen Distraktorentypen zur Berechnung von kritischen Differenzen:

Tab. 66a-B 3.10.3: Standardabweichung je zweier Subtestteile HRP									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1,40	1,20	1,21	1,40	1,21	1,06	1,42	1,29	1,30
GSV		0,99	1,00	1,19	1,00	0,85	1,21	1,09	1,10
GSA			1,01	1,20	1,01	0,86	1,22	1,09	1,10
SSN				1,39	1,20	1,05	1,41	1,29	1,30
SSV					1,01	0,86	1,22	1,10	1,11
SSA						0,71	1,07	0,95	0,96
LSN							1,43	1,30	1,25
LSV								1,18	1,19
LSA									1,20

Tab. 66b-B 3.10.3: Standardabweichung je zweier Subtestteile MOD									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1,98	1,72	2,08	1,72	2,16	2,01	2,06	2,11	2,20
GSV		1,46	1,82	1,46	1,90	1,75	1,80	1,85	1,95
GSA			2,19	1,82	2,27	2,12	2,17	2,21	2,31
SSN				1,46	1,90	1,75	1,80	1,84	1,94
SSV					2,34	2,19	2,24	2,29	2,39
SSA						2,04	2,09	2,14	2,24
LSN							2,15	2,19	2,29
LSV								2,23	2,33
LSA									2,43

Tab. 66c-B 3.10.3: Standardabweichung je zweier Subtestteile G									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1,98	1,72	2,08	1,90	2,04	2,14	2,05	1,84	2,12
GSV		1,46	1,82	1,64	1,78	1,88	1,79	1,58	1,86
GSA			2,19	2,01	2,14	2,25	2,16	1,94	2,23
SSN				1,82	1,96	2,06	1,97	1,76	2,04
SSV					2,10	2,20	2,11	1,90	2,18
SSA						2,30	2,21	2,00	2,28
LSN							2,12	1,91	2,19
LSV								1,70	1,98
LSA									2,27

Tab. 66d-B 3.10.3: Standardabweichung je zweier Subtestteile M									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1,61	1,30	1,41	1,86	2,05	2,28	1,88	1,92	2,02
GSV		0,99	1,10	1,55	1,74	1,98	1,57	1,61	1,71
GSA			1,21	1,66	1,85	2,08	1,68	1,72	1,82
SSN				2,11	2,30	2,53	2,13	2,17	2,27
SSV					2,49	2,72	2,32	2,36	2,46
SSA						2,96	2,55	2,60	2,69
LSN							2,15	2,19	2,29
LSV								2,23	2,33
LSA									2,43

Tab. 66e-B 3.10.3: Standardabweichung je zweier Subtestteile I									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1,83	2,58	2,02	2,17	2,25	2,44	2,08	2,07	2,43
GSV		1,50	1,85	2,00	2,09	2,28	1,91	1,91	2,27
GSA			2,20	2,35	2,44	2,63	2,26	2,26	2,62
SSN				2,50	2,59	2,78	2,41	2,41	2,77
SSV					2,68	2,87	2,50	2,50	2,86
SSA						3,05	2,69	2,69	3,05
LSN							2,32	2,32	2,68
LSV								2,32	2,68
LSA									3,04

Anh. 1: B 3.10.3b:

Reliabilität je zweier Subtestteile zu allen Distraktorentypen zur Berechnung von kritischen Differenzen:

Tab. 67a-B 3.10.3: Reliabilität je zweier Subtestteile HRP									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1	.29	.29	.24	.15	.05	.31	.13	.24
GSV		1	.22	.22	.29	.01	.08	.12	.14
GSA			1	.35	.31	.21	.17	.22	.34
SSN				1	.37	.27	.41	.18	.34
SSV					1	.20	.22	.11	.30
SSA						1	.23	.11	.03
LSN							1	.33	.29
LSV								1	.13
LSA									1

Tab. 67b-B 3.10.3: Reliabilität je zweier Subtestteile MOD									
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1	.45	.46	.43	.45	.49	.50	.49	.62
GSV		1	.29	.20	.36	.18	.20	.27	.30
GSA			1	.59	.59	.66	.40	.61	.62
SSN				1	.53	.60	.34	.52	.57
SSV					1	.59	.53	.50	.59
SSA						1	.41	.54	.54
LSN							1	.57	.66
LSV								1	.64
LSA									1

Tab. 67c-B 3.10.3: Reliabilität je zweier Subtestteile G

	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1	.45	.46	.53	.51	.43	.60	.25	.40
GSV		1	.29	.32	.42	.16	.32	.18	.19
GSA			1	.59	.59	.56	.56	.28	.63
SSN				1	.56	.57	.48	.38	.51
SSV					1	.59	.48	.38	.32
SSA						1	.45	.31	.46
LSN							1	.35	.50
LSV								1	.30
LSA									1

Tab. 67d-B 3.10.3: Reliabilität je zweier Subtestteile M

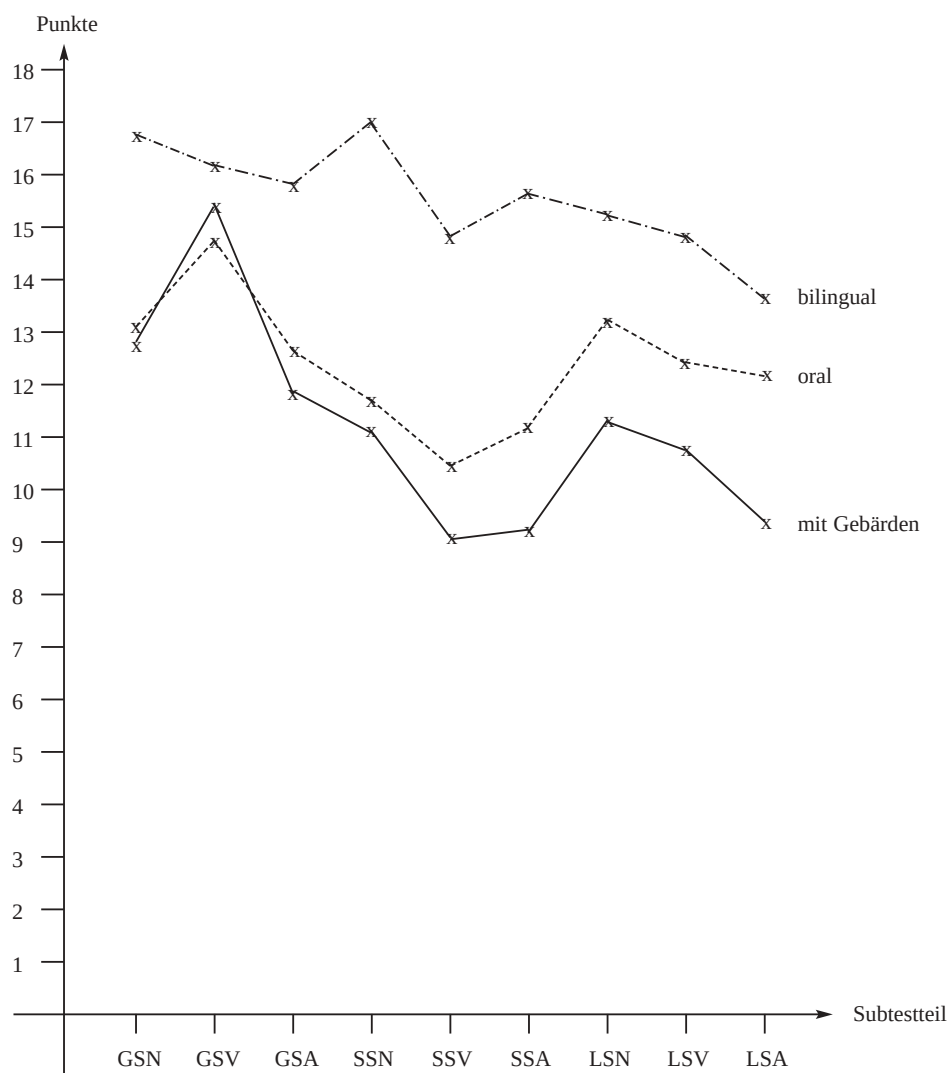
	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1	.33	.35	.51	.52	.51	.50	.45	.45
GSV		1	.27	.15	.24	.23	.04	.21	.20
GSA			1	.42	.40	.52	.41	.53	.56
SSN				1	.67	.69	.42	.52	.61
SSV					1	.64	.55	.50	.61
SSA						1	.40	.58	.60
LSN							1	.57	.66
LSV								1	.64
LSA									1

Tab. 67e-B 3.10.3: Reliabilität je zweier Subtestteile I

	GSN	GSV	GSA	SSN	SSV	SSA	LSN	LSV	LSA
GSN	1	.40	.51	.47	.50	.53	.47	.43	.45
GSV		1	.33	.37	.39	.41	.16	.29	.31
GSA			1	.58	.58	.60	.39	.56	.61
SSN				1	.73	.71	.57	.60	.64
SSV					1	.72	.60	.59	.55
SSA						1	.62	.52	.64
LSN							1	.60	.63
LSV								1	.65
LSA									1

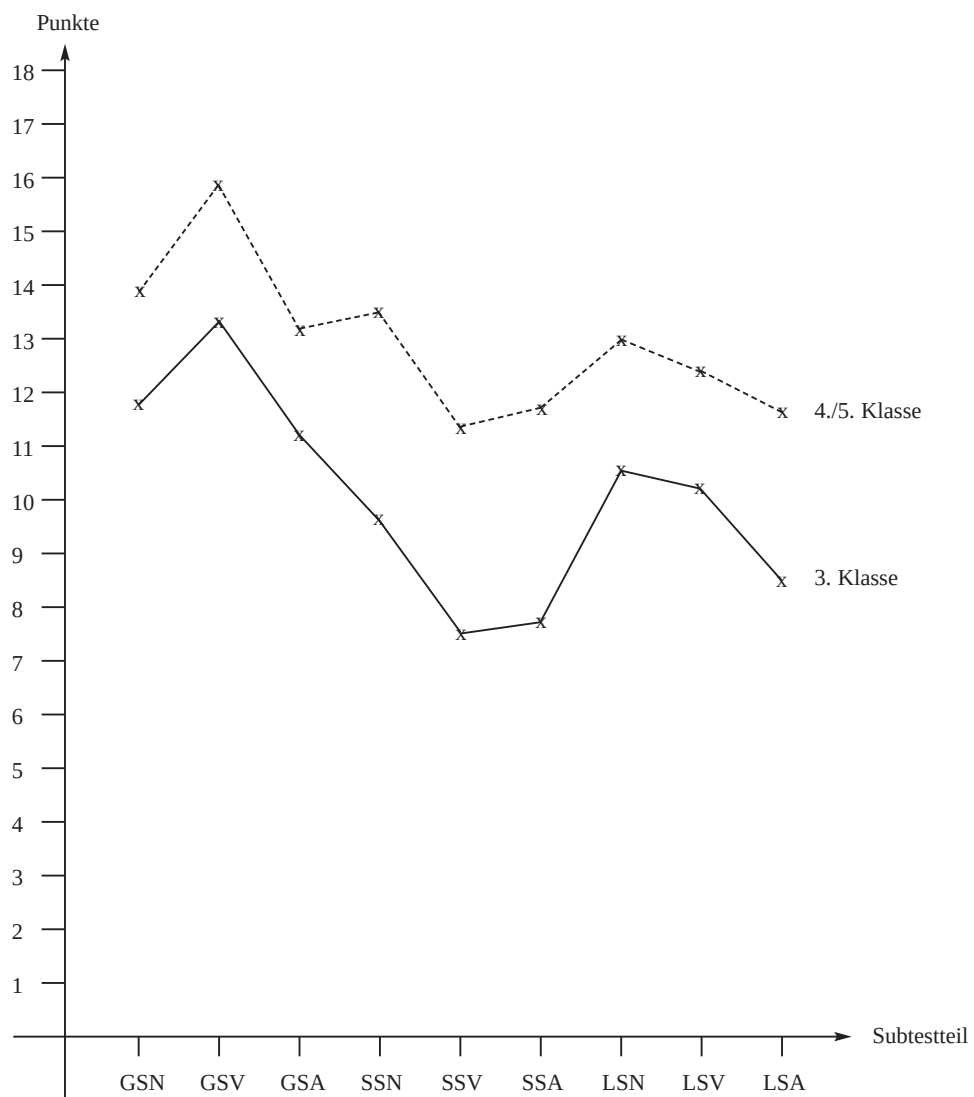
Anh. 1: C 1.7.4:

Abb. 23-C 1.7.4: Profilverläufe Konzepte



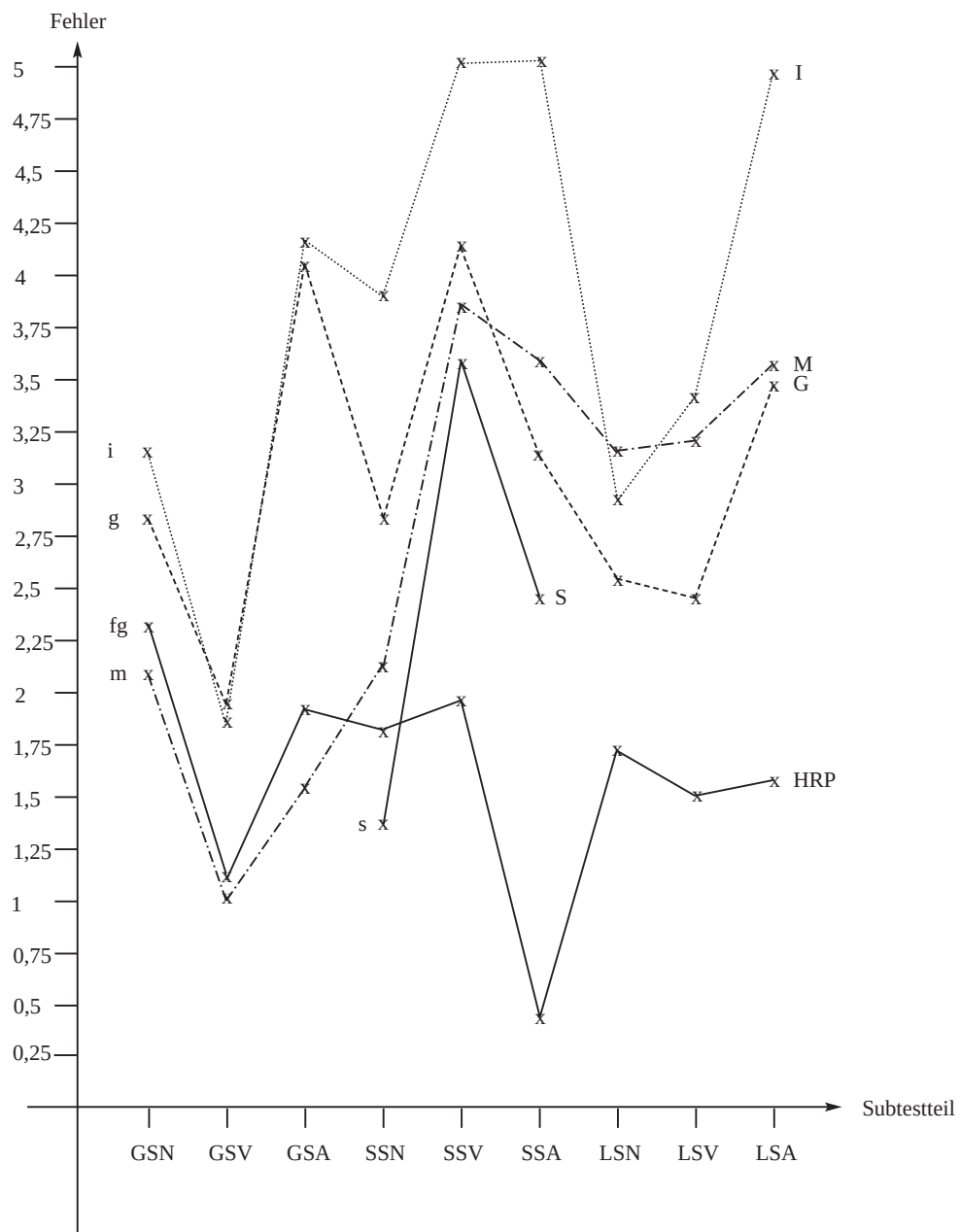
Anh. 1: C 1.7.5:

Abb. 24-C 1.7.5: Profilverläufe Klassen



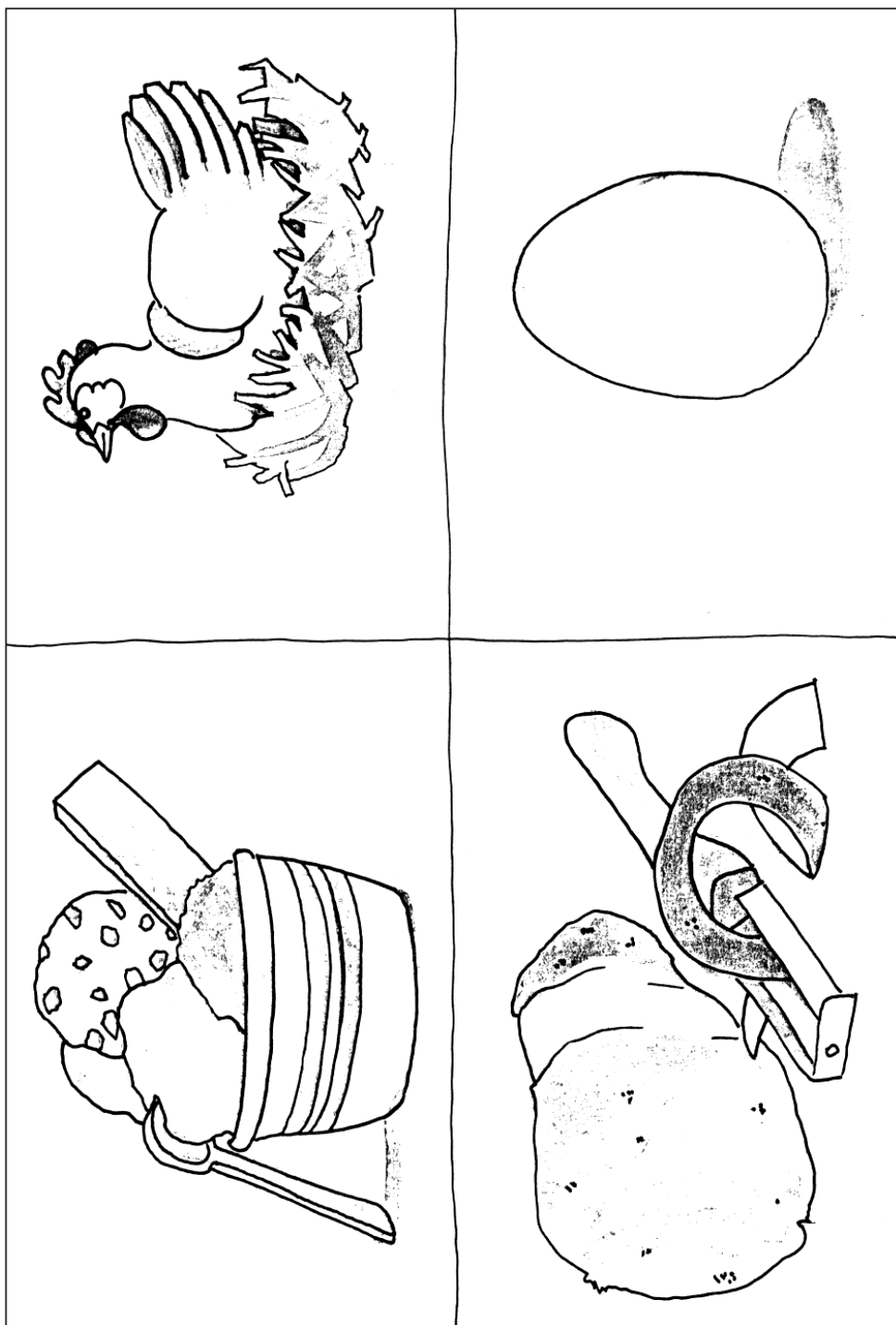
Anh. 1: C 1.8.1:

Abb. 25-C 1.8.1: Fehleranzahl



Anh. 1: Beispielseite Testheft I (Gebärden- und Lautsprache)

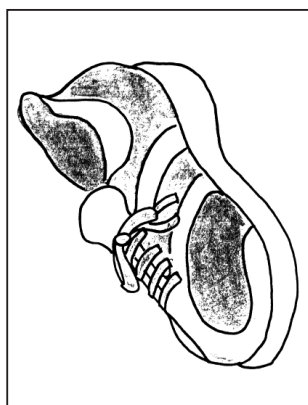
Nomen 1: Ei



Anh. 1: Beispielseite Testheft II (Schriftsprache)

Nomen 1: Schuh

Schuh	Uhr
Stiefel	Strumpf



Wir versichern, dass wir andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzen Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht haben.