

Gedächtnis und kognitive Leistungsfähigkeit

Entwicklung zweier psychometrischer Leistungstests und Validierung ihrer Messgegenstände mit Hilfe eines Modells des menschlichen Gedächtnisses

Vom Fachbereich Humanwissenschaften der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines Doktors der Philosophie (Dr. phil.)
genehmigte Dissertation

von

Gerd Heyde (Dipl.-Psych.), geboren in Groß-Gerau

Referent: Prof. Dr. Reinhard Leichner

Korreferent: Prof. Dr. Bruno Rüttinger

Eingereicht am 16. August 2006

Disputation am 21. Dezember 2006

Hochschulkennziffer: D 17

Darmstadt 2006

Meinen Eltern

Inhalt

Einleitung	1
Teil 1: Erweiterung und Neubewertung der theoretischen und empirischen Grundlagen des Aufmerksamkeitstests Inventar Komplexer Aufmerksamkeit (INKA)	3
1 Ausgangspunkt der Testentwicklung	3
2 Theoretische Fundierung	5
2.1 Aufmerksamkeit und Konzentration	6
2.2 Grundlegende Modelle der Aufmerksamkeit	7
2.2.1 Reizselektionsmodelle	7
2.2.1.1 Modelle der frühen Selektion	8
2.2.1.2 Modelle der späten Selektion	9
2.2.2 Kapazitätsmodelle der Aufmerksamkeit	10
2.2.2.1 Modell einer einheitlichen Aufmerksamkeitsressource	11
2.2.2.2 Modelle multipler Aufmerksamkeitsressourcen	13
3 Beschreibung des Inventars Komplexer Aufmerksamkeit	13
4 Entwicklung der Testaufgaben	15
4.1 Generative Regeln	16
4.2 Linkshänderversion	17
5 Konstruktion der Skala	18
5.1 Aufgabenanalyse nach der Klassischen Testtheorie	18
5.2 Aufgabenanalyse nach der Probabilistischen Testtheorie	21
5.2.1 Das Rasch-Modell	21
5.2.2 Überprüfung von Testaufgaben auf Gültigkeit des Rasch-Modells	23
5.2.3 Ergebnisse der Überprüfung der Aufgaben des INKA	25
5.2.3.1 Teilungskriterium Geschlecht	26
5.2.3.2 Teilungskriterium Alter	27
5.2.3.3 Teilungskriterium Schulbildungsniveau	28
5.2.3.4 Teilungskriterium Testleistung	29
6 Gütekriterien	30
6.1 Objektivität	30

6.2	Reliabilität	30
6.3	Validität	31
6.3.1	Faktorenanalytische Konstruktvalidierung	32
6.3.2	Konvergente und diskriminante Validität	41
6.3.2.1	Intelligenz-Struktur-Test 70 und Aufmerksamkeits-Belastungs- Test d2	42
6.3.2.2	Exkurs: Hauptschüler mit und ohne Berufsausbildung	44
6.3.2.3	Weitere Korrelationsstudien zur konvergenten und diskriminanten Validität	45
6.3.2.4	Die Untersuchung von Schmidt-Atzert et al.	50
6.3.3	Vorläufiges Fazit	55
6.4	Normierung	56
6.4.1	Leistungsunterschiede zwischen Personen unterschiedlichen Geschlechts	58
6.4.2	Leistungsunterschiede zwischen Personen unterschiedlichen Alters	59
6.4.3	Leistungsunterschiede zwischen Personen unterschiedlicher Schulbildung	63
6.4.4	Zusammenhang zwischen Schulbildung und Alter	65
6.4.5	Leistungsunterschiede zwischen Personen unterschiedlicher Händigkeit	70
6.5	Skalierung	73
6.6	Unverfälschbarkeit	73
6.7	Zumutbarkeit	73
6.8	Fairness	74
7	Durchführung des Tests	75
7.1	Instruieren der Testpersonen	75
7.2	Auswertung der Ergebnisse	76
7.3	Interpretation der Ergebnisse	76
 Teil 2: Entwicklung eines Tests zur Erfassung der kristallinen verbalen		
	Intelligenz, Inventar Lexikalischer Kompetenz (ILKO)	78
1	Ausgangspunkt der Testentwicklung	78
2	Theoretische Fundierung	79
2.1	Psychometrische Intelligenzmodelle	79
2.1.1	Die Anfänge	80
2.1.2	Die Modelle von Cattell-Horn und Carroll	81
2.2	Modelle der Sprachkompetenz	85

2.2.1 Sprachkompetenz nach Nodari	85
2.2.2 Sprachkompetenz nach Coseriu.....	87
2.2.3 Sprachkompetenz gemäß dem europäischen Referenzrahmen.....	89
3 Beschreibung des Tests	92
4 Entwicklung der Testaufgaben.....	93
4.1 Korpora geschriebener Gegenwartssprache.....	93
4.2 Auswahl der Zielwörter	94
4.3 Beschreibung der inhaltlichen Bedeutungen der Zielwörter	95
5 Konstruktion der Skala.....	96
5.1 Itemselektion nach Synonymen.....	100
5.2 Itemselektion nach Schwierigkeit.....	100
5.3 Itemselektion nach Trennschärfe.....	102
5.4 Semantische Struktur der Skala.....	104
6 Gütekriterien.....	108
6.1 Objektivität	109
6.2 Reliabilität	109
6.3 Validität	109
6.3.1 Inhaltsvalidität	109
6.3.2 Konstruktvalidität	110
6.4 Normierung.....	117
6.4.1 Leistungsunterschiede zwischen Personen unterschiedlichen Geschlechts.....	121
6.4.2 Leistungsunterschiede zwischen Personen unterschiedlichen Alters	122
6.4.3 Leistungsunterschiede zwischen Personen mit unterschiedlichen Schulfremdsprachen.....	124
6.4.4 Leistungsunterschiede zwischen Personen unterschiedlicher Studienfachgebiete.....	126
6.5 Nützlichkeit	128
6.6 Unverfälschbarkeit	136
6.7 Zumutbarkeit	136
6.8 Fairness.....	137
7 Durchführung des Tests	137
7.1 Instruieren der Testpersonen	138
7.2 Auswertung der Ergebnisse.....	138
7.3 Interpretation der Ergebnisse.....	139

Teil 3: Modellgeleitete Konstruktvalidierung des Inventars Lexikalischer	
Kompetenz und des Inventars Komplexer Aufmerksamkeit	140
1 Einleitung	140
2 Ein Gedächtnismodell zur Beschreibung und Einordnung der Tests.....	141
3 Das Drei-Komponenten-Modell des Arbeitsgedächtnisses von Baddeley und Hitch.....	143
3.1 Die phonologische Schleife	147
3.2 Der räumlich-visuelle Notizblock	148
3.3 Die zentrale Exekutive	148
3.4 Weiterentwicklung des Drei-Komponenten-Modells durch Baddeley	150
3.5 Modifizierung des Drei-Komponenten-Modells durch Logie.....	150
3.5.1 Die Struktur des räumlich-visuellen Notizblocks	151
3.5.2 Die Funktion des Langzeitgedächtnisses bei der Informationsaufnahme	152
3.5.3 Das modifizierte Modell.....	154
4 Die zur modellgeleiteten Konstruktvalidierung eingesetzten Tests.....	155
4.1 Ableitung der für die Bearbeitung der Tests relevanten kognitiven Prozesse.....	156
4.1.1 Inventar Lexikalischer Kompetenz.....	157
4.1.2 Inventar Komplexer Aufmerksamkeit	157
4.1.3 Figure Reasoning Test	158
4.1.4 Verbaler Kurz-Intelligenztest	159
4.1.5 Mehrfachwahl-Wortschatz-Intelligenztest.....	159
4.1.6 Mannheimer Rechtschreib-Test.....	159
4.1.7 Intelligenz-Struktur-Analyse: Beziehungen erschließen	160
4.1.8 Intelligenz-Struktur-Test 70.....	160
4.2 Zusammenfassung und Vergleich der relevanten kognitiven Prozesse.....	160
5 Hypothesen.....	163
6 Empirische Überprüfung der Hypothesen.....	164
6.1 Untersuchungsstichproben	164
6.2 Ergebnisse.....	166
Fazit	169
Literatur	172

Anhang	183
A1 ILKO-Testbogen	185
A2 Abstract	187
A3 Akademischer Lebenslauf des Verfassers	187
A4 Erklärung	187

Verzeichnis der Abbildungen

Abb. 1.1: Reizselektionsmodelle der Aufmerksamkeit	10
Abb. 1.2: Kapazitätsmodell der Aufmerksamkeit nach Kahneman	12
Abb. 1.3: Aufgabendesign der Rechtshänderversion des INKA	14
Abb. 1.4: Generative Regeln zur Konstruktion der Aufgaben des INKA	17
Abb. 1.5: Aufgabendesign der Linkshänderversion des INKA	17
Abb. 1.6: Logistische Funktion des Testmodells von Rasch	22
Abb. 1.7: Grafischer Modelltest, Teilungskriterium Geschlecht	26
Abb. 1.8: Grafischer Modelltest, Teilungskriterium Alter	27
Abb. 1.9: Grafischer Modelltest, Teilungskriterium Schulbildungsniveau	28
Abb. 1.10: Grafischer Modelltest, Teilungskriterium Testleistung	29
Abb. 1.11: Testwerte, die in die erste Validitätsuntersuchung Eingang fanden	35
Abb. 1.12: Diagramm der Eigenwerte der durch die Hauptkomponentenanalyse gewonnenen Faktoren	39
Abb. 1.13: Testwerte, die mit dem Ergebnis im INKA korreliert wurden	49
Abb. 1.14: Histogramm der Rohwerteverteilung in der Normstichprobe des INKA	57
Abb. 1.15: Balkendiagramm der mittleren Testleistungen im INKA bezogen auf die Variable ‚Alter‘	59
Abb. 1.16: Balkendiagramm der mittleren Testleistungen im INKA bezogen auf die Variable ‚Schulbildung‘	63
Abb. 1.17: Balkendiagramm der mittleren Testleistungen in den SPM bezogen auf die Variable ‚Schulbildung‘	68

Abb. 2.1: Vergleich der Gf-Gc Intelligenztheorie von Cattell-Horn mit der Three-Stratum Intelligenztheorie von Carroll	84

Abb. 2.2:	Dimensionen der Sprachkompetenz nach Nodari	86
Abb. 2.3:	Einteilung der Sprachkompetenz nach Coseriu.....	87
Abb. 2.4:	Kompetenzen eines Sprachverwenders gemäß des ‚Gemeinsamen europäischen Referenzrahmens für Sprachen: lernen, lehren, beurteilen‘	90
Abb. 2.5:	Die erste Aufgabe des Inventars Lexikalischer Kompetenz	92
Abb. 2.6:	Kriterien für die Auswahl der Zielwörter des ILKO aus der Liste der 30.000 häufigsten Wortformen der Textkorpora des IDS.....	94
Abb. 2.7:	Zielwörter der ersten Testvorform des ILKO.....	96
Abb. 2.8:	Personenbezogene Daten, die im Rahmen der Entwicklung des ILKO erhoben wurden	97
Abb. 2.9:	Beispiele für Items, zu denen korrekte Alternativantworten gefunden wurden.....	100
Abb. 2.10:	Testwerte, die in die erste Untersuchung der Konstruktvalidität des ILKO Eingang fanden	111
Abb. 2.11:	Histogramm der Rohwerteverteilung in der Normstichprobe des ILKO	120
Abb. 2.12:	Balkendiagramm der mittleren ILKO-Testwerte bezogen auf die Variable ‚Alter‘	122

Abb. 3.1:	Informationsfluss im modalen Modell des Gedächtnisses von Atkinson und Shiffrin.....	145
Abb. 3.2:	Drei-Komponenten-Modell des Arbeitsgedächtnisses von Baddeley und Hitch	146
Abb. 3.3:	Doppeldeutige Figur aus dem Experiment von Chambers und Reisberg	153
Abb. 3.4:	Modifiziertes Drei-Komponenten-Modell des Arbeitsgedächtnisses nach Logie	154

Verzeichnis der Tabellen

Tab. 1.1:	Soziodemografische Daten der Stichprobe zur Analyse der Aufgaben des INKA.....	20
Tab. 1.2:	Ergebnis der Analyse der Aufgaben des INKA nach der Klassischen Testtheorie.....	20
Tab. 1.3:	Logarithmen der produktnormierten Aufgabenparameterschätzungen auf der Basis der Testdaten der Personen der Gruppen A und B.....	26
Tab. 1.4:	Logarithmen der produktnormierten Aufgabenparameterschätzungen auf der Basis der Testdaten der Personen der Gruppen C und D.....	27
Tab. 1.5:	Logarithmen der produktnormierten Aufgabenparameterschätzungen auf der Basis der Testdaten der Personen der Gruppen E und F.....	28
Tab. 1.6:	Logarithmen der produktnormierten Aufgabenparameterschätzungen auf der Basis der Testdaten der Personen der Gruppen G und H.....	29
Tab. 1.7:	Mittelwerte und Standardabweichungen des Gesamtrohwerts des IST 70 bezogen auf die Schulbildung.....	33
Tab. 1.8:	Soziodemografische Daten der Stichprobe der ersten Validitätsuntersuchung.....	34
Tab. 1.9:	Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen.....	36
Tab. 1.10:	Anti-Image-Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen.....	37
Tab. 1.11:	Ergebnis der Hauptkomponentenanalyse.....	39
Tab. 1.12:	Kommunalitäten der Variablen und Ladungen der Variablen auf den rotierten Faktoren.....	40
Tab. 1.13:	Korrelationen und Redundanzen zwischen INKA und Test d2, IST 70-gesamt und Schulbildung.....	43
Tab. 1.14:	Soziodemografischer Aufbau der Stichprobe zur Berechnung der Mittelwertsunterschiede in der Testleistung im INKA bei Hauptschülern mit und ohne zusätzliche Berufsausbildung.....	44
Tab. 1.15:	Ergebnis des t-Tests mit der abhängigen Variablen ‚Testwert‘ und der Gruppierungsvariablen ‚Berufsausbildung‘.....	45
Tab. 1.16:	Soziodemografische Daten der Stichproben der weiteren Korrelationsstudien zur konvergenten und diskriminanten Validität.....	49

Tab. 1.17: Korrelationen und Redundanzen zwischen INKA und ISA, SPM, APM, FAIR und WST-ap	50
Tab. 1.18: Faktorladungen der Tests und aufgeklärte Testvarianz im 5-Faktoren-Modell.....	52
Tab. 1.19: Faktorladungen der Tests und aufgeklärte Testvarianz im 6-Faktoren-Modell.....	53
Tab. 1.20: Soziodemografische Daten der Normstichprobe des INKA	56
Tab. 1.21: Ergebnis des t-Tests mit der abhängigen Variablen ‚INKA-Testwert‘ und der Gruppierungsvariablen ‚Geschlecht‘	58
Tab. 1.22: Kategorienhäufigkeiten der Altersverteilung	60
Tab. 1.23: Altersgruppen, die zur Überprüfung der Leistungsunterschiede gebildet wurden	61
Tab. 1.24: Deskriptive Statistiken des INKA-Rohwerts für die 15 Altersgruppen	61
Tab. 1.25: Mittlere Ränge der INKA-Testleistungen für die 15 Altersgruppen.....	62
Tab. 1.26: Deskriptive Statistiken des INKA-Rohwerts für die Schulbildungsgruppen.....	64
Tab. 1.27: Mittlere Ränge der INKA-Testleistungen für die Schulbildungsgruppen.....	64
Tab. 1.28: Ergebnis des t-Tests mit der abhängigen Variablen ‚INKA-Testwert‘ und der Gruppierungsvariablen ‚Bildung: Gymnasiasten/Fachoberschüler – Akademiker‘	65
Tab. 1.29: Durchschnittliches Alter der Testpersonen bezogen auf die Schulbildung.....	66
Tab. 1.30: Ergebnis des t-Tests mit der abhängigen Variablen ‚Alter‘ und der Gruppierungsvariablen ‚Bildung: Gymnasiasten/Fachoberschüler – Akademiker‘	67
Tab. 1.31: Deskriptive Statistiken des SPM-Rohwerts für die vier Bildungsgruppen	69
Tab. 1.32: Ergebnis der einfaktoriellen Varianzanalyse mit der abhängigen Variablen ‚SPM-Testwert‘ und der Faktorvariablen ‚Schulbildung‘	69
Tab. 1.33: Ergebnis der Einzelvergleiche mit Hilfe des Scheffé-Tests	70
Tab. 1.34: Ergebnis des t-Tests mit der abhängigen Variablen ‚INKA-Testwert‘ und der Gruppierungsvariablen ‚Händigkeit‘	71
Tab. 1.35: Ergebnisse der t-Tests mit den abhängigen Variablen ‚INKA-Aufgabenwert‘ und der Gruppierungsvariablen ‚Händigkeit‘	72

Tab. 2.1:	Schüler/innen mit fremdsprachlichem Unterricht an den allgemein bildenden Schulen der Schuljahre 2000/01, 2001/02 und 2002/03 in der Bundesrepublik Deutschland.....	98
Tab. 2.2:	Herkunft und Sprachverwandtschaft der Zielwörter des ILKO	99
Tab. 2.3:	Schwierigkeiten der Items der Testvorform mit 41 Aufgaben.....	101
Tab. 2.4:	Soziodemografische Daten der Stichprobe zur Aufgabenanalyse	102
Tab. 2.5:	Ergebnisse der Analyse der Aufgaben des ILKO nach der Klassischen Testtheorie	103
Tab. 2.6:	Kennwerte der Aufgaben der Endform des ILKO	103
Tab. 2.7:	Anti-Image-Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen	105
Tab. 2.8:	Ergebnis der Hauptkomponentenanalyse der Skala	106
Tab. 2.9:	Kommunalitäten der Variablen und Ladungen der Variablen auf den rotierten Faktoren	107
Tab. 2.10:	Soziodemografische Daten der Stichprobe der Faktorenanalyse zur Konstruktvalidierung.....	110
Tab. 2.11:	Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen	112
Tab. 2.12:	Anti-Image-Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen	112
Tab. 2.13:	Ergebnis der Hauptkomponentenanalyse	113
Tab. 2.14:	Kommunalitäten der Variablen und Ladungen der Variablen auf den rotierten Faktoren	114
Tab. 2.15:	Soziodemografische Daten und Testergebnisse der Testpersonen mit einer anderen Muttersprache als Deutsch.....	118
Tab. 2.16:	Soziodemografische Daten der Personen der Normstichprobe des ILKO	119
Tab. 2.17:	Kennwerte der Aufgaben des ILKO, n = 581	120
Tab. 2.18:	Ergebnis des t-Tests mit der abhängigen Variablen ‚ILKO-Testwert‘ und der Gruppierungsvariablen ‚Geschlecht‘	121
Tab. 2.19:	Kategorienhäufigkeiten der Altersverteilung	123
Tab. 2.20:	Ergebnis des t-Tests mit der abhängigen Variablen ‚ILKO-Testwert‘ und der Gruppierungsvariablen ‚Alter‘	123
Tab. 2.21:	Zweite Schulfremdsprache der Personen, die Englisch als erste Schulfremdsprache angaben.....	124

Tab. 2.22: Deskriptive Statistiken des ILKO-Rohwerts für die vier Fremdsprachengruppen	125
Tab. 2.23: Ergebnis der einfaktoriellen Varianzanalyse mit der abhängigen Variablen ‚ILKO-Testwert‘ und der Faktorvariablen ‚Zweite Fremdsprache‘	125
Tab. 2.24: Ergebnis der Einzelvergleiche mit Hilfe des Scheffé-Tests	126
Tab. 2.25: Studienfachgebiete der Personen der Normstichprobe des ILKO.....	127
Tab. 2.26: Deskriptive Statistiken des ILKO-Rohwerts für die sechs Fachgebiete	127
Tab. 2.27: Mittlere Ränge der INKA-Testleistungen für die sechs Fachgebiete.....	128

Tab. 3.1: Unterschiede der kognitiven Informationsverarbeitung bei der Bearbeitung der Aufgaben der Tests	162
Tab. 3.2: Hypothesen über die Zusammenhänge zwischen den eingesetzten Tests.....	164
Tab. 3.3: Soziodemografische Daten der Gesamtstichprobe zur modellgeleiteten Konstruktvalidierung.....	165
Tab. 3.4: Testbatterien, -personen, -orte und -zeiträume.....	165
Tab. 3.5: Ergebnis der empirischen Überprüfung der Zusammenhangshypothesen	166
Tab. 3.6: Korrelative Zusammenhänge zwischen ILKO, INKA und den Untertests des IST-70	167

Verzeichnis der Abkürzungen

(soweit nicht im Text erklärt)

AM – arithmetisches Mittel	p – Wahrscheinlichkeit
df – Freiheitsgrade	r – Korrelation
k. A. – keine Angabe	Red _(yx) – Redundanz
KI – Konfidenzintervall	s – Standardabweichung
Max. – Maximum	SE – Standardfehler
Min. – Minimum	Sig. – Signifikanz

Einleitung

„The strong relationship between working memory and intelligence paves the way for a better understanding of psychometric ability concepts through theories of cognition”

(Süß, Oberauer, Wittmann, Wilhelm & Schulze 2002, S. 284).

Gegenstand der Arbeit

Die vorliegende Arbeit beschreibt die Entwicklung zweier psychometrischer Leistungsskalen, das *Inventar Komplexer Aufmerksamkeit* (INKA; Heyde 1995) und das *Inventar Lexikalischer Kompetenz* (ILKO), die theoretische Einbettung ihrer Messgegenstände und ihre Validierung.

Die erste Überprüfung der Gültigkeit der beiden Verfahren orientiert sich an dem klassischen Konzept der Konstruktvalidierung mit Hilfe von Tests, die explizit denselben Messanspruch besitzen (konvergente Validität) und solchen, die explizit einen anderen Messanspruch haben (diskriminante Validität). Da dieser gängige Ansatz jedoch nicht frei von Problemen ist, wird zusätzlich eine modellgeleitete Konstruktvalidierung durchgeführt. Zugrunde gelegt wird hierfür ein Modell der kognitiven Informationsspeicherung und -verarbeitung.

Aus der kognitionspsychologischen Perspektive gesehen messen die beiden Tests jeweils unterschiedliche funktionale Aspekte des menschlichen Gedächtnisses: Das INKA erweist sich als ein Test der selektiven visuellen Aufmerksamkeit, der auf Grund seiner Aufgabenstruktur auch die Funktionen des Arbeitsgedächtnisses erfasst, nämlich die temporäre Speicherung und Verarbeitung von Informationen und die Koordinierung dieser Prozesse. Das ILKO hingegen misst den Abruf von Inhalten des semantischen Langzeitgedächtnisses. Aus der Perspektive der psychometrischen Intelligenztheorien ist INKA ein Test, der neben Aufmerksamkeitsleistungen auch die Basismechanismen der fluiden Intelligenz erfasst; ILKO ist ein Maß der kristallinen Intelligenz.

Aufbau der Arbeit

Teil I stellt die Entwicklung des Aufmerksamkeitstests *Inventar Komplexer Aufmerksamkeit* dar und liefert auf der Basis neuer empirischer Daten eine aktualisierte Bewertung seiner

psychometrischen Qualität. Zu Beginn werden grundlegende Modelle der Aufmerksamkeit beschrieben, um den Messgegenstand des Tests theoretisch einzuordnen. Anschließend wird die Entwicklung der Testaufgaben und der Skala dargestellt. Den Kern der Überprüfung der psychometrischen Qualität des Verfahrens bildet die Bestimmung der konvergenten und der diskriminanten Validität im „klassischen“ Sinne. Mit Hilfe der Standardisierungsstichprobe von insgesamt 8935 Personen erfolgt abschließend eine Untersuchung von Leistungsunterschieden im Test, bezogen auf das Geschlecht, das Alter und die Schulbildung der Testpersonen.

Teil 2 dokumentiert die Entwicklung eines neuen Tests zur Messung des Wortschatzes, das *Inventar Lexikalischer Kompetenz*. Der Test soll als Indikator der kristallinen Intelligenz sensu Cattell (1963) und der Sprachkompetenz bei Personen mit höherer schulischer Bildung dienen. Zur theoretischen Einordnung seines Messgegenstands werden psychometrische Intelligenzmodelle und relevante Modelle der Sprachkompetenz dargestellt. Es folgt die Beschreibung der Entwicklung der Testaufgaben mit Hilfe der Korpora geschriebener Gegenwartssprache des Instituts für Deutsche Sprache, Mannheim, und die Darstellung der Konstruktion der Skala, inklusive einer Analyse ihrer semantischen Struktur. Erste Hinweise für die Gültigkeit des ILKO werden mit der Herleitung seiner Inhaltsvalidität und einer faktorenanalytischen Konstruktvalidierung geliefert. Abschließend werden Leistungsunterschiede in Bezug auf das Geschlecht, das Alter, die erlernten Schulfremdsprachen und das Studienfachgebiet der Testpersonen analysiert.

Teil 3 stellt die modellgeleitete Konstruktvalidierung der Tests INKA und ILKO dar. Zunächst werden das Drei-Komponenten-Modell des Arbeitsgedächtnisses von Alan Baddeley und Graham Hitch (1974) und seine durch Robert Logie (1995) modifizierte Variante vorgestellt. Anschließend werden die zur Validierung eingesetzten Tests dargestellt und die kognitiven Prozesse, die bei der Bearbeitung ihrer Aufgaben ablaufen, modellorientiert beschrieben. Die Beschreibungen der kognitiven Prozesse dienen als Grundlage für die Bildung von Hypothesen über Zusammenhänge zwischen den mentalen Leistungen, die bei der Bearbeitung der Testverfahren erbracht werden. Die Hypothesen werden abschließend einer empirischen Überprüfung unterzogen.

Teil 1: Erweiterung und Neubewertung der theoretischen und empirischen Grundlagen des Aufmerksamkeitstests Inventar Komplexer Aufmerksamkeit (INKA)

1 Ausgangspunkt der Testentwicklung

Die Entwicklung des Inventars Komplexer Aufmerksamkeit (Heyde, 1995) wurde ursprünglich angeregt durch die grundlegende Kritik an klassischen Konzentrationstests, die auf dem Durchstreichprinzip von Bourdon (vgl. Bourdon, 1955) beruhen. Bei diesen Tests sollen gleichmäßig einfache visuelle Diskriminierungsaufgaben so schnell wie möglich bearbeitet werden. Dabei werden im Allgemeinen zwei Kennwerte erhoben, die Anzahl der bearbeiteten Stimuli (z. B. Buchstaben, Zahlen, Zeichen) und die Anzahl der dabei gemachten Fehler. Bei der abschließenden Berechnung der Testwerte tritt ein gravierendes Problem auf, das im Folgenden in Anlehnung an Schmidt-Atzert, Büttner & Bühner (2004) exemplarisch erläutert werden soll: Ein Speed-Konzentrationstest besteht z. B. aus 100 Stimuli.

Testperson 1 entscheidet sich dafür, schnell zu arbeiten. Sie bearbeitet alle 100 Stimuli und macht dabei 50 Fehler. Testperson 2 hingegen entscheidet sich dafür, langsam und sorgfältig zu arbeiten. Sie bearbeitet nur 50 Stimuli, diese aber fehlerfrei. Da sich für beide Personen ein „fehlerkorrigierter“ Gesamtwert von 50 ergibt (Anzahl aller bearbeiteten Stimuli minus Anzahl der fehlerhaft bearbeiteten Stimuli), wird beiden Personen die gleiche Konzentrationsleistung bescheinigt. Die individuellen Testleistungen bedürfen aber offenbar vollkommen unterschiedlicher Interpretationen. Oehlschlägel und Moosbrugger (1991a) konnten für den Aufmerksamkeits-Belastungs-Test d2 (Brickenkamp, 1987) sogar zeigen, dass durch ein schnelles, zufälliges Markieren von Buchstaben in den Suchreihen fehlerkorrigierte Gesamtwerte resultierten, die den „Testpersonen“ die Fähigkeit zu konzentriertem Arbeiten bescheinigten. Ihre Erkenntnisse kleideten Sie in die provokante Frage „Konzentrationsleistung ohne Konzentration?“.

Der fachliche Disput, der dieser Erkenntnis folgte (Oehlschlägel & Moosbrugger, 1991a; Brickenkamp, 1991a; Oehlschlägel & Moosbrugger, 1991b; Brickenkamp, 1991b; Oehlschlägel & Moosbrugger, 1993; Brickenkamp, 1993), die Tatsache, dass auch andere Ansätze zur Messung der Konzentrationsfähigkeit mit nicht unerheblichen Mängeln behaftet

sind (Fay & Stumpf, 1992; Schmidt-Atzert, Büttner & Bühner, 2004) und die bis dato unzureichenden Versuche, durch die Entwicklung neuer Testkonzepte die Messung der Konzentrationsfähigkeit zu verbessern (Schäfer & Moosbrugger, 1993; Moosbrugger & Oehlschlägel, 1996), waren der Anlass, mit INKA ein neuartiges Aufgabenprinzip zur Messung der Konzentrationsleistung vorzustellen.

Die klassischen Tests zur Erfassung der Konzentrationsfähigkeit, seien es Durchstreichtests, Sortieraufgaben oder Tests, bei denen mathematische Operationen ausgeführt werden sollen, bestehen aus uniformen Aufgaben mit einem gleichmäßig geringen Schwierigkeitsgrad. Die Aufgaben sollen unter Zeitdruck bearbeitet werden. Als Indikator für die Ausprägung der Konzentrationsfähigkeit wird die Anzahl der korrekt bearbeiteten Stimuli bzw. Aufgaben herangezogen. Für die Bearbeitungsleistung ist also nicht die Aufgabenschwierigkeit sondern die Bearbeitungsschnelligkeit der Testperson von ausschlaggebender Bedeutung. Betrachtet man diese diagnostische Information, so ist fraglich, ob sie für alle Fragestellungen, die die Konzentrationsfähigkeit betreffen, Relevanz besitzt. Häufig ist nicht die Schnelligkeit einer Person bei der konzentrierten Bearbeitung einfacher Aufgaben von Interesse sondern das absolute Niveau ihrer Konzentrationsfähigkeit.

Fay und Stumpf (1992) weisen auf einen weiteren wichtigen Kritikpunkt an Speed-Konzentrationstests hin: Die Testleistung hängt bei diesen Verfahren in hohem Maße von der aktuellen Motivationslage der Testperson ab. Ursächlich hierfür sind die geringe intellektuelle Beanspruchung und die hohe Monotonie, die bei der Bearbeitung der uniformen, gleichmäßig einfachen Aufgaben besteht. Aus den Testergebnissen lassen sich nach ihrer Meinung allenfalls Aussagen über die geringstmögliche Konzentrationsleistung treffen.

Mit dem Inventar Komplexer Aufmerksamkeit soll ein deutlich anderer Weg zur Messung der Konzentrationsfähigkeit mit Hilfe von visuellen Diskriminierungsleistungen beschritten werden. Die wichtigsten Charakteristika des Tests sind:

1. INKA enthält 18 separate Aufgaben. Jede Aufgabe besteht aus eine Reihe von Stimuli. Die Testperson muss eine Stimulireihe *insgesamt* fehlerfrei bearbeiten, damit sie einen Rohwertpunkt erhält. Diese Aufgabenstellung verhindert, dass auch durch schnelles, unkonzentriertes Arbeiten Rohwertpunkte erzielt werden können.

2. Der Schwierigkeitsgrad der Aufgaben steigt im Laufe der Testbearbeitung an. Hierdurch erfolgt eine Messung des Fähigkeitsniveaus (power). D. h., Testpersonen mit hoher Konzentrationsfähigkeitskapazität werden, im Gegensatz zu Personen mit geringer Konzentrationsfähigkeitskapazität, auch Aufgaben mit hohem Schwierigkeitsgrad fehlerfrei bearbeiten können.
3. Die wechselnden Inhalte der Aufgaben, die für die Steigerung des Schwierigkeitsgrades genutzt werden, und die unterschiedlichen Aufgabenschwierigkeiten selbst bewirken, dass die Bearbeitung des Tests nicht monoton ist. Die aktuelle Motivationslage der Testperson wird von den Inhalten der Aufgaben weniger negativ als herkömmlich beeinflusst.

Bevor die Entwicklung des Inventars Komplexer Aufmerksamkeit dargestellt wird, soll vorab eine Klärung dessen erfolgen, was unter den psychologischen Konstrukten ‚Aufmerksamkeit‘ und ‚Konzentration‘ verstanden wird. Die Definitionen sollen helfen, den Messgegenstand des INKA theoretisch einzuordnen.

2 Theoretische Fundierung

Aufmerksamkeitsleistungen sind ein zentraler Gegenstand der psychologischen Forschung. Sie umfassen sowohl die Auswahl bestimmter Aspekte der physischen Umwelt als auch die Auswahl von Inhalten des Gedächtnisses, um sie weiterzuverarbeiten.

Aufmerksamkeitsleistungen werden klassifiziert als selektiv, geteilt, kurzfristig, dauerhaft, bewusst/kontrolliert und unbewusst/automatisiert. Der Begriff Aufmerksamkeit lässt sich allgemein definieren als die mentale Fähigkeit, verhaltensrelevante Stimuli, Gedächtnisinhalte, kognitive Operationen und Reaktionen aus einer Masse von Stimuli, Gedächtnisinhalten, kognitiven Operationen und Reaktionen zu selektieren, deren verbleibender Rest nicht verhaltensrelevant ist (Raz, 2004; Vecera & Luck, 2002). Oder nach den Worten von Cohen, Aston-Jones, und Gilzenrat (2004, S. 71): „Attention is the emergent property of the cognitive system that allows it to successfully process some resources of information to the exclusion of others, in the service of achieving some goals to the exclusion of others“.

2.1 Aufmerksamkeit und Konzentration

Schmidt-Atzert, Büttner und Bühner (2004) zitieren eine bemerkenswerte Untersuchung an Laien, die aufzeigt, dass es beim Alltagsgebrauch der deutschen Begriffe Aufmerksamkeit und Konzentration deutliche Unterschiede gibt: 68 Studenten wurden gebeten, Situationen zu schildern, bei denen sie selbst oder andere konzentriert, unkonzentriert, aufmerksam oder unaufmerksam waren. Als typische Konzentrationssituationen wurden von den Studenten ‚Prüfung‘ und ‚Lesen‘ genannt. In beiden Situationen wird die Verarbeitung aller aufgenommenen Informationen verlangt. Typische Aufmerksamkeitsituationen sind hingegen ‚Gespräch‘ oder ‚Teilnahme am Verkehr‘. Hier spielt die Selektion von relevanten Informationen eine herausragende Rolle. Es gibt jedoch auch Situationen, bei denen Aufmerksamkeit und Konzentration wechselseitig genannt wurden, z. B. ‚Vortrag‘. Das ist leicht nachvollziehbar, da je nach konkreter Gegebenheit entweder über eine Informationsselektion Störreize ausgeschaltet werden müssen oder Denkarbeit geleistet werden muss, um den Sinn der aufgenommenen Informationen zu erfassen. Zusätzlich sollten die Studenten für jede genannte Situation einstufen, wie konzentriert *und* wie aufmerksam die betroffene Person war. Hätten Konzentration und Aufmerksamkeit weitgehend die gleiche Bedeutung, müssten beide Einstufungen hoch korrelieren. Die resultierenden vier Korrelationskoeffizienten lagen jedoch nur zwischen $r = -.03$ ($\text{Red}_{(yx)} = 0,09\%$) und $r = .34$ ($\text{Red}_{(yx)} = 11,56\%$). Laien kennen also typische Aufmerksamkeits- und Konzentrationssituationen, die sich klar voneinander unterscheiden. Es besteht für sie auch praktisch kein Zusammenhang zwischen ‚aufmerksam sein‘ und ‚konzentriert sein‘.

In der deutschsprachigen wissenschaftlichen Psychologie wird neben dem Begriff ‚Aufmerksamkeit‘ auch der Begriff ‚Konzentration‘ verwendet. Es existieren allerdings bislang keine allgemeinverbindlichen Definitionen dieser beiden psychologischen Begriffe. Häufig werden sie als Bezeichnungen für Konzepte verwendet, die sich inhaltlich weitgehend überschneiden, manchmal sogar als Synonym für dasselbe Konzept (vgl. Fay & Stumpf, 1994; Schmidt-Atzert, Büttner und Bühner, 2004; Schmidt-Atzert, Bühner & Enders, 2006).

In der maßgebenden englischsprachigen psychologischen Forschung zu Aufmerksamkeitsleistungen findet fast ausschließlich der Begriff ‚attention‘ Verwendung. Um ihn näher zu spezifizieren, wird zwischen selective, divided und sustained attention

(selektive, geteilte und Daueraufmerksamkeit) unterschieden. In der vorliegenden Arbeit wird der Begriff ‚Konzentration‘ als Synonym für ‚selektive Aufmerksamkeit‘ verwendet.

2.2 Grundlegende Modelle der Aufmerksamkeit

Mitte des letzten Jahrhunderts kam mit der „kognitiven Revolution“ wieder verstärktes Interesse für die Erforschung der menschlichen Informationsverarbeitung auf (Johnson, Proctor & Vu, 2004). Zu dieser Zeit legte man den kognitiven Prozessen eine einfache Metapher zugrunde: Das Gehirn wurde mit einem Computer verglichen, der über einen Informationsspeicher und Verarbeitungskomponenten mit limitierter Kapazität verfügte. Die Struktur seiner Programme entsprach einfachen Flowcharts, die Schritt für Schritt durchlaufen werden mussten. Erst später kamen Modelle auf, die das Gehirn mit einem hoch vernetzten System gleichsetzten, dessen einzelne Komponenten vielfältig miteinander interagierten, und zwar nicht nur seriell sondern auch parallel. Allen Modellen ist die Annahme gemeinsam, dass die kognitive Informationsverarbeitung über eine begrenzte Kapazität verfügt und dass Aufmerksamkeitsleistungen der Selektion von Informationen bzw. kognitiven Prozessen dienen, um die kapazitätsbegrenzte Informationsverarbeitung vor Überlastung zu schützen (Siéoff, 2001; van der Heijden, 2001).

Die grundlegenden Modelle der Aufmerksamkeit werden in so genannte Reizselektionsmodelle und Kapazitätsmodelle der Aufmerksamkeit unterteilt. Beide Modellklassen sollen kurz dargestellt werden.

2.2.1 Reizselektionsmodelle

Reizselektionsmodelle basieren auf Experimenten zum dichiotischen Hören, die z. B. von Cherry (1953) und Moray (1959) durchgeführt wurden: Versuchspersonen wurden per Kopfhörer gleichzeitig zwei unterschiedliche Wortlisten dargeboten, für jedes Ohr eine. Die Aufgabe der Personen war, sich auf ein Ohr zu konzentrieren und die Wörter, die auf dieses Ohr trafen, durch begleitendes Nachsprechen mit einem „Schatten“ zu versehen (shadowing task). Die Wörter, die auf das andere Ohr trafen, sollten ignoriert werden. Die Ergebnisse dieser Versuche führten zu den ersten kognitiven Modellen der Aufmerksamkeit. Sie lassen

sich in zwei Gruppen aufteilen: die Modelle der frühen Selektion, d. h. Selektion vor Mustererkennung, und die Modelle der späten Selektion, d. h. Selektion nach Mustererkennung (Bundesen, 2001; Vecera & Luck, 2002).

2.2.1.1 Modelle der frühen Selektion

Das Filtermodell (filter model) von Broadbent (1958)

Die Untersuchungen zum dichotischen Hören zeigten, dass die Wörter des nicht beachteten Kanals von den Versuchspersonen nicht erinnert werden konnten (bzw. nicht gelernt wurden). Dieser Effekt zeigte sich sogar dann, wenn ein Wort bis zu 35-mal nacheinander wiederholt wurde (Moray, 1959). Wurde auf dem nicht beachteten Kanal die Sprache gewechselt oder wurden die Wörter rückwärts gelesen, war auch dies nicht erinnerbar. Die Veränderung von nicht inhaltsbezogenen physikalischen Merkmalen, wie der Wechsel weibliche/männliche Stimme oder große/geringe Lautstärke der Darbietung, konnte jedoch sehr wohl erinnert werden. (Dieses Ergebnis ist ein Indiz dafür, dass im Wahrnehmungsprozess die physikalischen Merkmale einer akustischen Information als erste analysiert werden.) Die Befunde führten zur Entwicklung des Filtermodells der Aufmerksamkeit. Ihm liegt die Annahme zugrunde, dass die Informationen aus den nicht beachteten Kanälen in einer frühen Phase der Informationsaufnahme durch einen selektiven Filter geblockt werden. Die Informationen aus dem ausgewählten Kanal hingegen werden zur Weiterverarbeitung in einen Verarbeitungskanal mit begrenzter Kapazität durchgelassen. Der Vorgang der Auswahl wird mit der Zuwendung von Aufmerksamkeit gleichgesetzt. Er ist nötig, um den kapazitätsbegrenzten Weiterverarbeitungskanal vor Überlastung zu schützen.

Das Filter-Dämpfungsmodell (filter-attenuation model) von Treisman (1960)

Trotz Ignorierung einer Botschaft ist es dennoch möglich, dass bestimmte Teile ihres Inhalts ins Bewusstsein vordringen können. Dieses Wahrnehmungsphänomen wird als Cocktailpartyeffekt bezeichnet, weil es bei der Konzentration auf ein Gespräch, z. B. bei einer Cocktailparty, möglich ist, z. B. den eigenen Namen wahrzunehmen, wenn er in einem anderen, nicht beachteten Gespräch genannt wird. Um dieses und ähnliche Phänomene erklären zu können, wurde die Filtertheorie zur Filter-Dämpfungstheorie weiterentwickelt:

Der selektive Filter wurde durch spezialisierte Filter-Dämpfungs-komponenten ersetzt, die zwar die Informationen aus der ausgewählten Quelle bevorzugt weiterleiten, jedoch auch anderen Informationen das Vordringen in den Weiterverarbeitungs-kanal ermöglichen. Der Informationsdämpfung wird hierzu eine semantische Analyse nachgeschaltet. Es wird angenommen, dass eingehende Informationen, die eine hohe subjektive Relevanz bzw. einen hohen subjektiven Bekanntheitsgrad besitzen, eher Inhalte im Langzeitgedächtnis aktivieren als Informationen, die einen geringen Bekanntheitsgrad besitzen. Übersteigt das Ausmaß an Aktivierung eine bestimmte Schwelle, so gelangt der Gedächtnisinhalt ins Bewusstsein.

2.2.1.2 Modelle der späten Selektion

Die Modelle der späten Selektion, z. B. von Deutsch und Deutsch (1963), gehen davon aus, dass alle über die Sinnesorgane eintreffenden Informationen Konzepte im Langzeitgedächtnis aktivieren, die mit ihnen verbundenen sind. (Es werden also alle eingehenden Informationen semantisch analysiert.) Die Aktivierung der Konzepte wird zusätzlich durch eine erfahrungs- bzw. erwartungsgesteuerte Beurteilung ihrer aktuellen Relevanz und Priming-Prozesse erhöht. (Priming bedeutet, dass ein Konzept durch die Aktivierung von mit ihm zusammenhängenden Konzepten in gewissem Maße selbst aktiviert wird.) Übersteigt die Gesamtaktivierung eines Konzepts eine bestimmte Schwelle, dann wird es für die Weiterverarbeitung zugänglich.

Der Unterschied zwischen den Modellen der frühen Selektion und den Modellen der späten Selektion sind die Merkmale der Informationen, die als Grundlage für die Auswahl der Stimuli zur Weiterverarbeitung dienen: Bei den Modellen der frühen Selektion basiert die Auswahl auf physikalischen Merkmalen, bei den Modellen der späten Selektion auf semantischen Merkmalen (Siéoroff, 2001). Treismans Modell steht zwischen beiden Modellen, da die Auswahl der Informationen sowohl auf physikalischen als auch semantischen Merkmalen beruht.

Eine Gegenüberstellung der Reizselektionsmodelle enthält Abbildung 1.1.

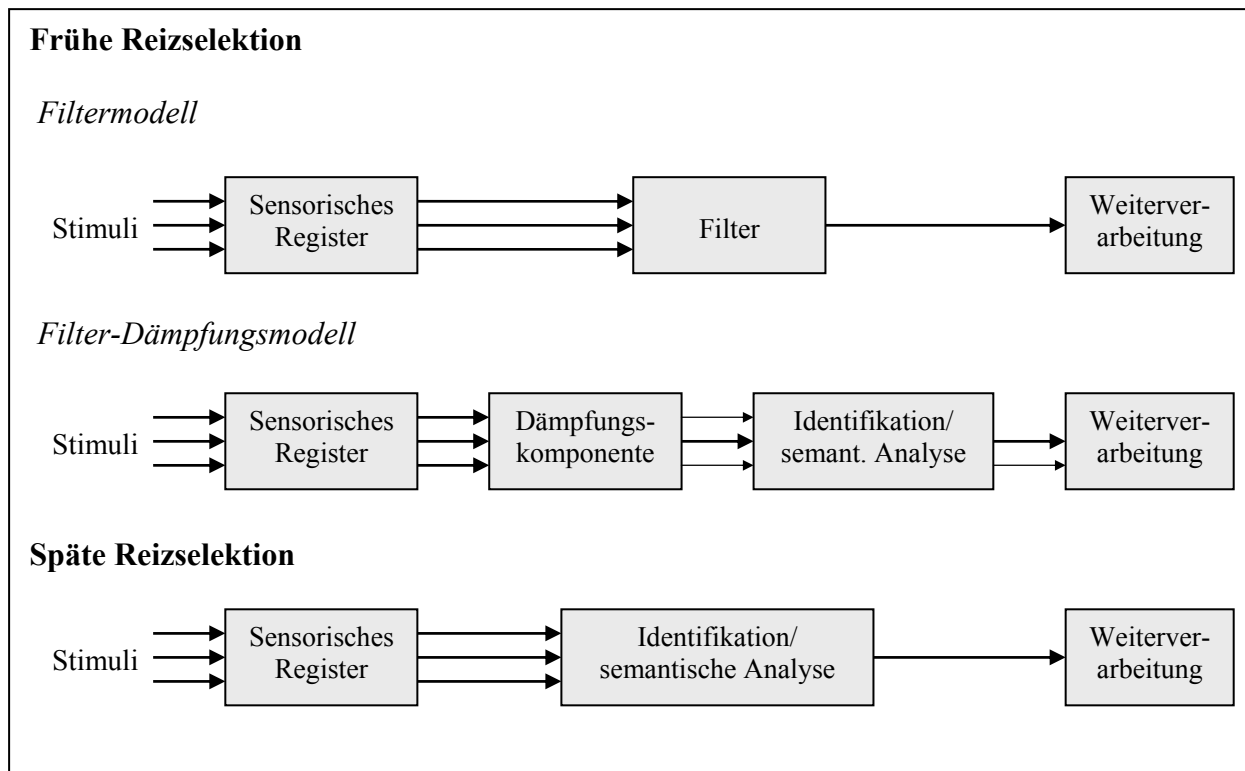


Abb. 1.1: Reizselektionsmodelle der Aufmerksamkeit

2.2.2 Kapazitätsmodelle der Aufmerksamkeit

Die Selektionsmodelle der Aufmerksamkeit konnten bestimmte Phänomene, wie den des Leistungsabfalls bei gleichzeitiger Ausführung zweier Aufgaben (geteilte Aufmerksamkeit) oder den der unbewussten/automatisierten Aufnahme und Weiterverarbeitung von Informationen, nicht hinreichend erklären. Um diesem Umstand Rechnung zu tragen, wurden so genannte Kapazitätsmodelle entwickelt, bei denen die Ressourcen der Informationsverarbeitung den anstehenden Aufgaben über Aufmerksamkeitsprozesse zugeteilt werden.

Die grundlegende Annahme der Kapazitätsmodelle der Aufmerksamkeit ist, dass die Ressourcen der bewussten kognitiven Informationsverarbeitung begrenzt sind. Folglich hängt die Anzahl der Aufgaben, die parallel bearbeitet werden können, davon ab, wie viele Ressourcen für ihre Bearbeitung benötigt werden. Die begrenzten Ressourcen werden über Aufmerksamkeitsprozesse den aktuell zu bearbeitenden Aufgaben flexibel zugeteilt.

Kapazitätsmodelle der Aufmerksamkeit werden mit Experimenten untersucht, bei denen eine Primäraufgabe und eine Sekundäraufgabe gleichzeitig zu bearbeiten sind: Die Schwierigkeit der Primäraufgabe ist variabel (z. B. einfache/schwierige Anagramme), die Schwierigkeit der Sekundäraufgabe bleibt gleich (z. B. einen Knopf drücken, wenn ein Licht erscheint). Wird die Schwierigkeit der Primäraufgabe erhöht, fällt die Leistung bei der Bearbeitung der Sekundäraufgabe ab (z. B. verlängerte Reaktionszeit beim Erscheinen des Lichts). Dieser Effekt wird als Hinweis dafür gesehen, dass die Bearbeitung der Primäraufgabe nun einen größeren Teil der begrenzten kognitiven Ressourcen erfordert.

2.2.2.1 Modell einer einheitlichen Aufmerksamkeitsressource

Das erste und bislang bekannteste Kapazitätsmodell stammt von Kahneman (1973), das Modell der einheitlichen Aufmerksamkeitsressource (*unitary-resource model of attention*). Aufmerksamkeit ist in seinem Modell eine kognitive Ressource mit begrenzter Kapazität, die einer Vielzahl von mentalen Aktivitäten zugeteilt werden kann. Sind die Anforderungen an die Aufmerksamkeitsressource so hoch, dass ihre Kapazität überstiegen wird, verschlechtert sich die erbrachte Leistung. Welchen mentalen Aktivitäten Aufmerksamkeit zugeteilt wird, bestimmt ein zentrales Zuteilungsverfahren (*allocation policy*). Das Zuteilungsverfahren unterliegt vier Einflussgrößen:

1. Dem Arousal

Nach dem Yerkes-Dodson-Gesetz (Yerkes & Dodson, 1908) ist bei der Bearbeitung von mittelschweren und schweren Aufgaben die Leistungsfähigkeit bei mittlerem Arousal maximal.

2. Überdauernden Dispositionen

Sie legen fest, welchen Informationen bevorzugt Ressourcen zugeteilt werden. Z. B. werden beim Hören des eigenen Namens sofort Ressourcen diesem Ereignis zugeteilt, da man i. d. R. anschließend reagieren muss. (Eine Erklärung des Cocktailpartyeffekts.)

3. Momentanen Intentionen

Bewusste strategische Entscheidungen darüber, was in diesem Moment wichtig ist und was nicht wichtig ist.

4. Einer Beurteilung darüber, wie viele Ressourcen zur Bearbeitung der einzelnen Aufgaben benötigt werden

Je schwieriger eine Aufgabe ist, desto mehr Ressourcen müssen eingesetzt werden, um sie zu bewältigen und desto weniger Ressourcen können anderen Aufgaben zugeteilt werden.

Das Modell von Kahneman lässt zu, dass bewusste kognitive Prozesse, die wenig Aufmerksamkeitskapazität beanspruchen, parallel ablaufen können und dass Prozesse, die durch Übung automatisiert wurden, keine Kapazität mehr beanspruchen. (Letzteres belegt, dass Aufmerksamkeit nicht mit wacher Bewusstheit gleichgesetzt werden darf.)

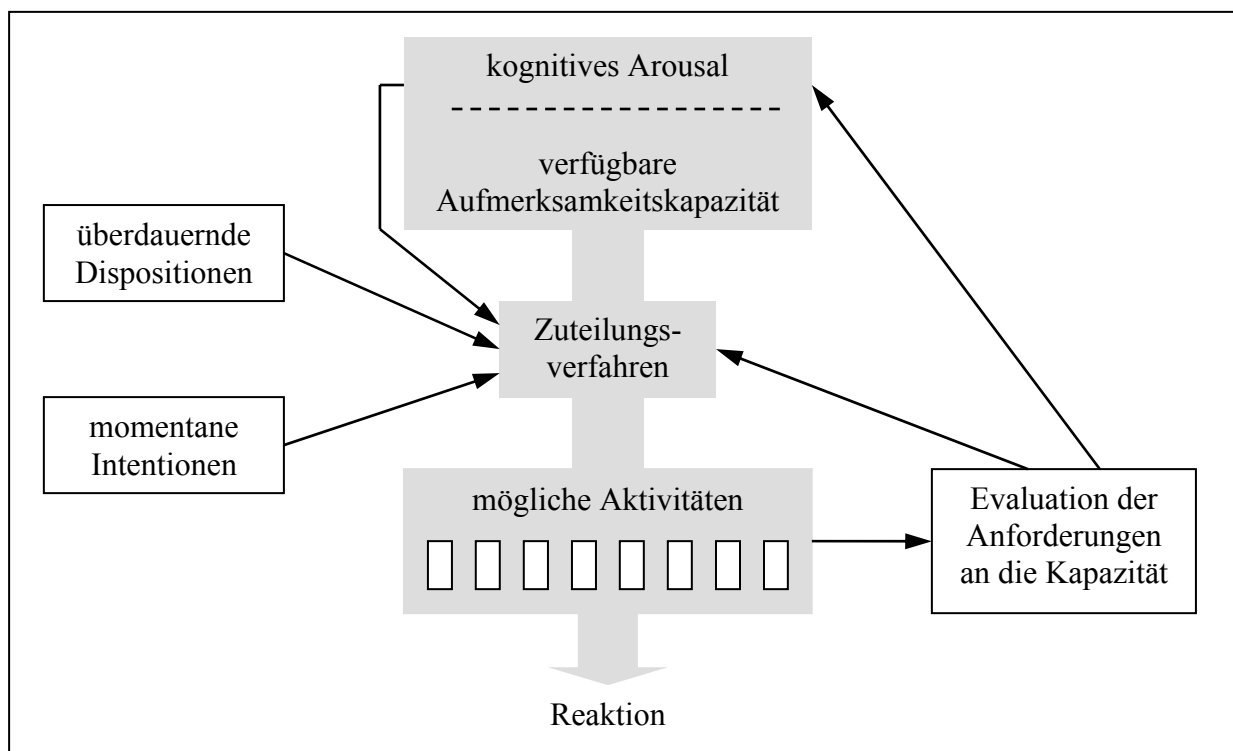


Abb. 1.2: Kapazitätsmodell der Aufmerksamkeit nach Kahneman

2.2.2.2 Modelle multipler Aufmerksamkeitsressourcen

Die Idee, dass nur eine einzige einheitliche Ressource für die Bewältigung aller möglichen Arten von Aufgaben existiert, konnte ihrerseits bestimmte kognitive Phänomene nicht erklären: Wird z. B. während der gleichzeitigen Bearbeitung von zwei verschiedenen Aufgaben eine dritte Aufgabe hinzugefügt, so verschlechtert sich nur bei derjenigen der ersten beiden Aufgaben die Bearbeitungsleistung, bei der Interferenzeffekte mit der dritten Aufgabe auftreten. Oder, werden zwei Aufgaben mit Stimuli unterschiedlicher Modalität aber gleicher Schwierigkeit bearbeitet, so verschlechtert sich die Bearbeitungsleistung, wenn die Stimulimodalität der einen Aufgabe, ohne ihre Schwierigkeit zu verändern, an die der anderen angeglichen wird. Derartige Ergebnisse legten die Vermutung nahe, dass die Bearbeitung verschiedenartiger Aufgaben oder Aufgabenbestandteile verschiedenartige kapazitätsbegrenzte kognitive Ressourcen in Anspruch nimmt, und führten zur Entwicklung von Modellen multipler Aufmerksamkeitsressourcen. Eines dieser Modelle wurde z. B. 1984 von Wickens vorgestellt (Johnson & Proctor, 2004).

3 Beschreibung des Inventars Komplexer Aufmerksamkeit

Das Inventar Komplexer Aufmerksamkeit folgt inhaltlich den Kapazitätsmodellen der Aufmerksamkeit. Es ist ein Test, der die selektive Aufmerksamkeit mit Hilfe von Aufgaben erfasst, bei denen visuelle Reize diskriminiert werden müssen. Die Komplexität der insgesamt 18 Aufgaben nimmt kontinuierlich zu und damit steigen auch die Aufgabenschwierigkeiten kontinuierlich an (siehe Teil 1, Kapitel 5.1).

Alle 18 Testaufgaben sollen in einem vorgegebenen Zeitraum von 15 Minuten nacheinander bearbeitet werden. Für die Bearbeitung der einzelnen Aufgaben gibt es keine zusätzlichen Zeitbegrenzungen. Die Gesamtbearbeitungszeit ist so gestaltet, dass eine Testperson, die über eine sehr hohe Aufmerksamkeitskapazität verfügt, alle Aufgaben fehlerfrei bearbeiten kann. Bei geringerer Aufmerksamkeitskapazität verlangsamt sich die Bearbeitungsgeschwindigkeit oder es steigt die Anzahl der Fehler (nicht entdeckte Zielreize). Beide Effekte führen zu einer Verringerung des erreichbaren Rohwerts, da nur für diejenigen Aufgaben ein Rohwertpunkt vergeben wird, die insgesamt fehlerfrei bearbeitet wurden. (Kombinationen aus verlangsamer Bearbeitungsgeschwindigkeit und gleichzeitiger hoher Anzahl an Fehlern konnten bei

Testauswertungen durch den Autor i. d. R. nicht festgestellt werden.) Im Gegensatz zu den gängigen Aufmerksamkeitstests, die als Speed-Tests konzipiert sind, ist INKA also ein Power-Test.

Jede Testaufgabe des INKA besteht aus einer Reihe zufällig angeordneter Konsonanten, der ‚Suchreihe‘. Die Testperson soll in der Reihe bestimmte einzelne Konsonanten und/oder Konsonantenpärchen suchen. Welche Konsonanten und/oder Konsonantenpärchen zu suchen sind, ergibt sich aus Buchstabenvorgaben, die mit Hilfe einer Umwandlungstabelle transformiert werden müssen. Die Vorgaben befinden sich neben den Reihen in der Spalte ‚Vorgaben‘. Die Umwandlungstabelle befindet sich oberhalb der Aufgaben.

Umwandlungstabelle:																				
B	C	D	F	G	H	J	K	L	M	N	P	Q	R	S	T	V	W	X	Y	Z
Z	R	K	G	L	T	P	W	H	B	M	Q	Y	N	F	D	C	S	J	X	V
Vorgaben	Suchreihen	Buchstaben																		
F, K	VFJCTQVXWMLKJDZRQGTRDSLJTRZGSYDKSBSTRJBCJKWTRBJLFCJ																			
B, TH	RFLPHZRKLMHGFDFVNVXGJDTHTVXFSQWZXYKLNHBGFRFDTHM																			

Abb. 1.3: Aufgabendesign der Rechtshänderversion des INKA

Zur Bearbeitung einer Aufgabe muss die Testperson also zunächst die Buchstabenvorgaben, die sich neben der entsprechenden Suchreihe befinden, mit Hilfe der Umwandlungstabelle transformieren.

Bei der ersten Aufgabe in Abbildung 1.3 wird durch Umwandeln aus der Vorgabe ‚F‘ ein ‚G‘ und aus dem ‚K‘ ein ‚W‘. Bei der zweiten Aufgabe wird aus dem ‚B‘ ein ‚Z‘ und aus der Buchstabenkombination ‚TH‘ wird die Buchstabenkombination ‚DT‘.

Die umgewandelten Vorgaben (Zielreize) müssen memoriert und anschließend in der Suchreihe gefunden werden. Die Suchreihen sind hierbei immer von links nach rechts zu bearbeiten.

Bei der ersten Aufgabe in Abbildung 1.3 muss sich die Testperson also den Buchstaben ‚G‘ und den Buchstaben ‚W‘ merken und versuchen, diese in der Suchreihe zu finden.

Wenn die Testperson einen Zielreiz in der Suchreihe gefunden hat (in den Aufgaben zur Demonstration fett hervorgehoben) muss sie das dadurch anzeigen, indem sie den Buchstaben notiert, der sich *unmittelbar links neben dem entdeckten Zielreiz* in der Suchreihe befindet.

Bei der Aufgabenbearbeitung ist entscheidend, dass die Lösungsbuchstaben in der genauen Abfolge ihrer Anordnung in der Suchreihe notiert werden. Zum Notieren steht neben der Suchreihe in der Spalte ‚Buchstaben‘ hierfür Platz zur Verfügung. Nur eine vollständige Lösungsbuchstabenreihe wird als richtige Lösung einer Aufgabe gewertet.

Für die erste Aufgabe in Abbildung 1.3 lautet die korrekte Lösungsbuchstabenreihe ‚XQZK‘; für die zweite Aufgabe lautet sie ‚HGJWF‘.

Ein falsch notierter Lösungsbuchstabe kann von den Testpersonen als ungültig gekennzeichnet werden, indem er z. B. durchgestrichen wird. Bis auf die Buchstabeneinträge darf die Testperson bei der Aufgabenbearbeitung nichts auf den Testbogen schreiben. Sie darf auch keine (Hilfs-)Markierungen in den Suchreihen anbringen.

4 Entwicklung der Testaufgaben

Bei der Entwicklung der Testaufgaben wurde durch die Verwendung von visuellen Diskriminierungsaufgaben der Anforderung Rechnung getragen, dass individuell unterschiedlich ausgeprägte Fertigkeiten möglichst keinen Einfluss auf die

Bearbeitungsleistungen haben sollen. Bei Rechenverfahren zur Messung der Konzentrationsfähigkeit, um nur ein Beispiel zu nennen, kann der individuelle Übungsgrad bei der Ausführung mathematischer Operationen die Bearbeitungsleistung systematisch beeinflussen.

Die Notwendigkeit jede Suchreihe vollständig und fehlerfrei bearbeiten zu müssen, um einen Rohwertpunkt zu erzielen, und die Gestaltung der Lösungsbuchstabenreihen schließen aus, dass Testpersonen durch zufälliges Notieren von Buchstaben zu richtigen Lösungen kommen können, also eine „Konzentrationsleistung ohne Konzentration“ erbringen. Hierzu zwei Erläuterungen:

Die Aufgabenlösungen bestehen aus Reihen von 6 bzw. 7 Buchstaben. Eine Regelmäßigkeit ihrer Länge wurde vermieden, damit die Testpersonen keinen Hinweis auf die Güte ihrer Aufgabenbearbeitung erhalten. Lösungsbuchstabenreihen, die bei jeder Aufgabe dieselbe Länge besitzen, geben einen direkten Hinweis auf einen Fehler bei der Bearbeitung, wenn eine ermittelte Reihe länger oder kürzer ist als die vorherigen.

Falls eine Testperson wüsste, wie viele Lösungsbuchstaben (Konsonanten) eine Aufgabe enthält, betrüge für sie die Wahrscheinlichkeit, bei einer Reihe von 7 Buchstaben durch Zufall die richtige zu notieren $1/180.088.541$ ($1/21^7$), bei einer Reihe von 6 Buchstaben $1/85.766.121$ ($1/21^6$). Da die Testperson überdies nicht wissen kann, wie viele Buchstaben eine Lösung enthält, verringert sich in der Praxis die Wahrscheinlichkeit durch Raten zur richtigen Lösung zu kommen noch erheblich.

4.1 Generative Regeln

Der Aufgabenkonstruktion lagen hypothesengeleitet generative Regeln zugrunde. Generative Regeln ermöglichen es, auf ökonomische Weise einen beliebig großen Aufgabenpool mit Aufgaben unterschiedlicher Schwierigkeit zu erstellen. Die Regeln sind in Abbildung 1.4 dargestellt.

1. In den Aufgaben finden nur Konsonanten Verwendung.
2. Die Suchreihen bestehen aus einer zufälligen Kombination von Buchstaben.
3. Die Vorgaben bestehen aus einem, zwei oder drei Elementen.
4. Die Elemente sind einzelne Buchstaben oder Buchstabenpärchen.
5. Zur Transformation der Vorgaben wird eine Umwandlungstabelle vorgegeben.

Abb. 1.4: Generative Regeln zur Konstruktion der Aufgaben des INKA

4.2 Linkshänderversion

In Voruntersuchungen wurde festgestellt, dass sich für Linkshänder, die die Urversion des Tests (die Rechtshänderversion) bearbeiteten, das Eintragen der Lösungsbuchstaben umständlich gestaltete. Dieser Sachverhalt führte zur Entwicklung eines speziellen grafischen Aufgabendesigns für Linkshänder, um eine Benachteiligung bei der Testbearbeitung gegenüber Rechtshändern auszuschließen: Die Aufgaben und ihre Reihenfolge wurden von der Rechtshänderversion übernommen. Die Spalten ‚Vorgaben‘ und ‚Buchstaben‘ wurden vertauscht, so dass Linkshänder die Lösungsbuchstaben auf der für sie ergonomisch günstigeren linken Seite des Testbogens eintragen können. Zwei Beispiele für das Aufgabendesign der Linkshänderversion des INKA zeigt Abbildung 1.5.

Buchstaben	Suchreihen	Vorgaben
	VFJCTQXMLKJDZRQTRNPKLWJTRZSYXDKSBSTRJBCJKXTRBLFCJ	<i>K, B</i>
	RFLPHRKLMHGDFVZNXGJHTVXGFSQWXYKLMZJNBHTGFRFHM	<i>F, HT</i>

Abb. 1.5: Aufgabendesign der Linkshänderversion des INKA

Die geringfügige Veränderung des Aufgabendesigns zog eine geringfügige Veränderung der Instruktion auf der ersten Seite des Linkshänder-Testbogens nach sich.

Die Linkshänderversion wird von Linkshändern im Rahmen der Testungen gerne genutzt. Die mittleren Gesamtestleistungen unterscheiden sich bei Rechts- und Linkshänderversion nicht. Auch weisen die Schwierigkeiten der einzelnen Aufgaben beider Versionen keine signifikanten Unterschiede auf (siehe Teil 1, Kapitel 6.4.5).

5 Konstruktion der Skala

Die Erkenntnisse zur optimalen Aufgabengestaltung, die in Voruntersuchungen mit Testvorformen gewonnenen wurden, bildeten die Grundlage für die Formulierung der in Abbildung 1.4 dargestellten generativen Regeln zur Konstruktion der Testaufgaben. Mit Hilfe der generativen Regeln wurden 18 Aufgaben entwickelt und zu einem Test zusammengestellt. Die Zielsetzung war, ein gut differenzierendes Verfahren zur Messung der Konzentrationsfähigkeit zu erstellen. Durch die systematische Steigerung der Komplexität der Buchstabenvorgaben wurden die Aufgabenschwierigkeiten systematisch erhöht. Die ersten Aufgaben des Tests wurden bewusst sehr einfach gestaltet. Sie dienen in erster Linie zum Üben und Eingewöhnen in die Aufgabenstellung, um möglicherweise bestehende unterschiedliche Ausgangspositionen zwischen den Testpersonen anzugleichen.

Der 18 Aufgaben enthaltende Test wurde zunächst 350 Testpersonen vorgelegt, um Aufgabenanalysen nach der Klassischen Testtheorie und nach der Item Response Theory (Rasch-Modell) durchzuführen.

5.1 Aufgabenanalyse nach der Klassischen Testtheorie

Bei der Entwicklung von Tests nach der Klassischen Testtheorie werden für die Aufgabenanalyse vor allem zwei Kennwerte berechnet, die Schwierigkeit und die Trennschärfe der Testaufgaben (Lienert & Raatz, 1994).

Die Aufgabenschwierigkeit (p) ist definiert als der prozentuale Anteil der richtigen Antworten in einer Analysestichprobe der Größe n . Eine Aufgabe mit einer Schwierigkeit von $p = 50\%$ differenziert optimal zwischen den Testpersonen der Stichprobe und hat die besten Voraussetzungen für eine hohe Trennschärfe. Für einen Power-Test ist es jedoch wichtig, dass seine Aufgaben in ihrer Schwierigkeit variieren. Nur so kann er gut zwischen den Personen der Zielpopulation differenzieren (Krauth, 1995): Aufgaben mit hoher Schwierigkeit differenzieren besser zwischen Personen mit hoher Merkmalsausprägung. Aufgaben mit mittlerer Schwierigkeit differenzieren besser zwischen Personen mit mittlerer Merkmalsausprägung. Aufgaben mit niedriger Schwierigkeit differenzieren besser zwischen Personen mit niedriger Merkmalsausprägung.

Die Aufgabentrennschärfe (r_{it}) gibt einen Hinweis darauf, ob die Aufgaben tatsächlich etwas Ähnliches erfassen wie der Gesamttest. Hierzu wird der korrelative Zusammenhang zwischen den Ergebnissen der Testpersonen in Aufgabe i und ihren Gesamttestwerten berechnet. Weil im Gesamttestwert jedoch die Aufgabe i enthalten ist, ist dieser Korrelationskoeffizient künstlich erhöht. Deshalb wird er üblicherweise part-whole-korrigiert berechnet: Der Gesamttestwert wird für die Korrelation aus allen Aufgaben ohne Aufgabe i berechnet. Aufgaben mit Trennschärfen von $r_{it} \leq .25$ werden im Allgemeinen für die Weiterverwendung als wenig geeignet angesehen. Die Frage, welche Trennschärfekoeffizienten als ausreichend hoch angesehen werden können, sollte aber immer auch im Zusammenhang mit der Art des Tests und den Indizes der restlichen Aufgaben beantwortet werden (Moosbrugger & Hartig, 2003). Bei einem Power-Test, der Aufgaben unterschiedlicher Schwierigkeit enthalten sollte, können die Trennschärfekoeffizienten nicht auf einem gemeinsamen hohen Niveau liegen. Hohe Trennschärfen können nur erreicht werden, wenn die Aufgaben weder allzu leicht noch allzu schwierig sind (Moosbrugger, 1992).

Tabelle 1.2 enthält die Schwierigkeitskoeffizienten und die part-whole-korrigierten Trennschärfekoeffizienten der Aufgaben des INKA. Die Ergebnisse wurden an einer Stichprobe von 350 Testpersonen gewonnen, die den Test im Rahmen einer Testbatterie bearbeiteten, die von einer hessischen Großstadt zur Auswahl von Bewerbern für Ausbildungsplätze aller Ausbildungsrichtungen durchgeführt wurde. Den soziodemografischen Aufbau der Stichprobe gibt Tabelle 1.1 wieder:

	männlich	weiblich	gesamt
Geschlecht	165	185	350
Schulbildung			
Hauptschule	10	18	28
mittlerer Bildungsweg	47	82	129
Abitur, Fachhochschulreife	108	85	193
Alter (in Jahren)	AM = 22,2 s = 6,4 Min.: 15 Max.: 39	AM = 17,7 s = 2,0 Min.: 14 Max.: 25	AM = 19,8 s = 5,1 Min.: 14 Max.: 39

Tab. 1.1: Soziodemografische Daten der Stichprobe zur Analyse der Aufgaben des INKA

Aufgabe	Schwierigkeit	Trennschärfe	Aufgabe	Schwierigkeit	Trennschärfe
1	.90	.23	10	.46	.44
2	.72	.34	11	.51	.53
3	.80	.45	12	.46	.49
4	.59	.42	13	.41	.51
5	.42	.37	14	.31	.37
6	.63	.48	15	.16	.37
7	.65	.49	16	.08	.30
8	.45	.36	17	.07	.21
9	.65	.49	18	.02	.13

Tab. 1.2: Ergebnis der Analyse der Aufgaben des INKA nach der Klassischen Testtheorie

Die geringen Schwierigkeiten der ersten drei Aufgaben waren gewollt und vorhersehbar. Die Schwierigkeiten der letzten vier Aufgaben wurden in dieser Höhe nicht erwartet. Folgende Gründe sprachen jedoch dafür, die Aufgaben nach der Aufgabenanalyse nicht aus dem Test auszuschließen:

1. die Vermeidung von Deckeneffekten bei sehr leistungsstarken Testpersonen,
2. die, abgesehen von den letzten beiden Aufgaben, akzeptablen Trennschärfekoeffizienten und vor allem
3. die Ergebnisse der Aufgabenanalysen nach dem Modell von Rasch, die im Folgenden vorgestellt werden.

5.2 Aufgabenanalyse nach der Probabilistischen Testtheorie

Die ersten 350 Datensätze, die mit dem INKA gewonnen wurden, dienten auch zur Überprüfung, ob das logistische 1-Parameter Modell der Probabilistischen Testtheorie (Rasch-Modell) für die Aufgaben des Tests Gültigkeit besitzt. Die Überprüfung erfolgte an der Universität Wien, Institut für Psychologie, mit dem Rechenprogramm MTEST (Dimitz & Formann, 1979).

Zunächst sollen kurz die Grundzüge des Rasch-Modells dargestellt werden, um zu verdeutlichen, welche Eigenschaften ein Rasch-skaliertes Test besitzt.

5.2.1. Das Rasch-Modell

Das Rasch-Modell ist ein probabilistisches Testmodell. Es ist eindimensional, d. h. es bezieht sich auf die Messung einer einzelnen Persönlichkeitseigenschaft, und zweikategorial, d. h. es gilt für Aufgaben, die ein dichotomes Antwortformat besitzen. Das Modell baut auf der logistischen Funktion auf, die hier den Zusammenhang zwischen dem beobachtbaren Verhalten (das Lösen einer Aufgabe) und der interessierenden Persönlichkeitsdimension (in Abhängigkeit von der korrespondierenden Aufgabeneigenschaft) beschreibt. Im Gegensatz zur Klassischen Testtheorie (KTT) wird der gemessene Wert nicht einfach mit dem wahren Wert (plus einem variablen Messfehlerwert) gleichgesetzt sondern es wird die Existenz einer nicht beobachtbaren latenten Persönlichkeitsdimension postuliert. Die Beobachtung der Indikatoren bzw. Symptome dieser latenten Dimension besitzt nach dem Modell von Rasch einen probabilistischen Charakter.

$$P(\theta, I|\beta, \delta) = \frac{e^{\beta - \delta}}{1 + e^{\beta - \delta}}$$

$P(\theta, I|\beta, \delta)$ - Wahrscheinlichkeit der Lösung einer Aufgabe mit der Schwierigkeit δ für eine Person mit der Fähigkeit β

Abb. 1.6: Logistische Funktion des Testmodells von Rasch

Gemäß der logistischen Funktion hängt für eine Testperson die Wahrscheinlichkeit des Lösens einer bestimmten Aufgabe nur von ihrer Fähigkeit und von der Schwierigkeit der Aufgabe ab.

Ist ein psychodiagnostisches Verfahren Rasch-skaliert, so zeichnet es sich durch folgende Eigenschaften aus (Fischer, 1974; Kubinger, 1988; Kubinger, 1995; Rost, 1996):

1. Seine Aufgaben sind *homogen*.

Homogenität bedeutet, dass alle Aufgaben auf dem Aufgabenparameter-Personenparameter-Kontinuum, das die zu messende latente Persönlichkeitsdimension repräsentiert, lokalisiert sind. Sie messen folglich alle ein und dieselbe Eigenschaft und unterscheiden sich nur in ihrer Schwierigkeit. Ihre Homogenität ist der Indikator für die gegebene Messgenauigkeit des Tests, also für das, was im Rahmen der Klassischen Testtheorie mit der Reliabilität beschrieben wird.

2. Seine Aufgaben sind voneinander *lokal stochastisch unabhängig*.

Lokale stochastische Unabhängigkeit bedeutet, dass die Kovariation zwischen den beobachtbaren Verhaltensweisen bei der Aufgabenbearbeitung (Lösung, keine Lösung) nur durch die ihnen gemeinsam zugrunde liegende latente Persönlichkeitsdimension begründet ist. Die Wahrscheinlichkeit, eine bestimmte Aufgabe zu lösen, hängt also nicht von der Wahrscheinlichkeit ab, eine andere Aufgabe zu lösen, sondern nur von der Fähigkeit der

Testperson und der Schwierigkeit der Aufgabe. Der Test liefert daher eine *erschöpfende Statistik*. Die Anzahl der gelösten Testaufgaben charakterisiert erschöpfend die Ausprägung der zu messenden latenten Dimension.

3. Der Test ist *spezifisch objektiv*.

Spezifische Objektivität bedeutet, dass ein Rasch-skaliertes Test stichprobenunabhängig ist. Der Vergleich von Aufgaben bezüglich ihrer Schwierigkeiten ist unabhängig davon, welche Personenstichprobe dafür verwendet wird. Sind innerhalb der Population, für die die Gültigkeit des Rasch-Modells für den Test festgestellt wurde, teilstichprobenspezifische Normierungen angezeigt, dann sind die Schwierigkeitsrelationen zwischen sämtlichen Items in allen Teilstichproben gleich. (Falls das Rasch-Modell für einen Test Gültigkeit besitzt, so bedeutet das also nicht, dass keine teilstichprobenspezifische Normierungen vorgenommen werden müssen.) Im Rahmen der Klassischen Testtheorie sind Aussagen über Testpersonen immer bezogen auf Aufgaben und deren Lösungshäufigkeit in der zugrunde liegenden Normstichprobe. Eine Aufgabe hat also keine Schwierigkeit an sich, sondern immer nur eine bestimmte Schwierigkeit in einer bestimmten Stichprobe. Im Rasch-Modell sind Personen- und Aufgabenparameter als getrennte und als trennbare Größen konzipiert.

5.2.2 Überprüfung von Testaufgaben auf Gültigkeit des Rasch-Modells

Legt man einem diagnostischen Verfahren die Axiome der Klassischen Testtheorie zugrunde, so ist empirisch nicht überprüfbar, ob das Modell für die Aufgaben des Verfahrens tatsächlich Gültigkeit besitzt. Die Gültigkeit des probabilistischen Testmodells von Rasch hingegen lässt sich empirisch überprüfen. Hierzu dienen so genannte Modelltests.

Die wesentliche Grundlage von Modelltests ist die spezifische Objektivität. Bei Gültigkeit des Modells müssen die Schätzungen der Parameter in verschiedenen Teilstichproben statistisch gleich sein. So dürfen sich die Personenparameterschätzungen in verschiedenen Aufgabenteilstichproben nicht systematisch unterscheiden. Umgekehrt müssen sich in verschiedenen Personenteilstichproben gleiche Schätzungen der Aufgabenparameter ergeben. Gängige Modelltests sind grafische Modelltests und Likelihoodquotiententests.

Grafische Modelltests

Wie bereits beschrieben, ist eine zentrale Annahme des Rasch-Modells die Homogenität von Personen und Aufgaben. Gehören alle getesteten Personen einer homogenen Population an und sind auch die Aufgaben homogen, dann ergeben sich in verschiedenen Personenteilstichproben (abgesehen von Zufallsschwankungen) dieselben Schätzwerte für die Aufgabenparameter.

Um das Ausmaß von Abweichungen auf einen ersten Blick zu überprüfen, dienen grafische Kontrollen. Hierzu wird die Gesamtstichprobe nach einem Kriterium, bei dem man annimmt, dass es einen Einfluss auf die Homogenität der Aufgaben haben könnte, in zwei Gruppen aufgeteilt. Als externe Teilungskriterien können z. B. das Geschlecht oder das Alter dienen, als internes Teilungskriterium der Testscore, i. d. R. hoher vs. niedriger Score. Mit jeder der beiden Teilgruppen werden anschließend Schätzungen der Aufgabenparameter vorgenommen und diese werden dann in einem zweidimensionalen Koordinatensystem abgetragen. Bei Modellkonformität der Aufgaben entsprechen sich die Parameter der einzelnen Aufgaben und die Punkte liegen im Koordinatensystem um eine Gerade, die durch den Ursprung geht und die Steigung 1 hat.

Likelihoodquotiententests

Zur statistischen Signifikanzprüfung werden Likelihoodquotiententests verwendet. Mit ihrer Hilfe wird überprüft, ob die beobachteten Daten durch die in verschiedenen Teilstichproben geschätzten Aufgabenparameter besser vorhergesagt werden können als durch die in der Gesamtstichprobe geschätzten Aufgabenparameter. Im Gegensatz zu Grafischen Modelltests ermöglichen sie nur eine globale Überprüfung des Aufgabenpools.

Likelihoodquotiententests basieren auf folgendem allgemeinen Prinzip: Zuerst wird die Wahrscheinlichkeit der gewonnenen Daten (unter Zugrundelegung der Parameterschätzungen) für das Modell, dessen Gültigkeit getestet werden soll (die Gesamtstichprobe), berechnet. Dann berechnet man die Wahrscheinlichkeit der gewonnenen Daten für Personenteilstichproben, die durch Teilungskriterien gewonnen werden, für die angenommen wird, dass sie einen Einfluss auf die Homogenität der Aufgaben haben könnten. Mit dem Chi-Quadrat-Test wird abschließend geprüft, ob Unterschiede zwischen den ermittelten

Wahrscheinlichkeiten statistische Signifikanz besitzen. Ist der Chi-Quadrat-Wert nicht signifikant, so können die errechneten Personen- und Aufgabenparameter als angemessene Schätzungen für Personenfähigkeit und Aufgabenschwierigkeit gelten.

5.2.3 Ergebnisse der Überprüfung der Aufgaben des INKA

Zur Überprüfung der Gültigkeit des Rasch-Modells für die Aufgaben des INKA wurden grafische Modelltests und Likelihoodquotiententests durchgeführt. Die Kriterien für die Aufteilungen der Stichprobe waren das Geschlecht, das Alter, das Schulbildungsniveau und die Testleistung. Die durchgeführten Modelltests ergaben sämtlich, dass alle 18 Aufgaben des INKA Rasch-skaliert sind.

5.2.3.1 Teilungskriterium Geschlecht

Aufgabenparameterschätzungen für die Teilstichproben weiblicher Personen (Gruppe A, n = 185) und männlicher Personen (Gruppe B, n = 165).

Aufgabe	Gruppe A	Gruppe B	Aufgabe	Gruppe A	Gruppe B
1	3,18	2,82	10	,09	,08
2	1,51	1,55	11	,46	,29
3	2,02	2,48	12	,35	-,04
4	,56	,94	13	-,06	-,21
5	,10	-,31	14	-,55	-,65
6	,98	1,18	15	-1,58	-1,96
7	,88	1,46	16	-2,80	-2,32
8	,25	-,10	17	-2,73	-2,33
9	1,11	1,35	18	-3,77	-4,22

Tab. 1.3: Logarithmen der produktnormierten Aufgabenparameterschätzungen auf der Basis der Testdaten der Personen der Gruppen A und B

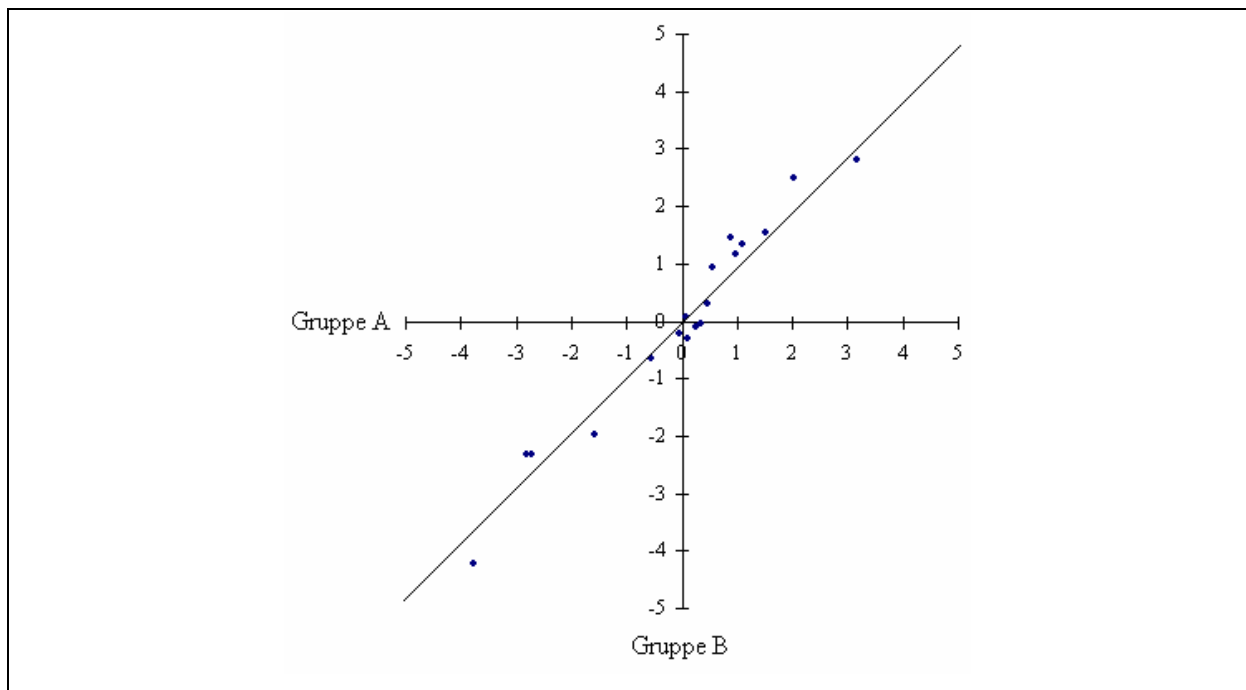


Abb. 1.7: Grafischer Modelltest, Teilungskriterium Geschlecht

Die Chi-Quadrat-verteilte Testgröße des Likelihoodquotiententests beträgt 23,20 (df = 17). Dieser Wert ist nicht signifikant.

5.2.3.2 Teilungskriterium Alter

Aufgabenparameterschätzungen für die Teilstichproben Personen älter als 20 Jahre (Gruppe C, n = 82) und Personen bis 20 Jahre (Gruppe D, n = 268).

Aufgabe	Gruppe C	Gruppe D	Aufgabe	Gruppe C	Gruppe D
1	2,90	3,04	10	,46	-,04
2	1,26	1,59	11	,20	,42
3	3,24	2,04	12	,00	,21
4	,63	,76	13	-,33	-,09
5	-,45	,01	14	-,82	-,53
6	1,17	1,03	15	-2,15	-1,63
7	1,38	1,08	16	-2,27	-2,69
8	-,10	,13	17	-2,35	-2,59
9	1,03	1,27	18	-3,80	-4,03

Tab. 1.4: Logarithmen der produktnormierten Aufgabenparameterschätzungen auf der Basis der Testdaten der Personen der Gruppen C und D

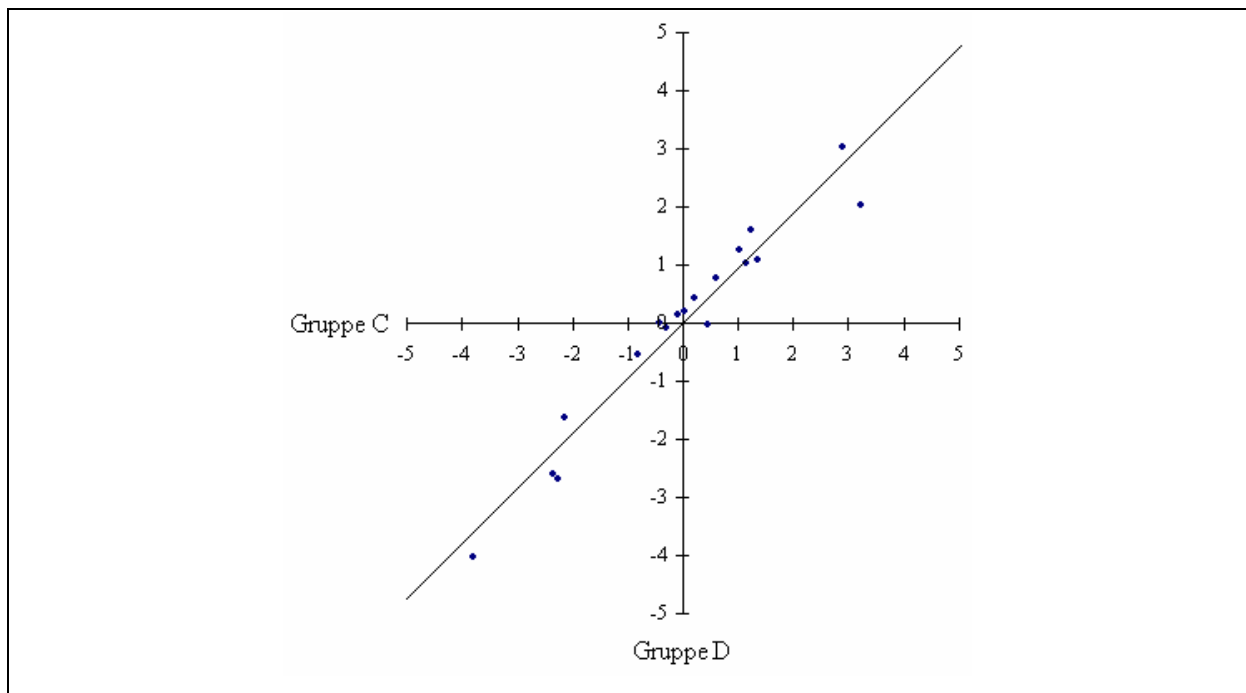


Abb. 1.8: Grafischer Modelltest, Teilungskriterium Alter

Die Chi-Quadrat-verteilte Testgröße des Likelihoodquotiententests beträgt 20,04 (df = 17). Dieser Wert ist nicht signifikant.

5.2.3.3 Teilungskriterium Schulbildungsniveau

Aufgabenparameterschätzungen für die Teilstichproben Hauptschulabschluss oder mittlerer Bildungsweg (Gruppe E, n = 157) und Abitur, Fachhochschulreife oder Hochschulabschluss (Gruppe F, n = 193).

Aufgabe	Gruppe E	Gruppe F	Aufgabe	Gruppe E	Gruppe F
1	2,88	3,18	10	,04	,18
2	1,34	1,79	11	,19	,68
3	2,26	2,21	12	,28	,02
4	,63	,92	13	-,20	,00
5	-,20	,09	14	-,60	-,56
6	1,02	1,18	15	-1,62	-2,07
7	1,24	1,07	16	-2,36	-3,00
8	,03	,19	17	-2,30	-3,01
9	1,13	1,39	18	-3,77	-4,26

Tab. 1.5: Logarithmen der produktnormierten Aufgabenparameterschätzungen auf der Basis der Testdaten der Personen der Gruppen E und F

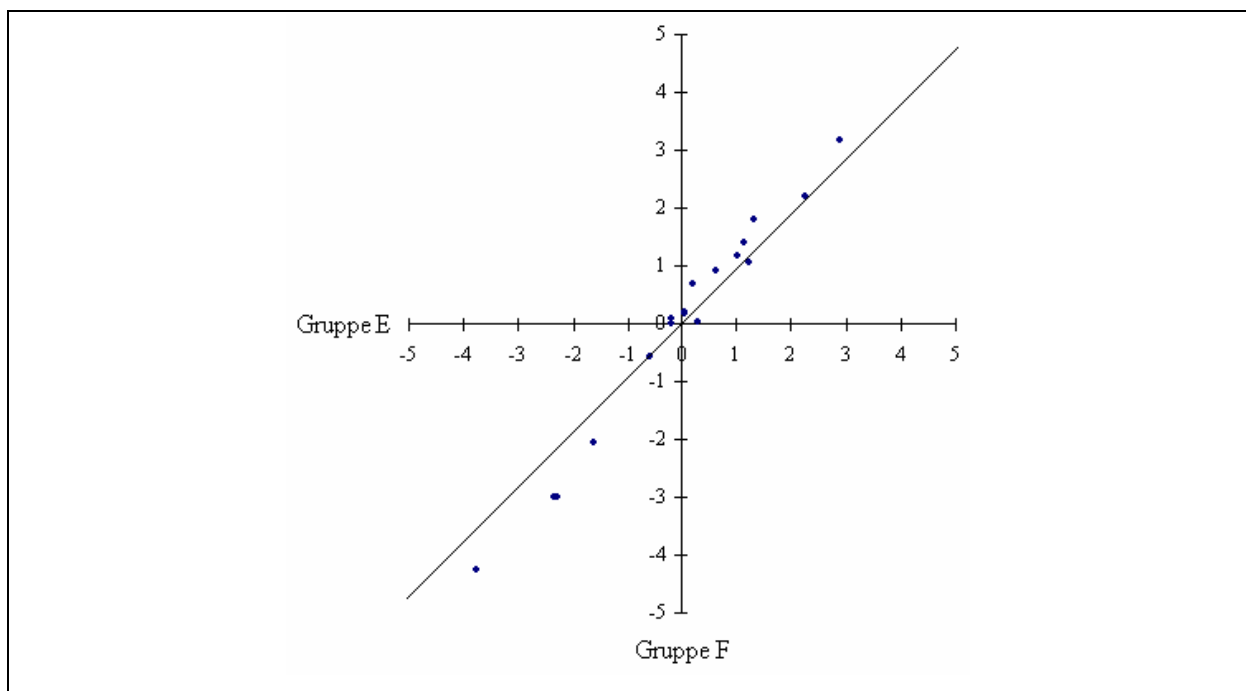


Abb. 1.9: Grafischer Modelltest, Teilungskriterium Schulbildungsniveau

Die Chi-Quadrat-verteilte Testgröße des Likelihoodquotiententests beträgt 19,11 (df = 17). Dieser Wert ist nicht signifikant.

5.2.3.4 Teilungskriterium Testleistung

Aufgabenparameterschätzungen für die Teilstichproben Personen mit niedrigem Rohwert (bis 9, Gruppe G, n = 212) und Personen mit hohem Rohwert (über 9, Gruppe H, n = 138).

Aufgabe	Gruppe G	Gruppe H	Aufgabe	Gruppe G	Gruppe H
1	2,90	3,83	10	,10	,05
2	1,67	1,17	11	,28	,55
3	2,41	1,70	12	,11	,24
4	,80	,61	13	-,46	,28
5	-,01	-,24	14	-,59	-,61
6	1,08	1,03	15	-1,94	-1,66
7	1,13	1,15	16	-2,85	-2,41
8	,24	-,17	17	-2,28	-2,68
9	1,23	1,22	18	-3,82	-4,07

Tab. 1.6: Logarithmen der produktnormierten Aufgabenparameterschätzungen auf der Basis der Testdaten der Personen der Gruppen G und H

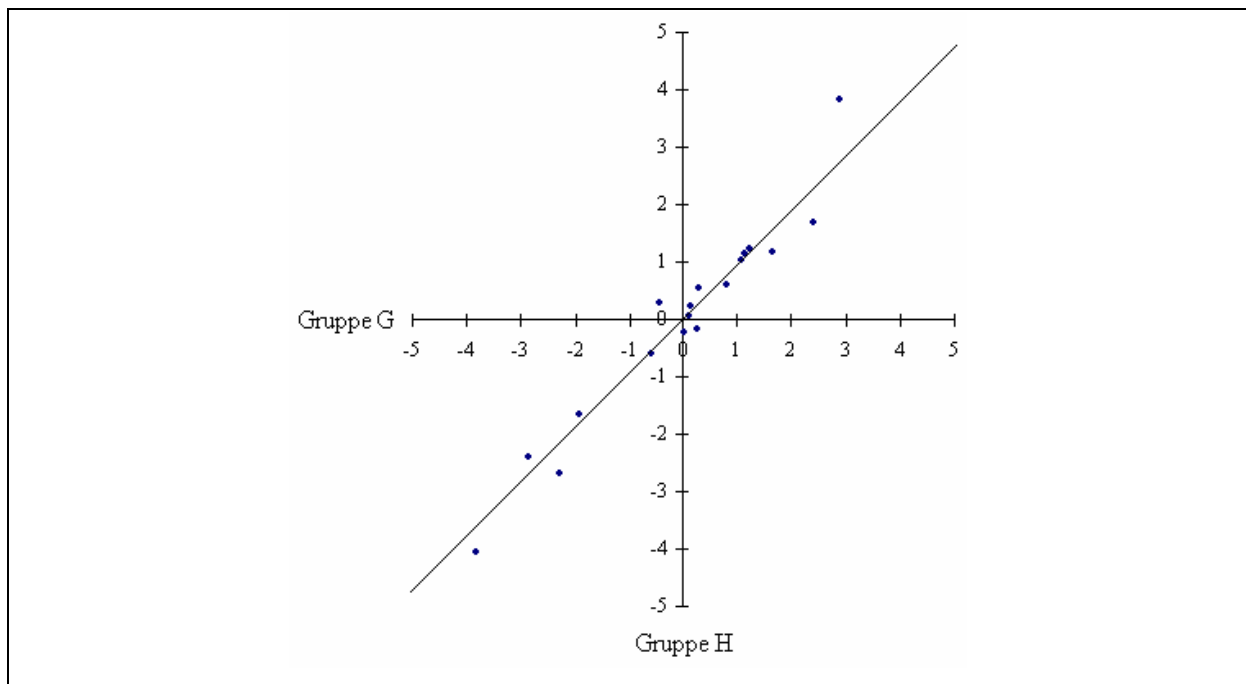


Abb. 1.10: Grafischer Modelltest, Teilungskriterium Testleistung

Die Chi-Quadrat-verteilte Testgröße des Likelihoodquotiententests beträgt 22,64 (df = 17). Dieser Wert ist nicht signifikant.

Im Allgemeinen werden bei der Konstruktion eines Tests nach dem Modell von Rasch aus einem relativ großen Itempool sukzessive diejenigen Aufgaben ausgeschieden, die den Modellvoraussetzungen nicht entsprechen. Die Modelltests werden nach jeder Testverkürzung wiederholt. Für den letztlich verbleibenden Itempool muss dann die Modellgültigkeit noch einmal mit Hilfe einer neuen und unabhängigen Stichprobe geprüft werden. Es werden also im Regelfall A-posteriori-Anpassungen des Tests an das Modell vorgenommen (Kubinger, 2003b). Bei der Überprüfung der 18 Aufgaben der ersten Version des INKA ergab sich bereits, dass alle Aufgaben die Modellvoraussetzungen erfüllen. Der Test konnte also in seiner ursprünglichen Form beibehalten werden.

6 Gütekriterien

In Anlehnung an Kubinger (1996) und das Testkuratorium (1986) sollen folgende Gütekriterien für das Inventar Komplexer Aufmerksamkeit überprüft werden: Objektivität, Reliabilität, Validität, Normierung, Skalierung, Unverfälschbarkeit, Zumutbarkeit und Fairness.

6.1 Objektivität

Das Inventar Komplexer Aufmerksamkeit ist ein voll standardisiertes Verfahren. Sowohl die Durchführung des Tests als auch die Auswertung und Interpretation der Testergebnisse sind genau festgelegt und für den Testleiter in einer Handanweisung beschrieben. Befolgt der Testleiter die Vorschriften der Handanweisung, so kann das Testergebnis als unabhängig von seinem Verhalten bezeichnet werden.

6.2 Reliabilität

Im Rahmen der Klassischen Testtheorie beschreibt die Reliabilität den Grad an Genauigkeit, mit dem ein Test ein personengebundenes Merkmal misst. Der Reliabilitätskoeffizient gibt an, in welchem Maße Messwerte übereinstimmen, die unter gleichen Bedingungen an ein und derselben Person gewonnen wurden (Lienert & Ratz, 1994). Zu seiner Bestimmung werden

entweder Paralleltests eingesetzt (Paralleltest-Reliabilität) oder es wird derselbe Test nach einer gewissen Zeit erneut vorgegeben (Retest-Reliabilität). Beide Verfahren sind mit Problemen behaftet: Erinnerungs- und Übungseffekte können die Korrelation der Testergebnisse verfälschen. Beim Einsatz von Paralleltests muss zusätzlich sichergestellt sein, dass beide Testvarianten äquivalent sind (Fisseni, 1990). Ein weiteres Verfahren, das zur Reliabilitätsbestimmung eingesetzt wird, ist die Berechnung der inneren Konsistenz: Die Bearbeitungsergebnisse von Bestandteilen des Tests (Aufgabengruppen bzw. Aufgaben) werden als Messwiederholungen aufgefasst und miteinander korreliert (Lienert & Raatz, 1994).

Zur Bestimmung der inneren Konsistenz des INKA wurde der Alpha-Koeffizient von Cronbach mit den Daten der Stichprobe zur Aufgabenanalyse ($n = 350$) berechnet. Er beträgt $r = ,81$. Dieser Wert darf als akzeptabel gelten (vgl. Lienert & Raatz, 1994; Bartram & Lindley, 1994), zumal die Streuung der Schwierigkeitskoeffizienten der Aufgaben des INKA sehr groß ist ($,02 \leq p \leq ,90$).

In Teil 1, Kapitel 5.2.3 wurde aufgezeigt, dass INKA Rasch-skaliert ist: Seine Aufgaben sind alle auf dem Aufgabenparameter-Personenparameter-Kontinuum lokalisiert, das die zu messende latente Persönlichkeitsdimension repräsentiert. Sie sind also homogen und unterscheiden sich nur in ihrer Schwierigkeit. Die Homogenität ist im Rahmen der Probabilistischen Testtheorie der Indikator für die Messgenauigkeit eines Verfahrens, also für das, was im Rahmen der Klassischen Testtheorie mit der inneren Konsistenz beschrieben wird.

6.3 Validität

Zur Überprüfung der Validität wurde das Inventar Komplexer Aufmerksamkeit einer faktorenanalytischen Konstruktvalidierung (Fischer, 1974) und mehreren Studien zur konvergenten und diskriminanten Validität unterzogen. Im Teil 3 der vorliegenden Arbeit werden zudem die Ergebnisse einer modellgeleiteten Konstruktvalidierung des INKA ausführlich dargestellt.

6.3.1 Faktorenanalytische Konstruktvalidierung

In die Berechnungen gingen die Testergebnisse von 615 Personen ein, die INKA im Rahmen einer Testbatterie zusammen mit dem Intelligenz-Struktur-Test 70 (IST 70; Amthauer, 1973) und dem Aufmerksamkeits-Belastungs-Test d2 (Test d2; Brickenkamp, 1987) bearbeiteten. Die Testbatterie wurde von einer hessischen Großstadt zur Auswahl von Bewerbern für Ausbildungsplätze aller Ausbildungsrichtungen (gewerblich-technisch und kaufmännisch-verwaltend) durchgeführt.

IST 70

Der IST 70 ist ein standardisiertes diagnostisches Verfahren zur Bestimmung der Intelligenz. Er liefert einen Befund über das Intelligenzniveau und soll einen Einblick in ihre Struktur ermöglichen (Amthauer, 1973; Brickenkamp, 1975). Der Test besteht aus insgesamt 180 Aufgaben; die reine Bearbeitungszeit beträgt 72 Minuten. Die Aufgaben gliedern sich in vier Bereiche mit folgenden Untertests:

1) Verbale Intelligenz

- Satzergänzung (SE) – „Urteilsbildung“
- Wortauswahl (WA) – „Erfassen von sprachlichen Bedeutungsgehalten“
- Analogien (AN) – „Kombinationsfähigkeit“
- Gemeinsamkeiten (GE) – „Sprachliche Abstraktionsfähigkeit“

2) Merkfähigkeit

- Merkaufgaben (ME) – „Merkfähigkeit“

3) Rechnerische Intelligenz

- Rechenaufgaben (RA) – „Praktisch-rechnerisches Denken“
- Zahlenreihen (ZR) – „Theoretisch-rechnerisches Denken“

4) Räumliche Vorstellung

- Figurenauswahl (FA) – „Vorstellungsfähigkeit“
- Würfelaufgaben (WÜ) – „Räumliches Vorstellen-Können“

Im Manual der 4. Auflage werden Standardwerte (Z-Werte) sowohl für die Rohwerte der einzelnen Untertests als auch für den Gesamtrohwert angegeben, die sich auf das Alter der Testpersonen beziehen (Spannweite: 12. bis 60. Lebensjahr). Schulbildungsspezifische Vergleichswerte werden nur am Rande für den Gesamtrohwert angegeben. Sie zeigen eine deutliche Bildungsabhängigkeit der Testleistung (Tabelle 1.7).

Schulbildung	AM	s	n
Sonderschule oder Volksschule ohne Abschluss	48	24	511
Volksschul- bzw. Hauptschulabschluss	77	23	6051
Abgang vor mittlerer Reife	86	22	537
Mittlere Reife oder Höhere Schule	102	20	2950
Abitur	116	18	1959
Universität oder Hochschulabschluss	123	17	956

Tab. 1.7: Mittelwerte und Standardabweichungen des Gesamtrohwerts des IST 70 bezogen auf die Schulbildung

Der Leistungsverlauf (Gesamtrohwert, Altersbereich von 12 bis 55 Jahren) zeigt einen steilen Anstieg vom 12ten Lebensjahr zum Leistungsmaximum im 19ten Lebensjahr und einen anschließenden kontinuierlichen Abfall dieses Leistungszugewinns um ca. 80% bis zum 55sten Lebensjahr auf (Querschnittsuntersuchung).

Test d2

Der Test d2 ist ein standardisierter Konzentrationstest, der auf dem Durchstreichprinzip nach Bourdon beruht. Er misst die Schnelligkeit und Genauigkeit der Unterscheidung ähnlicher visueller Reize (Brickenkamp, 1975; Brickenkamp, 1987). In 14 Zeilen, die aus jeweils 47 Zeichen bestehen, müssen bestimmte Zielzeichen (ein ‚d‘, das mit 2 Strichen in unterschiedlicher Position versehen ist) entdeckt und durchgestrichen werden. Die Bearbeitungszeit ist für jede Zeile auf 20 Sekunden begrenzt; die Gesamtbearbeitungsdauer beträgt 4 Minuten und 40 Sekunden. Der Testwert (GZ-F) ergibt sich aus der Gesamtzahl

aller bearbeiteten Zeichen (GZ) abzüglich der Gesamtzahl der dabei gemachten Fehler (F). Für die Testrohwerte liegen schulbildungs- und altersbezogene Normen für männliche und weibliche Testpersonen vor.

Die Stichprobe, bei der die Testbatterie INKA, IST 70 und Test d2 eingesetzt wurde, baute sich soziodemografisch wie folgt auf (Tabelle 1.8):

	männlich	weiblich	gesamt
Geschlecht	302	313	615
Schulbildung			
Hauptschule	19	29	48
mittlerer Bildungsweg	96	150	246
Abitur, Fachhochschulreife	154	126	280
Hochschule	33	8	41
Alter (in Jahren)	AM = 23,1 s = 6,2 Min.: 15 Max.: 39	AM = 18,5 s = 3,0 Min.: 14 Max.: 34	AM = 20,7 s = 5,3 Min.: 14 Max.: 39

Tab. 1.8: Soziodemografische Daten der Stichprobe der ersten Validitätsuntersuchung

Folgende Testrohwerte gingen in die Auswertungen ein (Abbildung 1.11):

Intelligenz-Struktur-Test 70	Aufmerksamkeits- Belastungs-Test d2	Inventar Komplexer Aufmerksamkeit
IST-SE..... Satzergänzung IST-WA..... Wortauswahl IST-AN..... Analogien IST-GE..... Gemeinsamkeiten IST-ME Merkaufgaben IST-RA..... Rechenaufgaben IST-ZR Zahlenreihen IST-FA Figurenauswahl IST-WÜ..... Würfelaufgaben (jeweils Anzahl der richtig bearbeiteten Aufgaben)	d2 Gesamtzahl aller bearbeiteten Zeichen abzüglich aller Fehler (GZ-F)	INKA Anzahl der richtig bearbeiteten Aufgaben

Abb. 1.11: Testwerte, die in die erste Validitätsuntersuchung Eingang fanden

Die Variablen INKA, Test d2 und alle 9 Untertests des IST 70 sollen einer explorativen Faktorenanalyse unterzogen werden. Als Statistikprogramm zur Durchführung dieser und aller folgenden Berechnungen dient SPSS[®] für Windows[™], Version 6.0.1.

Als Erstes muss geklärt werden, ob sich die vorliegenden Daten für eine Faktorenanalyse eignen.

Nach Brosius (1998) sollte zwischen den Variablen, die in eine Faktorenanalyse einfließen, Multikollinearität vorliegen. Variablen, die nur gering mit den übrigen Variablen korrelieren, sollten bei der Faktorenanalyse unberücksichtigt bleiben. Zur Überprüfung der Multikollinearität schlägt er die Sichtung der Korrelationsmatrix und ggf. die Durchführung des Bartlett-Test auf Sphärizität vor.

Die Korrelationsmatrix gibt einen ersten Eindruck davon, welche Variablen stark und welche schwach miteinander korrelieren. Die Durchführung einer Faktorenanalyse wäre wenig sinnvoll, wenn alle ermittelten Korrelationskoeffizienten nur sehr geringe absolute Werte aufweisen würden. Gemeinsame Faktoren existieren nämlich nur für solche Variablen, die relativ stark miteinander korrelieren.

	INKA	Test d2	IST-SE	IST-WA	IST-AN	IST-GE	IST-ME	IST-RA	IST-ZR	IST-FA	IST-WÜ
INKA	1,00000										
Test d2	,44019	1,00000									
IST-SE	,31422	,29736	1,00000								
IST-WA	,36050	,32709	,60553	1,00000							
IST-AN	,37028	,35637	,66534	,68647	1,00000						
IST-GE	,36505	,35299	,61250	,60734	,63944	1,00000					
IST-ME	,35639	,39043	,34796	,32592	,36378	,37799	1,00000				
IST-RA	,40930	,39942	,61848	,56795	,61221	,53905	,33744	1,00000			
IST-ZR	,44290	,43353	,40805	,48468	,43716	,41696	,37485	,62528	1,00000		
IST-FA	,29733	,31128	,38520	,34099	,41188	,43433	,20257	,43945	,35606	1,00000	
IST-WÜ	,33627	,32124	,35686	,33673	,41693	,38311	,26333	,43068	,41267	,48843	1,00000

Tab. 1.9: Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen

Die für die vorliegende Stichprobe ausgewiesenen Korrelationskoeffizienten (Tabelle 1.9) streuen zwischen $r = ,20257$ und $r = ,68647$. Um auszuschließen, dass sich die Korrelationen nur zufällig in der Stichprobe ergeben haben, obwohl in der Grundgesamtheit keinerlei Zusammenhang zwischen den Variablen besteht, wurde der Bartlett-Test auf Sphärizität durchgeführt. Er testet die Hypothese, dass alle Korrelationskoeffizienten zwischen den Variablen in der Grundgesamtheit den Wert $,0$ haben. Die Testgröße ist ein Chi-Quadrat-Wert. Für die vorliegenden Korrelationskoeffizienten ergibt sich ein Chi-Quadrat von $\chi^2 = 2967,1885$, für das eine Wahrscheinlichkeit von $p = ,000$ ausgewiesen wird. Die Nullhypothese, alle Korrelationen zwischen den 11 Variablen seien in der Grundgesamtheit gleich Null, kann zurückgewiesen werden.

Weiter schlägt Brosius (1998) die Sichtung der Anti-Image-Korrelationsmatrix vor. Dieser Matrix liegen partielle Korrelationskoeffizienten zugrunde. Partielle Korrelationskoeffizienten ergeben sich, wenn die linearen Einflüsse der übrigen Variablen auf die beiden jeweils

betreffenden Variablen eliminiert werden. Es werden also die durch die übrigen Variablen nicht erklärten Teile miteinander korreliert. Wenn das Faktorenmodell geeignet ist, sollten die Werte nahe Null liegen. Tabelle 1.10 enthält die Matrix für die vorliegenden Variablen.

	INKA	Test d2	IST-SE	IST-WA	IST-AN	IST-GE	IST-ME	IST-RA	IST-ZR	IST-FA	IST-WÜ
INKA	,93360										
Test d2	-,21851	,91867									
IST-SE	,02861	,04184	,91589								
IST-WA	-,04351	,00302	-,15496	,91424							
IST-AN	-,02971	-,03664	-,24497	-,34416	,90907						
IST-GE	-,04766	-,04178	-,20691	-,20224	-,19119	,93567					
IST-ME	-,12200	-,19118	-,08389	,01600	-,05673	-,11175	,92220				
IST-RA	-,05117	-,05691	-,26653	-,05827	-,15900	-,03426	,03540	,90703			
IST-ZR	-,14323	-,13559	,05843	-,16593	,06749	,00974	-,12326	-,37867	,88806		
IST-FA	-,03244	-,07717	-,04350	,03686	-,04455	-,15107	,06940	-,10199	-,02218	,90655	
IST-WÜ	-,07626	-,04440	-,00522	,03792	-,10321	-,03074	-,03225	-,05249	-,12600	-,30716	,91443

Tab. 1.10: Anti-Image-Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen

Ein zusammenfassendes Maß für die Eignung des Faktorenmodells, in das die partiellen Korrelationskoeffizienten einfließen, stellt das Kaiser-Mayer-Olkin-Maß (KMO) dar. Es kann einen Maximalwert von 1 erreichen. Ein hoher KMO-Wert zeigt an, dass die Variablenauswahl für eine Faktorenanalyse gut geeignet ist. Im vorliegenden Fall ist der KMO-Wert = 0,91414. Nach Kaiser (zitiert nach Brosius, 1998, S. 647) sind Werte zwischen 0,9 und 1,0 als „fabelhaft“ zu beurteilen. Abschließend sind die MSA-Werte (Measures of Sampling Adequacy), die sich in der Hauptdiagonalen der Matrix befinden, zu betrachten. Sie entsprechen inhaltlich dem KMO-Maß, beziehen sich aber jeweils nur auf eine Variable. Auch hier sind Werte zwischen 0,9 und 1,0 als „fabelhaft“ zu beurteilen. Werte zwischen 0,8 bis unter 0,9 gelten als „recht gut“. Die Ergebnisse aller Untersuchungen zeigen, dass die 11 Variablen sehr gut zur Berechnung einer Faktorenanalyse geeignet sind.

Bortz (1993, S. 483) weist auf einen zusätzlichen Aspekt hin: „Um zu möglichst stabilen, vom Zufall weitgehend unbeeinflussten Faktorenstrukturen zu gelangen, sollte die verwendete Stichprobe möglichst groß und repräsentativ sein.“ Die vorliegende Stichprobengröße von

n = 615 reicht nach seinen weiteren Ausführungen für eine generalisierende Interpretation der Faktorenstruktur aus. Ihre Repräsentativität ist allerdings eingeschränkt.

Als Methode zur Verrechnung der Daten dient eine Hauptkomponentenanalyse mit anschließender orthogonaler Varimax-Rotation. Ihr Ergebnis enthält Tabelle 1.12. Faktoren, die mittels einer Hauptkomponentenanalyse gewonnen werden, sind voneinander unabhängig und erklären sukzessive maximale Varianz. Die Varimax-Rotation soll eine möglichst gute Einfachstruktur für die bedeutsamen Faktoren herstellen, um die Interpretierbarkeit der Faktoren zu verbessern. Einfachstruktur besagt, dass auf jedem Faktor einige Variablen möglichst hoch und andere möglichst niedrig und dass auf verschiedenen Faktoren verschiedene Variablen möglichst hoch laden sollen (Bortz, 1993).

Zur Bestimmung der Anzahl der bedeutsamen Faktoren dienen der Scree-Test und das Kaiser-Guttman-Kriterium. Beim Scree-Test wird das Eigenwertediagramm, das die Eigenwerte aller Faktoren in abfallender Größe grafisch darstellt, gesichtet. Beim kleinsten Wert rechts beginnend folgt man dem Verlauf der Eigenwerte bis die annähernde Konstanz ihrer Größen endet, was zu einem Knick im Eigenwerteverlauf führt. Das Kaiser-Guttman-Kriterium legt fest, dass nur Faktoren bei der Interpretation berücksichtigt werden, deren Eigenwert größer oder gleich 1 ist. Die datenreduzierende Funktion einer Hauptkomponentenanalyse ist nämlich nur dann gewährleistet, wenn ausschließlich Faktoren interpretiert werden, die mindestens eine Varianz von 1 aufklären, also mehr als jede einzelne z-standardisierte Variable (Bortz, 1993).

Wie man dem Eigenwertediagramm in Abbildung 1.12 entnehmen kann, ist der Verlauf der Eigenwerte von rechts nach links bis hin zum Faktor 4 annähernd konstant. Beim Übergang von Faktor 4 zum Faktor 3 steigt der Eigenwerteverlauf sichtbar an. Nach dem Scree-Test könnte man folglich die ersten drei Faktoren als bedeutsam erachten. Betrachtet man allerdings die numerischen Eigenwerte der Faktoren in Tabelle 1.11, so stellt man fest, dass Faktor 3 mit einem Eigenwert von $\lambda = ,94027$ weniger Varianz aufklärt als eine einzelne Variable. Für die Rotation wurden dennoch die ersten drei Faktoren ausgewählt, nicht zuletzt da sich durch eine orthogonale Faktorenrotation die Ladungen der Variablen in der Art auf den Faktoren verschieben, dass sich der Eigenwert des ersten Faktors zugunsten der Eigenwerte der anderen Faktoren reduziert. Die Eigenwerte nach Rotation sind in Tabelle

1.11 zusätzlich für die ersten drei Faktoren dargestellt. Faktor 3 besitzt nach Rotation nun einen Eigenwert von $\lambda = 1,72659$.

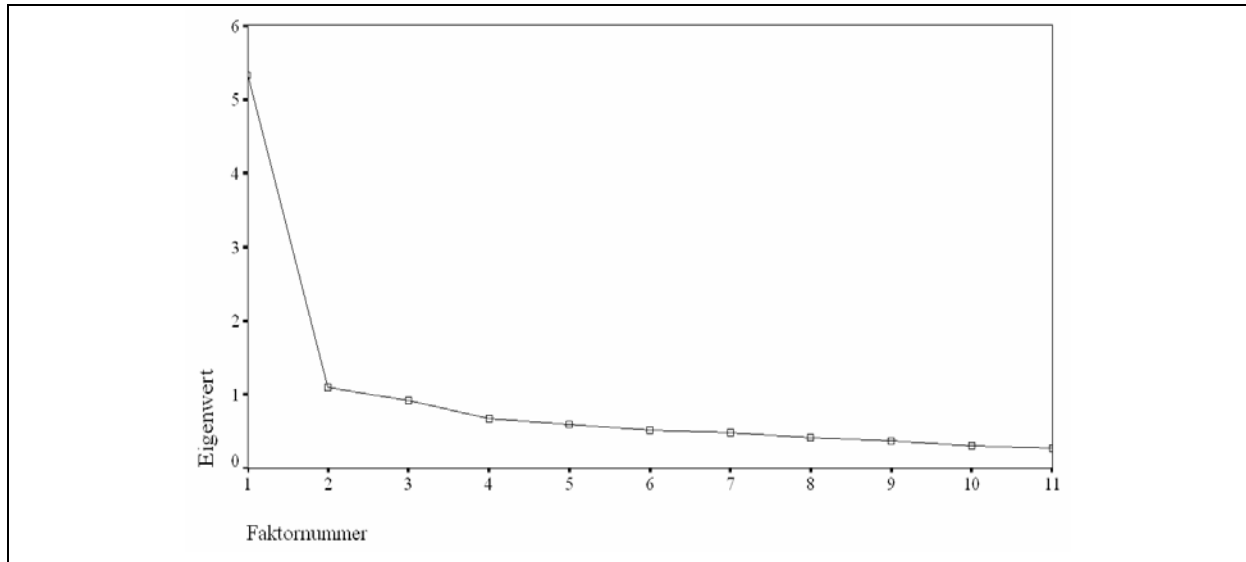


Abb. 1.12: Diagramm der Eigenwerte der durch die Hauptkomponentenanalyse gewonnenen Faktoren

Faktor	Eigenwert	Varianzaufklärung	VA kumuliert	Eigenwert nach Rotation
1	5,02050	45,6%	45,6%	3,06383
2	1,05932	9,6%	55,3%	2,22967
3	,94027	8,5%	63,8%	1,72659
4	,71452	6,5%	70,3%	
5	,64143	5,8%	76,1%	
6	,56704	5,2%	81,3%	
7	,51966	4,7%	86,0%	
8	,46644	4,2%	90,3%	
9	,42620	3,9%	94,1%	
10	,33958	3,1%	97,2%	
11	,30506	2,8%	100,0%	

Tab. 1.11: Ergebnis der Hauptkomponentenanalyse

Variable	Kommunalität	Ladung auf Faktor 1	Ladung auf Faktor 2	Ladung auf Faktor 3
INKA	,57729	,14020	,71218	,22459
Test d2	,63290	,16285	,75537	,18919
IST-SE	,68836	,79561	,17040	,16226
IST-WA	,69329	,78406	,25275	,12106
IST-AN	,73199	,79582	,24215	,20006
IST-GE	,60581	,72256	,17018	,23399
IST-ME	,51811	,28612	,65572	-,07916
IST-RA	,63386	,57506	,41893	,35731
IST-ZR	,52371	,31016	,55027	,35314
IST-FA	,71828	,27570	,08422	,79698
IST-WÜ	,69649	,17240	,23665	,78152

(höchste Ladungen grau unterlegt)

Tab. 1.12: Kommunalitäten der Variablen und Ladungen der Variablen auf den rotierten Faktoren

Folgende inhaltliche Interpretationen der rotierten Faktoren werden vorgeschlagen:

Faktor 1: ‚Verbale Intelligenz‘

Der erste Faktor bildet die sprachlichen Untertests des IST 70 ab (IST-SE, IST-WA, IST-AN, IST-GE). Die Rechenaufgaben des IST 70 (IST-RA) haben ebenfalls auf Faktor 1 ihre höchste Ladung. Da sich bei den Rechenaufgaben um Textaufgaben handelt, setzt ihr Verständnis, und damit ihre Lösung, ebenfalls sprachliche Fähigkeiten voraus.

Faktor 2: ‚Aufmerksamkeit und Gedächtnis‘

Auf dem zweiten Faktor laden die beiden Aufmerksamkeitstests INKA und Test d2 am höchsten. Bei beiden Aufmerksamkeitstests ist eine Gedächtnisleistung für die Bearbeitung der Aufgaben notwendig: Die Zielreize müssen memoriert werden, um sie in den Suchreihen

finden zu können. Insofern ist die Lokalisierung des Untertests Merkfähigkeit des IST 70 (IST-ME) auf Faktor 2 plausibel. Die Zahlenreihen des IST 70 (IST-ZR) laden auch auf Faktor 2 am höchsten. Dies könnte ebenfalls durch die erforderliche Gedächtnisleistung begründet sein: Um vermutete Veränderungsregeln auf ihre Gültigkeit hin zu überprüfen, muss man sie memorieren und auf die Zahlenreihen anwenden.

Faktor 3: ‚Räumliche Vorstellung‘

Der dritte Faktor setzt sich zusammen aus den beiden Untertests des IST 70 zum räumlichen Vorstellen, Figurenauswahl (IST-FA) und Würfelaufgaben (IST-WÜ).

Fazit

Das Ergebnis der explorativen Faktorenanalyse zeigt, dass das Inventar Komplexer Aufmerksamkeit selektive visuelle Aufmerksamkeitsleistungen und -aufgabenbedingt - kurzzeitige Merkfähigkeit erfasst.

6.3.2 Konvergente und diskriminante Validität

Hinweise auf die Validität eines Tests können Untersuchungen liefern, bei denen neben dem zu überprüfenden Verfahren weitere eingesetzt werden, die dasselbe personenbezogene Merkmal erfassen, und solche, die andere Merkmale messen. Konvergente Validität wäre gegeben, wenn INKA deutliche Zusammenhänge mit Tests aufweist, die ihrerseits das Konstrukt ‚Aufmerksamkeit‘ messen. Diskriminante Validität läge vor, wenn INKA zu Tests, die andere Konstrukte erfassen, keine Zusammenhänge zeigt.

Nachfolgend werden Zusammenhänge mit Verfahren untersucht, die die allgemeine intellektuelle Leistungsfähigkeit, die Aufmerksamkeit oder das sprachgebundene Wissen erfassen. Im Einzelnen sind dies:

- Intelligenz-Struktur-Test 70 (IST 70; Amthauer, 1973)
- Aufmerksamkeits-Belastungs-Test d2 (Test d2; Brickenkamp, 1987)
- Intelligenz-Struktur-Analyse (ISA; Fay, Trost & Gittler, 2001)
- Standard Progressive Matrices (SPM; Raven, 1996; Raven, Raven & Court, 1999)
- Advanced Progressive Matrices (APM; Raven, 1995; Raven, Raven & Court, 1998b)
- Frankfurter Aufmerksamkeits-Inventar (FAIR; Moosbrugger & Oehlschlägel, 1996)
- Wortschatztest-aktiv und passiv (WST-ap; Ibrahimovic & Bulheller, 2005)

6.3.2.1 Intelligenz-Struktur-Test 70 und Aufmerksamkeits-Belastungs-Test d2

Zur Bestimmung der konvergenten und diskriminanten Validität des INKA werden zunächst die Daten der Stichprobe, die auch als Grundlage für die Faktorenanalyse diente, herangezogen (siehe Teil 1, Kapitel 6.3.1). Es werden die korrelativen Zusammenhänge zwischen dem Testergebnis im INKA und dem IST 70-Gesamtwert und dem Ergebnis im Test d2 (GZ-F) untersucht. Weiterhin wird der Einfluss der Schulbildung der Testpersonen auf die Leistung im INKA, im Test d2 und im IST 70 analysiert. (Ein Aspekt, der genau genommen die Kriteriumsvalidität des Tests betrifft.) Für die Variable ‚Schulbildung‘ besteht die Annahme, dass sie einen deutlich stärkeren positiven Zusammenhang mit der Testleistung im IST 70 zeigt als mit der Leistung im INKA und der im Test d2. Der Grund für die Annahme: Der IST 70 misst die kognitive Leistungsfähigkeit überwiegend (7 von 9 Untertests) mit Hilfe von Aufgaben, deren Bewältigung unmittelbar von kulturspezifischem Wissen und der Beherrschung von Kulturtechniken abhängt (Allgemeinwissen, Umgang mit Sprache, Umgang mit Zahlenmaterial). Kulturspezifisches Wissen und Kulturtechniken werden in besonderem Maße in schulischen Bildungseinrichtungen vermittelt (siehe hierzu auch Tabelle 1.7). Die Leistung bei der Bearbeitung der Aufgaben des INKA und des Test d2 erfordern, abgesehen von der Kenntnis der lateinischen Buchstaben, kein besonderes Wissen oder Fertigkeiten, die im schulischen Unterricht vermittelt werden.

Da die Gesamtwerte des INKA annähernd normalverteilt sind (siehe Teil 1, Kapitel 6.4) und für den IST 70 und den Test d2 Standardnormen berechnet wurden, also von einer Normalverteilung der Testwerte ausgegangen werden darf, werden Korrelationen nach Pearson berechnet. Die Schulbildung gliedert sich in vier ordinale Kategorien auf (Hauptschule, Realschule, Gymnasium/Fachoberschule, Hochschule). Für die Ermittlung

ihrer korrelativen Zusammenhänge werden Rangkorrelationen nach Spearman berechnet. Zur besseren Interpretierbarkeit der Ergebnisse wird jeweils auch die Redundanz angegeben. Statistisch gesehen ist sie das Maß für die gemeinsame Varianz zweier Merkmale, ausgedrückt in Prozent. Sie gibt also an, wie viel Unterschiedlichkeit der einen Variablen durch die Unterschiede der anderen Variablen erklärt werden kann (Bortz, 1993). Die Redundanzen wurden auf eine Stelle hinter dem Komma aufgerundet. Alle dargestellten Korrelationen in Tabelle 1.13 sind hoch signifikant ($p = ,000$; 1-seitig).

	Test d2	IST 70-gesamt	Schulbildung
INKA	,4876	,4945	,3466
Red _(yx)	23,8%	24,5%	12,0%
Test d2		,5231	,4024
Red _(yx)		27,4%	16,2%
IST 70-gesamt			,6522
Red _(yx)			42,5%

Tab. 1.13: Korrelationen und Redundanzen zwischen INKA und Test d2, IST 70-gesamt und Schulbildung

Wie erwartet existiert im Sinne der konvergenten Validität ein hoher positiver Zusammenhang zwischen dem INKA und dem Test d2. Es zeigt sich aber auch ein hoher positiver Zusammenhang zwischen INKA und dem Gesamtwert des IST 70. Falls Aufmerksamkeit und Intelligenz unterschiedliche Konstrukte darstellen, könnte dieses Ergebnis gegen die diskriminante Validität des INKA sprechen. Bei der Schulbildung zeigt sich das erwartete Ergebnis: Die Redundanz zwischen der Schulbildung und INKA beträgt mit 12% weniger als ein Drittel der Redundanz zwischen der Schulbildung und IST 70-gesamt von 42,5%. Beim Test d2 zeigt sich ein ähnliches Ergebnis.

6.3.2.2 Exkurs: Hauptschüler mit und ohne Berufsausbildung

Dass INKA ein Leistungsmerkmal erfasst, das unabhängig von der schulischen Ausbildung ist, zeigt nicht nur die vorausgegangene Korrelationsstudie mit dem IST 70 sondern auch das Ergebnis einer Untersuchung, bei der INKA von Hauptschülern bearbeitet wurde, die entweder eine zusätzliche Berufsausbildung (Lehre) vorzuweisen hatten oder nicht. Die Daten wurden im Rahmen von Qualifizierungsmaßnahmen für Arbeitssuchende eines bundesweit tätigen Weiterbildungsträgers gewonnen. Den soziodemografischen Aufbau der Stichprobe gibt Tabelle 1.14 wieder:

	Hauptschüler	
	ohne zusätzliche Berufsausbildung	mit zusätzlicher Berufsausbildung
Geschlecht		
männlich	116	141
weiblich	136	134
gesamt	252	275
Alter (in Jahren)	AM = 30,4 s = 13,5 Min.: 16 Max.: 62	AM = 36,9 s = 11,6 Min.: 16 Max.: 61

Tab. 1.14: Soziodemografischer Aufbau der Stichprobe zur Berechnung der Mittelwertsunterschiede in der Testleistung im INKA bei Hauptschülern mit und ohne zusätzliche Berufsausbildung

Zur Überprüfung, ob signifikante Mittelwertsunterschiede in der Bearbeitungsleistung im INKA zwischen Hauptschülern mit und ohne zusätzliche Berufsausbildung bestehen, wurde ein t-Test berechnet. Der Levene-Test auf Varianzgleichheit ergab einen F-Wert von $F = 6,028$. Dieser Wert hat eine 2-seitige Wahrscheinlichkeit von $p = ,014$. Die Nullhypothese, dass die beiden Bildungsgruppen aus Populationen mit gleichen Varianzen

stammen, muss verworfen werden und es muss ein Test für unabhängige Stichproben mit ungleicher Varianz berechnet werden. Das Ergebnis enthält Tabelle 1.15.

Gruppenstatistiken

Hauptschüler	n	AM	s	Standardfehler des AM
ohne Berufsausbildung	252	5,8056	3,352	,211
mit Berufsausbildung	275	5,6182	3,662	,221

Test bei unabhängigen Stichproben mit ungleicher Varianz

t-Wert	df	Sig. (2-seitig)	Differenz der AM	SE der Diff.	95%-KI der Diff.
,61	525	,540	,1874	,306	-,413; ,788

Tab. 1.15: Ergebnis des t-Tests mit der abhängigen Variablen ‚Testwert‘ und der Gruppierungsvariablen ‚Berufsausbildung‘

Der ermittelte t-Wert ist nicht signifikant. Die Nullhypothese, dass beide Bildungsgruppen aus Populationen mit gleichen Mittelwerten stammen, darf beibehalten werden. Eine zusätzliche Berufsausbildung nach Abschluss der Hauptschule hat offensichtlich keinen Einfluss auf die Bearbeitungsleistung im INKA. Dieses Ergebnis ist ein weiterer Hinweis darauf, dass die INKA-Testleistung nicht durch die Beherrschung von Kulturtechniken oder durch kulturspezifisches Wissen beeinflusst wird.

6.3.2.3 Weitere Korrelationsstudien zur konvergenten und diskriminanten Validität

Die nachfolgenden Daten zur korrelationsstatistischen Validierung des Inventars Komplexer Aufmerksamkeit wurden ebenfalls im Rahmen von Qualifizierungsmaßnahmen für Arbeitssuchende eines bundesweit tätigen Weiterbildungsträgers gewonnen. Zusammen mit INKA kamen zum Einsatz:

- Intelligenz-Struktur-Analyse (ISA; Fay, Trost & Gittler, 2001)
- Standard Progressive Matrices, Set A, B, C, D und E (SPM; Raven, 1996; Raven, Raven & Court, 1999)
- Advanced Progressive Matrices, Set I und II (APM; Raven, 1995; Raven, Raven & Court, 1998b)
- Frankfurter Aufmerksamkeits-Inventar (FAIR; Moosbrugger & Oehlschlägel, 1996)
- Wortschatztest-aktiv und passiv (WST-ap; Ibrahimovic & Bulheller, 2005)

ISA

Die Intelligenz-Struktur-Analyse ist ein standardisiertes schriftliches Testverfahren zur Diagnose von Teilbereichen der intellektuellen Befähigung und der generellen intellektuellen Leistungsfähigkeit. Sie besteht aus neun Untertests mit insgesamt 177 Aufgaben. Die Untertests erfassen verschiedene Facetten der sprachlichen Fähigkeiten (Untertests: Sätze ergänzen, Gemeinsamkeiten finden, Beziehungen erschließen, Begriffe bilden), der mathematisch-numerischen Fähigkeiten (Untertests: Zahlenreihen fortsetzen, Praktisches Rechnen), des räumlichen und figuralen Vorstellungsvermögens (Untertests: Würfel erkennen, Figuren zusammensetzen), sowie das mittelfristige Gedächtnis (Untertest: Waren merken). Die Bearbeitungszeit des Gesamttests beträgt, exklusive der Instruktionen, 90 Minuten.

Progressive Matrices von Raven: SPM, APM

Die Progressive Matrices von Raven messen die Beobachtungsgabe und das logische Denken mit Hilfe von sprachfreien, kulturunabhängigen Aufgaben. Die Testautoren beschreiben den Messgegenstand der Tests als die „Fähigkeit, Beziehungen zu stiften“ (Raven, Raven & Court, 1998a). Die „Fähigkeit Beziehungen zu stiften“ wird inhaltlich beschrieben als die Fähigkeit, „Ordnung aus dem Chaos zu erzeugen, ein neues Verständnis zu entwickeln, über das Vorgegebene hinaus zu gehen, das nicht auf den ersten Blick Erkennbare wahrzunehmen und (weitgehend nonverbale) Konstrukte zu schaffen, die die Behandlung komplexerer Probleme mit vielen voneinander abhängigen Variablen vereinfachen“ (Raven, Raven & Court, 1998a, S. 16). Operationalisiert wird das Konstrukt mit Hilfe von grafischen

Aufgabenstellungen, die regelmäßige Veränderungen in zwei Dimensionen gleichzeitig aufweisen. Die grafischen Darstellungen sind unvollständig und die Testperson hat aus vorgegebenen Antwortmöglichkeiten das ergänzende Teil auszuwählen.

SPM

Die Standard Progressive Matrices bestehen aus fünf Sets (A, B, C, D, E) mit jeweils 12 grafischen Aufgabenstellungen. Die Sets steigen in ihrer Schwierigkeit systematisch an (Raven, Raven & Court, 1999). Beim Set A müssen geometrische Muster, die eine Leerstelle aufweisen, mit Hilfe einer von sechs Mustervorgaben stimmig ergänzt werden. Set B enthält 2x2-Matrizen, die aus geometrischen Figuren bestehen. Die Figur in der untersten Zeile rechts fehlt. Mit Hilfe von einer von sechs Figurenvorgaben muss die letzte Zeile logisch richtig ergänzt werden. Set C, Set D und Set E enthalten 3x3-Matrizen, die aus geometrischen Figuren bestehen. Wiederum fehlt die Figur in der untersten Zeile rechts. Hier muss mit Hilfe einer von acht Figurenvorgaben die letzte Zeile logisch richtig ergänzt werden. Die Bearbeitungszeit der SPM ist nicht begrenzt. Sie beträgt i. d. R. 30 Minuten.

APM

Die Advanced Progressive Matrices bestehen aus zwei Sets (Set I, Set II) mit 12 bzw. 36 grafischen Aufgabenstellungen. Sie stellen eine schwierigere Variante der SPM dar, die entwickelt wurde, um bessere Unterscheidungen im oberen Leistungsbereich treffen zu können. Set I und Set II sind in separaten Testheften enthalten. Set I dient zu Übungszwecken bzw. um entscheiden zu können, ob die SPM oder die APM für die weitere Beurteilung der intellektuellen Leistungsfähigkeit besser geeignet sind (Raven, Raven & Court, 1998b). Für die Bearbeitung von Set I und Set II sind insgesamt 60 Minuten zu veranschlagen.

FAIR

Das Frankfurter Aufmerksamkeits-Inventar erhebt den Anspruch „zur Untersuchung des individuellen Aufmerksamkeitsverhaltens“ geeignet zu sein (Moosbrugger & Oehlschlägel,

1996, S. 5). Das Testprinzip ist folgendes: In 32 Zeilen, die aus jeweils 20 geometrischen Zeichen bestehen, sollen innerhalb von sechs Minuten bestimmte Zeichen von links nach rechts identifiziert und mit einer durchgehenden Linie verbunden werden. Als Testkennwerte werden erfasst:

- Gesamtwert (G): Menge aller bearbeiteten Zeichen während der Testdauer
- Leistungswert (L): Menge der konzentriert bearbeiteten Zeichen (Richtige)
- Qualitätswert (Q): Anteil der konzentrierten Urteile an allen abgegebenen Urteilen (L/G)
- Kontinuitätswert (K): Ausmaß der kontinuierlich aufrecht erhaltenen Konzentration (QxL)

WST-ap

Der Wortschatztest-aktiv und passiv von Ibrahimovic und Bulheller (2005) wird von den Testautoren als die deutschsprachige Version der Mill-Hill Vocabulary Scale von Raven, Court und Raven (1994) bezeichnet. Wortschatztests sind ein klassisches Maß der kristallinen Intelligenz, also des kulturspezifischen Wissens. Der Test besteht aus zwei Teilen. Im ersten Teil wird der aktive Wortschatz durch ein freies Antwortformat erfasst: 30 vorgegebene Wörter sollen schriftlich inhaltlich definiert werden. Im zweiten Teil wird der passive Wortschatz durch eine Wiedererkennungslleistung (multiple choice-Format) überprüft: Zu 30 vorgegebenen Wörtern sollen aus jeweils 6 dazugehörigen Begriffen alle ausgewählt werden, die mit dem vorgegebenen Wort inhaltlich übereinstimmen. Das Verfahren liegt in den Parallelformen A und B vor. Die Gesamtbearbeitungsdauer beträgt für beide Formen jeweils 23 Minuten.

Im Sinne der diskriminanten Validität dürften allenfalls niedrige positive Korrelationen zwischen dem INKA und der ISA, den SPM und den APM bestehen. Dasselbe gilt auch für den Zusammenhang zwischen dem INKA mit dem WST-ap. Im Sinne der konvergenten Validität müsste sich ein deutlicher positiver Zusammenhang zwischen dem INKA und dem FAIR zeigen.

Die Stichproben, mit deren Hilfe die Daten gewonnen wurden, hatten folgenden soziodemografischen Aufbau (Tabelle 1.16; FOS steht für Fachoberschule):

	ISA	SPM	APM	FAIR	WST-ap	
					Form A	Form B
Geschlecht						
männlich	380	211	88	229	345	285
weiblich	353	198	81	221	348	292
Schulbildung						
Hauptschule	139	26	24	112		
Realschule	282	134	100	172	364	355
Gymn., FOS	304	221	42	130	329	222
Hochschule	8	6	3	14		
<i>keine Angabe</i>		22		22		
gesamt	733	409	169	450	693	577
Alter (in Jahren)	AM = 16,9 s = 3,5 Min.: 14 Max.: 54	AM = 17,4 s = 4,5 Min.: 14 Max.: 58	AM = 17,2 s = 4,9 Min.: 14 Max.: 53	AM = 17,6 s = 5,5 Min.: 14 Max.: 53	AM = 16,2 s = 1,45 Min.: 14 Max.: 19	AM = 15,8 s = 1,25 Min.: 14 Max.: 19

Tab. 1.16: Soziodemografische Daten der Stichproben der weiteren Korrelationsstudien zur konvergenten und diskriminanten Validität

Folgende Testrohwerter gingen in die Berechnungen ein (Abbildung 1.13):

ISA	SPM	APM	FAIR	WST-ap
Gesamtwert (ges.)	Gesamtwert (ges.)	Gesamtwert (ges.)	Leistungswert (L) Qualitätswert (Q) Kontinuitätswert (K)	Form A (A) Form B (B)

Abb. 1.13: Testwerte, die mit dem Ergebnis im INKA korreliert wurden

Es wurden parametrische Korrelationen berechnet (Pearsonscher Korrelationskoeffizient). Die Ergebnisse für den Wortschatztest-aktiv und passiv werden zitiert nach Ibrahimovic und Bulheller (2005). Alle Korrelationen sind hoch signifikant ($p = ,000$; 1-seitig).

	ISA	SPM	APM	FAIR			WST-ap	
	ges.	ges.	ges.	L	Q	K	A	B
INKA	,6084	,5070	,4715	,4755	,3083	,4950	,268	,200
Red _(yx)	37,0%	25,7%	22,2%	22,6%	9,5%	24,5%	7,2%	4,0%

Tab. 1.17: Korrelationen und Redundanzen zwischen INKA und ISA, SPM, APM, FAIR und WST-ap

Abbildung 1.17 zeigt, dass die Korrelationen des INKA mit den Intelligenztests ISA, SPM und APM auf dem unerwartet hohen Niveau liegen, das sich schon beim IST 70-Gesamtwert gezeigt hatte (siehe Teil 1, Kapitel 6.3.2.1). Beim FAIR entspricht der Leistungswert (L) den gängigen Maßen von Konzentrationstests (Anzahl aller bearbeiteter Stimuli abzüglich der Anzahl der hierbei gemachten Fehler) und ist als das Äquivalent des Testwertes des Test d2 (GZ-F) zu sehen. Der Zusammenhang zwischen INKA und FAIR-L liegt auf demselben erwarteten Niveau wie der Zusammenhang zwischen INKA und Test d2. Zum Wortschatztest-aktiv und passiv zeigen sich für INKA die erwarteten niedrigen korrelativen Zusammenhänge.

Alle Zusammenhangswerte sind vor dem Hintergrund zu betrachten, dass die Validität der eingesetzten Verfahren als gegeben angenommen und Minderungskorrekturen nicht berechnet wurden (Lienert & Ratz, 1994).

6.3.2.4 Die Untersuchung von Schmidt-Atzert et al.

Ähnliche Zusammenhänge der Testleistung im INKA mit Maßen der Intelligenz und der Aufmerksamkeit konnten Schmidt-Atzert, Bühner und Enders (2006) ermitteln. Sie untersuchten die Struktur von Leistungen in Konzentrationstests mit Hilfe konfirmatorischer Faktorenanalysen. In ihre Berechnungen gingen die Testergebnisse von 117 Studenten ein. Die Studenten bearbeiteten insgesamt 11 Tests zur Erfassung von Konzentration und verwandten Konstrukten (z. B. Informationsverarbeitungsgeschwindigkeit) sowie weitere Tests, die zur anschließenden Validierung des Faktorenmodells herangezogen wurden. Zur Validierung dienten Tests, die das „relativ konstruktferne Merkmal Intelligenz“ (Schmidt-

Atzert, Bühner und Enders, 2006, S. 34) sowie spezifische Fähigkeiten erfassten, die für die Bearbeitung der eingesetzten Konzentrationstests bedeutsam sein konnten (verbale, numerische und figurale Fähigkeiten, Merkfähigkeit, motorische Schnelligkeit).

Folgende Konzentrationstests wurden eingesetzt:

- Inventar Komplexer Aufmerksamkeit (INKA; Heyde, 2000)
- Aufmerksamkeits-Belastungs-Test d2 (Test d2; Brickenkamp, 2002)
- Frankfurter Aufmerksamkeitsinventar (FAIR; Moosbrugger & Oehlschlegel, 1996)
- Konzentrations-Leistungs-Test 6–13, Form A (KLT-R; Lukesch & Mayrhofer, 2001)
- Revisions-Test (REV-T; Marschner, 1972)
- Berliner-Intelligenz-Struktur-Test (BIS; Jäger, Süß & Beauducel, 1997), je zwei verbale, numerische und figurale Tests zur *Bearbeitungsgeschwindigkeit*: Klassifizieren von Wörtern (BIS-KW), Unvollständige Wörter (BIS-UW), X-Größer (BIS-XG), Teil-Ganzes (BIS-TG), Buchstaben-Durchstreichen (BIS-BD), Zahlen-Symbol-Test (BIS-ZS)

Folgende Tests dienen zur anschließenden Validierung:

- *Verbale Fähigkeiten*: Mehrfachwahl-Wortschatz-Intelligenztest, Form B (MWT-B; Lehrl, 1999)
- *Rechenfertigkeit*: Diagnostischer Rechentest für 3. Klassen (DRE3; Samstag, Sander & Schmidt, 1971)
- *Figurale Fähigkeiten*: Gleiche-Figuren-Test (GFT; unveröffentlicht)
- *Intelligenz (Reasoning)*: Berliner-Intelligenz-Struktur-Test (BIS; Jäger, Süß & Beauducel, 1997), je zwei verbale, numerische und figurale Tests zur *Verarbeitungskapazität*: Wortanalogien (BIS-WA), Tatsache/Meinung (BIS-TM), Zahlenreihen (BIS-ZN), Schätzen (BIS-SC), Analogien (BIS-AN), Charkow (BIS-CH)
- *Merkfähigkeit*: Berliner-Intelligenz-Struktur-Test (BIS; Jäger, Süß & Beauducel, 1997), Orientierungs-Gedächtnis (BIS-OG), Zahlen-Paare (BIS-ZP) und Sinnvoller Text (BIS-ST)
- *Motorische Schnelligkeit (MOT)*: erfasst mittels der Kopie einer Testauswertungsschablone (Zielobjekte im Test d2), auf der lediglich die Kreise durchzustreichen waren

Die präferierte Lösung stellte ein Modell mit folgenden fünf Faktoren dar: Konzentration (KON), numerische Fähigkeiten (N), verbale Fähigkeiten (V), figurale Fähigkeiten (F) und Gedächtnis (GED). Tabelle 1.18 enthält die Ladungen der einzelnen Tests auf den Faktoren.

Die außerordentlich hohe Ladung des Zahlen-Symbol-Tests (BIS-ZS) auf dem ersten Faktor veranlassten Schmitz-Atzert et al. das Modell noch einmal *ohne* diesen Test zu prüfen, um die Frage zu klären, ob der Konzentrationsfaktor alleine durch den BIS-ZS gebildet wird. Nach Ausschluss des BIS-ZS konnte der Faktor nicht mehr repliziert werden. (Die Ladungen nach Ausschluss des BIS-ZS sind in Klammern wiedergegeben. Signifikante Ladungen sind fett gedruckt. R^2 gibt die durch die latenten Variablen aufgeklärte Testvarianz an.)

Tests	Faktoren					R^2
	KON	N	V	F	GED	
REV-T	,58 (-)	,76 (,92)	- (-)	- (-)	- (-)	,91 (,84)
BIS-XG	,14 (-)	,81 (,76)	- (-)	- (-)	-	,67 (,58)
KLT-R	,13 (-)	,70 (,68)	- (-)	- (-)	,58 (,60)	,85 (,82)
Test d2	,46 (-)	- (-)	- (-)	,65 (,80)	- (-)	,64 (,63)
FAIR	,37 (-)	- (-)	- (-)	,59 (,70)	- (-)	,48 (,49)
BIS-ZS	,92 (-)	- (-)	- (-)	,17 (-)	,26 (-)	,95 (-)
INKA	,04 (-)	- (-)	- (-)	,57 (,50)	,31 (,36)	,43 (,39)
BIS-BD	,32 (-)	- (-)	- (-)	,32 (,44)	- (-)	,20 (,19)
BIS-UW	,17 (-)	- (-)	,61 (,63)	- (-)	,24 (,26)	,46 (,47)
BIS-KW	,21 (-)	- (-)	,75 (,78)	- (-)	- (-)	,60 (,61)
BIS-TG	,15 (-)	- (-)	,55 (,57)	- (-)	- (-)	,32 (,32)

Tab. 1.18: Faktorladungen der Tests und aufgeklärte Testvarianz im 5-Faktoren-Modell

Zur Validierung der Faktoren wurden anschließend die Nicht-Konzentrationstests in die Modellbildung einbezogen. Das hierfür präferierte Modell umfasste die bereits gesicherten fünf Faktoren KON, N, V, F und GED und einen zusätzlichen Intelligenzfaktor (INT).

Tabelle 1.19 zeigt, dass die Ladungen der neu hinzugefügten Tests auf den sechs Faktoren weitgehend den Erwartungen entsprachen. (Die Tests zur Validierung sind kursiv gedruckt. Die Ladungen unter Ausschluss des BIS-ZS sind in Klammern wiedergegeben. Signifikante Ladungen sind fett gedruckt. R^2 gibt die durch die latenten Variablen aufgeklärte Testvarianz an.)

Tests	Faktoren						R^2
	KON	N	V	F	INT	GED	
REV-T	,72 (-)	,60 (,89)	- (-)	- (-)	- (-)	- (-)	,87 (<i>,79</i>)
BIS-XG	,30 (-)	,82 (,78)	- (-)	- (-)	- (-)	- (-)	,77 (<i>,61</i>)
KLT-R	-,13 (-)	,60 (,68)	- (-)	- (-)	- (-)	,74 (<i>,58</i>)	,81 (<i>,80</i>)
<i>DRE3</i>	- (-)	,59 (,60)	- (-)	- (-)	- (-)	,43 (<i>,41</i>)	,54 (<i>,52</i>)
Test d2	,58 (-)	- (-)	- (-)	,50 (,75)	- (-)	- (-)	,59 (<i>,57</i>)
FAIR	,50 (-)	- (-)	- (-)	,53 (,74)	- (-)	- (-)	,54 (<i>,54</i>)
BIS-ZS	,76 (-)	- (-)	- (-)	,05 (-)	- (-)	,16 (-)	,75 (-)
INKA	-,15 (-)	- (-)	- (-)	,49 (,46)	,35 (,38)	,32 (<i>,26</i>)	,49 (<i>,48</i>)
BIS-BD	,39 (-)	- (-)	- (-)	,29 (,47)	- (-)	- (-)	,24 (<i>,23</i>)
<i>GFT</i>	,50 (-)	- (-)	- (-)	,34 (<i>,59</i>)	- (-)	- (-)	,37 (<i>,35</i>)
<i>MOT</i>	,39 (-)	- (-)	- (-)	- (,36)	- (-)	- (-)	,15 (<i>,13</i>)
BIS-UW	,14 (-)	- (-)	,51 (,58)	- (-)	- (-)	,33 (<i>,25</i>)	,45 (<i>,40</i>)
BIS-KW	,37 (-)	- (-)	,77 (,89)	- (-)	- (-)	- (-)	,72 (<i>,79</i>)
BIS-TG	,26 (-)	- (-)	,47 (,54)	- (-)	- (-)	- (-)	,29 (<i>,29</i>)
<i>MWT-B</i>	- (-)	- (-)	,67 (,53)	- (-)	- (-)	,43	,45 (<i>,28</i>)
<i>BIS-ZP</i>	- (-)	,18 (<i>,35</i>)	- (-)	- (-)	- (-)	,43 (<i>,23</i>)	,21 (<i>,17</i>)
<i>BIS-OG</i>	- (-)	- (-)	- (-)	,32 (,31)	- (-)	,15 (<i>,09</i>)	,13 (<i>,10</i>)
<i>BIS-ST</i>	- (-)	- (-)	,37 (,45)	- (-)	- (-)	,34 (<i>,18</i>)	,25 (<i>,24</i>)
<i>BIS-ZN</i>	- (-)	,33 (,42)	- (-)	- (-)	,58 (,48)	- (-)	,44 (<i>,41</i>)
<i>BIS-SC</i>	- (-)	,31 (,37)	- (-)	- (-)	,43 (,36)	- (-)	,28 (<i>,27</i>)
<i>BIS-TM</i>	- (-)	- (-)	,51 (,51)	- (-)	,37 (,37)	- (-)	,39 (<i>,39</i>)
<i>BIS-WA</i>	- (-)	- (-)	,44 (,38)	- (-)	,36 (,47)	- (-)	,33 (<i>,36</i>)
<i>BIS-AN</i>	- (-)	- (-)	- (-)	,31 (,41)	,55 (,48)	- (-)	,40 (<i>,40</i>)
<i>BIS-CH</i>	- (-)	- (-)	- (-)	,30 (,27)	,48 (,55)	- (-)	,32 (<i>,37</i>)

Tab. 1.19: Faktorladungen der Tests und aufgeklärte Testvarianz im 6-Faktoren-Modell

Aus den Tabellen 1.18 und 1.19 ist zu ersehen, dass sich in beiden Faktorenmodellen die Struktur der Ladungen des INKA deutlich von der der übrigen Konzentrationstests unterscheidet (mit Ausnahme des KLT-R):

Im 5-Faktoren-Modell lädt INKA nur signifikant auf dem Faktor F (figurale Fähigkeiten), eine deutliche Ladung besteht außerdem auf dem Faktor GED (Gedächtnis). Besonders hervorzuheben ist hierbei, dass INKA, im Gegensatz zu den übrigen Konzentrationstests, nur eine sehr kleine, nicht signifikante Ladung auf dem Faktor zeigt, den Schmidt-Atzert et al. als Konzentrationsfähigkeit (KON) interpretieren.

Nach der Erweiterung des 5-Faktoren-Modells um einen Intelligenzfaktor, auf dem erwartungsgemäß alle Untertests des BIS zur Verarbeitungskapazität (Reasoning) signifikant laden, verändern sich die Ladungsstrukturen der Konzentrationstests nicht wesentlich, bis auf eine Ausnahme: INKA zeigt nun auch eine signifikante Ladung auf dem hinzugefügten Intelligenzfaktor.

Die durch die Studie von Schmidt-Atzert et al. gewonnenen Erkenntnisse bestätigen die Ergebnisse, die sich bislang bei der Konstruktvalidierung des INKA gezeigt haben. Die Bearbeitungsleistung im INKA steht im Zusammenhang mit:

- den Bearbeitungsleistungen im Test d2 und im FAIR
- Gedächtnisleistungen und
- Intelligenzleistungen (fluide Intelligenz)

Dass INKA nicht auf dem Faktor ‚Konzentration‘ lädt, könnte auf seine andersartige Aufgabenstruktur und seine andersartig operationalisierte Bearbeitungsleistung zurückzuführen sein. Alle Tests, die hohe Ladungen auf dem Konzentrationsfaktor zeigen, weisen folgende Gemeinsamkeiten auf: Sie geben einfache und gut wahrnehmbare Reize vor. Sie verlangen die Anwendung von einfach zu merkenden Regeln sowie die Koordination von einfachen Teiloperationen. Und, die Testleistungen werden über das (fehlerkorrigierte) *Arbeitstempo* operationalisiert (Schmidt-Atzert, Bühner und Enders, 2006). INKA hingegen ist ein Test, der aus komplexen kognitiven Aufgaben besteht, die multiple Kodierungs-, Memorierungs-, Such- und Indizierungsprozesse umfassen. Seine Testleistung wird nicht über das Arbeitstempo sondern über die *Aufgabenschwierigkeit* operationalisiert.

Abschließend sei darauf hingewiesen, dass Schmidt-Atzert et al. (2006, S. 39) ihre inhaltliche Interpretation des Konzentrationsfaktors selbst relativieren: „Nicht befriedigend zu beantworten ist die Frage, ob der Konzentrationsfaktor tatsächlich für Konzentration steht oder ob er besser mit einem der Konzentration nahe stehenden Konstrukt zu erklären ist. Alle Tests, die auf dem Faktor laden, sind Speedtests und sie bestehen aus einfachen kognitiven Aufgaben.“ Eine inhaltliche Interpretation als ‚Bearbeitungsgeschwindigkeit‘, also der operativen Fähigkeit, der der BIS-ZS im Berliner Intelligenzstruktur-Test (Jäger, Süß & Beauducel, 1997) zugeordnet ist, lassen sie außer Betracht.

6.3.3 Vorläufiges Fazit

Die bislang dargestellten Untersuchungen zur Validität des Inventars Komplexer Aufmerksamkeit geben klare Hinweise darauf, dass der Test komplexe visuelle Aufmerksamkeitsleistungen erfasst, die nicht durch kulturspezifisches Wissen oder die Beherrschung von Kulturtechniken beeinflusst werden. Im Rahmen der Untersuchungen ergaben sich weiterhin deutliche Indizien dafür, dass die kognitive Leistung, die bei der Bearbeitung der Aufgaben des INKA erbracht werden muss, im Gegensatz zu gängigen Aufmerksamkeits-tests, mit Aspekten der intellektuellen Leistungsfähigkeit bzw. kognitiven Kapazität in Zusammenhang steht. Der Gründe hierfür können in der andersartigen Aufgabenstellung beim INKA und der andersartig operationalisierten Bearbeitungsleistung liegen. Alle gängigen Aufmerksamkeits-tests sind Speed-Tests: Sie haben gleichmäßig einfache Aufgabenstellungen und die Testleistungen werden über das (fehlerkorrigierte) *Arbeitstempo* der Testpersonen operationalisiert. INKA hingegen ist ein Power-Test: Seine Aufgabenstellung ist komplex, die Aufgabenschwierigkeiten steigen systematisch an und die Testleistung wird über die *Aufgabenschwierigkeit* operationalisiert.

Eine modellgeleitete Analyse der kognitiven Prozesse, die der Aufgabenbearbeitung im INKA zugrunde liegen, wird im Teil 3 der vorliegenden Arbeit dargestellt. In diesem Zusammenhang werden auch weitere Belege für die Konstruktvalidität des Verfahrens geliefert.

6.4 Normierung

Die Normstichprobe des INKA umfasste zum Zeitpunkt des Abschlusses der vorliegenden Arbeit insgesamt 8935 Personen. Sie wurde in den Jahren 1995 bis 2004 im Rahmen der Berufsberatung und Personalauswahl durch eine nordrhein-westfälische Personalberatung und in geringem Maße (< 10%) im Rahmen der Bewerberauswahl einer hessischen Großstadt gewonnen.

Die Stichprobe baut sich folgendermaßen soziodemografisch auf (Tabelle 1.20):

	männlich	weiblich	gesamt
Geschlecht	4810 (53,8%)	4125 (46,2%)	8935 (100%)
Schulbildung			
Hauptschüler	1044 (21,7%)	668 (16,2%)	1712 (19,2%)
Realschüler	1918 (39,9%)	1876 (45,5%)	3794 (42,5%)
Gymnasiasten/Fachoberschüler	1357 (28,2%)	1397 (33,9%)	2754 (30,8%)
Akademiker	178 (3,7%)	115 (2,8%)	293 (3,3%)
ohne Abschluss/keine Angabe	313 (6,5%)	69 (1,8%)	382 (4,3%)
Alter (in Jahren)	AM = 21,2 s = 8,8 Min.: 13 Max.: 63	AM = 21,7 s = 10,9 Min.: 13 Max.: 59	AM = 21,5 s = 9,8 Min.: 13 Max.: 63

Tab. 1.20: Soziodemografische Daten der Normstichprobe des INKA

Der Mittelwert der INKA-Testrohwerte liegt für die Gesamtstichprobe bei AM = 7,67, die Standardabweichung bei s = 3,43. Der zweiseitige Kolmogorov-Smirnov-Test auf

Normalverteilung der Rohwerte ergibt ein K-S-Z von 8,1853. Dieser Wert besitzt eine Wahrscheinlichkeit von $p = ,000$. Die Nullhypothese, dass die Testrohwerte normalverteilt sind, muss damit zurückgewiesen werden. Brosius (1998) weist jedoch darauf hin, dass trotz dieser geringen Irrtumswahrscheinlichkeit es möglich ist, dass die Werte in der Grundgesamtheit annähernd normalverteilt sind. Die annähernde Normalverteilung genügt bereits als Voraussetzung für viele statistische Prozeduren. Um einen Eindruck von der Ähnlichkeit der empirischen Verteilung mit der Normalverteilung zu gewinnen, schlägt er vor, z. B. ein Histogramm mit Normalverteilungskurve zu erstellen.

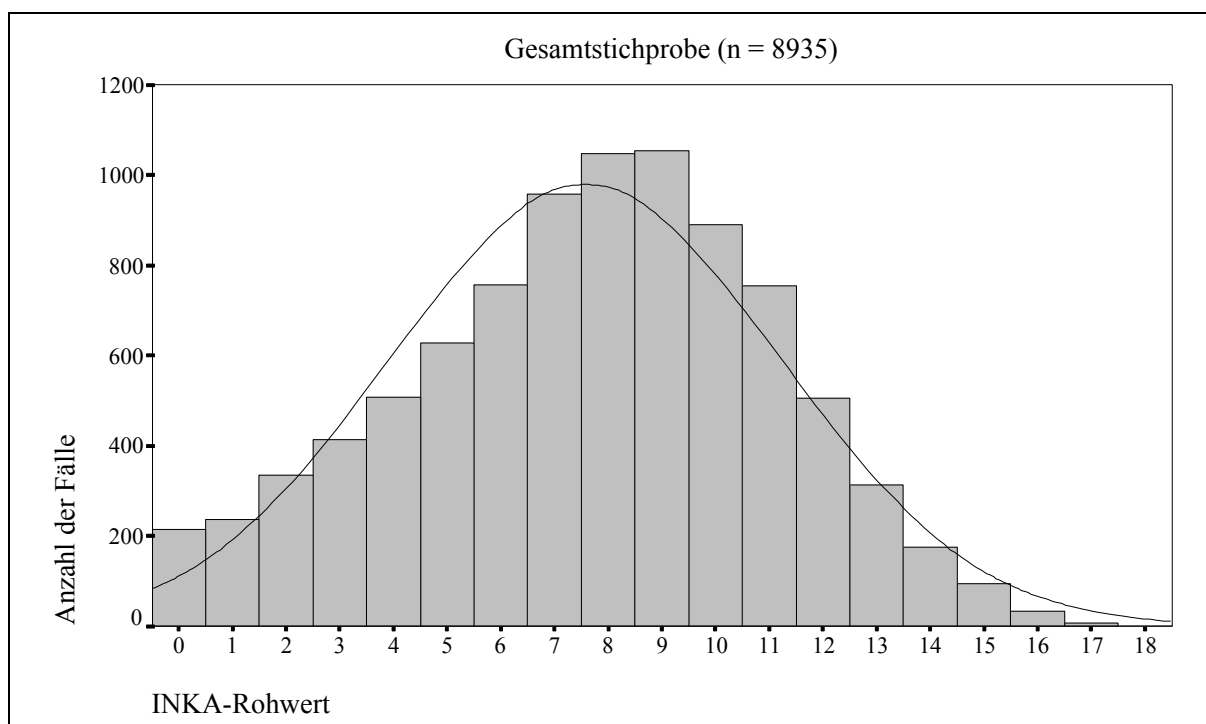


Abb. 1.14: Histogramm der Rohwertverteilung in der Normstichprobe des INKA

Das Histogramm der Rohwertverteilung in der Normierungsstichprobe in Abbildung 1.14 zeigt eine annähernde Normalverteilung der Werte. Gegenüber der Standardnormalverteilung ist die gefundene Verteilung flacher (Exzess = -,422) und aufgrund der deutlich steigenden Aufgabenschwierigkeiten (siehe Teil 1, Kapitel 5.1) erwartungsgemäß linksschief (Schiefe = -,188).

Zur Klärung der Frage, ob sich die Testergebnisse zwischen Teilgruppen der Testpersonen signifikant voneinander unterscheiden und damit teilgruppenspezifische Normen erstellt werden müssen, werden nachfolgend Mittelwertsvergleiche durchgeführt. Die Gruppierungsvariablen sind das Geschlecht, das Alter, die Schulbildung und die Händigkeit der Testpersonen.

6.4.1 Leistungsunterschiede zwischen Personen unterschiedlichen Geschlechts

Zur Überprüfung, ob signifikante Mittelwertsunterschiede in der Bearbeitungsleistung im INKA zwischen weiblichen und männlichen Testpersonen bestehen, wurde ein t-Test berechnet. Der Levene-Test auf Varianzgleichheit ergab einen F-Wert von $F = ,520$. Dieser Wert hat eine 2-seitige Wahrscheinlichkeit von $p = ,471$. Die Nullhypothese, dass die weiblichen und die männlichen Testpersonen aus Populationen mit gleichen Varianzen stammen, darf beibehalten werden. Das Ergebnis des t-Tests enthält Tabelle 1.21.

Gruppenstatistiken

Geschlecht	n	AM	s	Standardfehler des AM
weiblich	4124	7,9345	3,427	,053
männlich	4810	7,4069	3,432	,049

Test bei unabhängigen Stichproben mit gleicher Varianz

t-Wert	df	Sig. (2-seitig)	Differenz der AM	SE der Diff.	95%-KI der Diff.
7,25	8932	,000	,5277	,073	,385 - ,670

Tab. 1.21: Ergebnis des t-Tests mit der abhängigen Variablen ‚INKA-Testwert‘ und der Gruppierungsvariablen ‚Geschlecht‘

Der ermittelte t-Wert ist hoch signifikant. Die Nullhypothese, dass weibliche und männliche Testpersonen aus Populationen mit gleichen Mittelwerten stammen, muss verworfen werden. Weibliche Testpersonen schneiden signifikant besser ab als männliche Testpersonen.

6.4.2 Leistungsunterschiede zwischen Personen unterschiedlichen Alters

Zur Klärung der Frage, ob Leistungsunterschiede bei der Bearbeitung des INKA zwischen Personen unterschiedlichen Alters bestehen, wurde zunächst ein Balkendiagramm erstellt, das die mittlere Testleistung für jede Altersstufe grafisch darstellt (Abbildung 1.15).

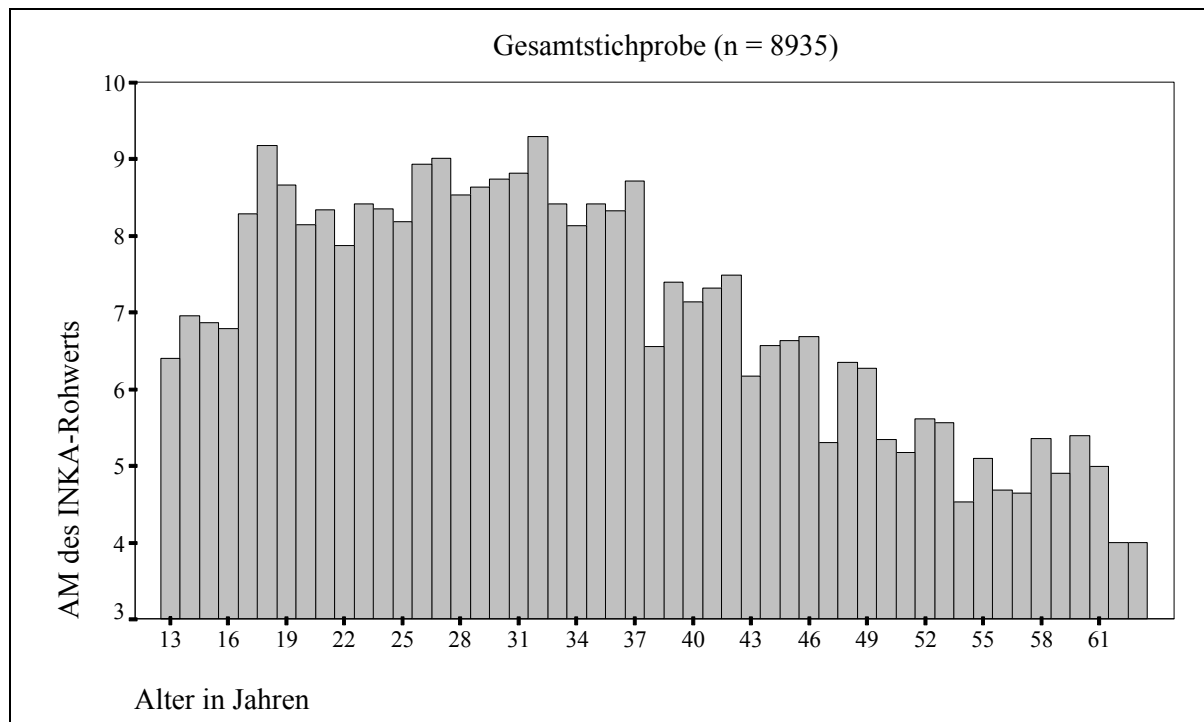


Abb. 1.15: Balkendiagramm der mittleren Testleistungen im INKA bezogen auf die Variable ‚Alter‘

Wie zu erkennen ist, steigt die Leistungsfähigkeit der Testpersonen tendenziell von 13 bis 17 Jahren an, hält sich auf diesem Niveau bis zu einem Alter von 37 Jahren und fällt anschließend bis zum 63sten Lebensjahr kontinuierlich ab. Dieser Leistungsverlauf ist kongruent mit den Erkenntnissen des altersbezogenen Verlaufs der fluiden Intelligenz sensu Cattell (Horn & Cattell, 1967; Horn & Blankson, 2005). Horn und Blankson (2005, S. 63) führen hierzu näher aus: „The systems* involved in retaining information in immediate awareness, concentration, and reasoning with novel problems decline, on average, in adulthood” (*cognitive systems, Anm. d. Verf.); und “...an ability to concentrate seems to

determine ... solving the difficult, abstract problems that characterize Gf'. Denselben Abfall der kognitiven Leistungsfähigkeit mit zunehmendem Alter zeigen auch Studien zum psychologischen Konstrukt ‚Arbeitsgedächtnis‘ (Phillips & Hamilton, 2001). Ein Modell des Arbeitsgedächtnisses wird im Teil 3 der vorliegenden Arbeit im Rahmen der modellgeleiteten Konstruktvalidierung des INKA und des ILKO ausführlich dargestellt.

Alter	13	14	15	16	17	18	19	20	21	22	23	24	25
n	5	654	1724	1222	1066	934	376	218	195	184	194	164	137
Alter	26	27	28	29	30	31	32	33	34	35	36	37	38
n	128	117	122	106	90	77	57	67	67	58	48	61	54
Alter	39	40	41	42	43	44	45	46	47	48	49	50	51
n	66	42	41	47	40	46	38	39	46	42	41	49	51
Alter	52	53	54	55	56	57	58	59	60	61	62	63	
n	47	50	28	21	16	17	17	10	5	2	1	1	

Tab. 1.22: Kategorienhäufigkeiten der Altersverteilung

Zur Überprüfung, ob die sichtbaren Mittelwertsunterschiede in der Bearbeitungsleistung zwischen den Testpersonen unterschiedlichen Alters Signifikanz besitzen, wurden in einem ersten Schritt Altersgruppen gebildet (Tabelle 1.23). Um die Standardfehler der Mittelwerte möglichst klein zu halten, bei gleichzeitiger guter Differenzierung der altersbezogenen Leistungsunterschiede, wurde der minimale Gruppenumfang auf etwa 300 Personen festgelegt.

Da die Gruppe der 13jährigen nur fünf Personen umfasste, wurde sie bei der Prüfung der Mittelwertsunterschiede nicht berücksichtigt. Aufgrund fehlender Altersangaben konnten die Ergebnisse von weiteren sieben Personen auch nicht in die Berechnungen einfließen.

Gruppe	Alter in Jahren	Anzahl	Anteil	Gruppe	Alter in Jahren	Anzahl	Anteil
1	14	654	7,3%	9	24-25	301	3,4%
2	15	1724	19,3%	10	26-28	367	4,1%
3	16	1222	13,7%	11	29-31	273	3,1%
4	17	1066	11,9%	12	32-36	297	3,3%
5	18	934	10,5%	13	37-42	311	3,5%
6	19	376	4,2%	14	43-49	292	3,3%
7	20-21	414	4,6%	15	50-63	315	3,5%
8	22-23	377	4,2%	gesamt		8923	100%

Tab. 1.23: Altersgruppen, die zur Überprüfung der Leistungsunterschiede gebildet wurden

Altersgruppe	n	AM	s	SE des AM	95%-KI für das AM
14	654	6,9572	2,9999	,1173	6,7268 - 7,1875
15	1724	6,8706	3,0309	,0730	6,7275 - 7,0138
16	1222	6,7872	3,2588	,0932	6,6043 - 6,9701
17	1066	8,2889	3,4718	,1063	8,0803 - 8,4976
18	934	9,1820	3,2249	,1055	8,9749 - 9,3891
19	376	8,6622	3,6008	,1857	8,2971 - 9,0274
20-21	414	8,2295	3,4986	,1719	7,8915 - 8,5675
22-23	377	8,1671	3,5327	,1819	7,8094 - 8,5249
24-25	301	8,2724	3,3604	,1937	7,8913 - 8,6536
26-28	367	8,8256	3,5052	,1830	8,4658 - 9,1854
29-31	273	8,7216	3,2124	,1944	8,3389 - 9,1044
32-36	297	8,5084	3,4933	,2027	8,1095 - 8,9073
37-42	311	7,4791	3,4742	,1970	7,0915 - 7,8667
43-49	292	6,2671	3,3864	,1982	5,8771 - 6,6572
50-63	315	5,2095	3,1583	,1780	4,8594 - 5,5597
gesamt	8923	7,6519	3,4380	,0364	7,5806 - 7,7233

Tab. 1.24: Deskriptive Statistiken des INKA-Rohwerts für die 15 Altersgruppen

Aus den deskriptiven Statistiken in Tabelle 1.24 lässt sich bereits erkennen, dass die Nullhypothese, die mittlere Testleistungen der unterschiedlichen Altersgruppen seien in der Grundgesamtheit gleich, offenbar nicht zutrifft. Der Levene-Test auf Gleichheit der Gruppenvarianzen ergab einen F-Wert von $F = 3,8235$ ($df_1 = 14$, $df_2 = 8908$). Dieser Wert besitzt eine 2-seitige Wahrscheinlichkeit von $p = ,000$. Die Nullhypothese, dass alle Gruppenvarianzen in der Grundgesamtheit gleich seien, muss zurückgewiesen werden. Statt einer Varianzanalyse muss zur statistischen Überprüfung der Mittelwertsunterschiede ein verteilungsfreies Verfahren, wie z. B. der Kruskal-Wallis-Test, eingesetzt werden (Bortz, 1993).

Das Kruskal-Wallis-H, das annähernd χ^2 -verteilt ist, überprüft die Nullhypothese, dass alle Stichproben derselben Grundgesamtheit entstammen. Zur Überprüfung der Hypothese werden zuerst alle Werte in eine gemeinsame Rangordnung gebracht. Anschließend wird mit einem χ^2 -Test überprüft, ob die durchschnittlichen Rangwerte in den einzelnen Stichproben gleich groß sind (Brosius, 1998). Tabelle 1.25 enthält die mittleren Ränge der Testleistungen im INKA für die vorliegenden Altersgruppen.

Altersgruppe	n	mittlerer Rang	Altersgruppe	n	mittlerer Rang
14	654	3864,34	24-25	301	4957,07
15	1724	3842,51	26-28	367	5366,39
16	1222	3806,39	29-31	273	5255,05
17	1066	4971,94	32-36	297	5147,85
18	934	5618,33	37-42	311	4345,59
19	376	5257,88	43-49	292	3449,77
20-21	414	4913,29	50-63	315	2686,48
22-23	377	4834,08	gesamt	8923	

Tab. 1.25: Mittlere Ränge der INKA-Testleistungen für die 15 Altersgruppen

Für die mittleren Ränge ergibt sich ein hinsichtlich verbundener Ränge korrigierter χ^2 -Wert von 805,3173 ($df = 14$). Dieser Wert besitzt eine Wahrscheinlichkeit von $p = ,000$. Die

Nullhypothese muss zurückgewiesen werden: Die mittleren Ränge der Testleistung im INKA sind, bezogen auf die Altersgruppen, in der Grundgesamtheit nicht gleich.

6.4.3 Leistungsunterschiede zwischen Personen unterschiedlicher Schulbildung

Um einen ersten Überblick darüber zu erhalten, ob Unterschiede in der Bearbeitungsleistung beim INKA zwischen Personen mit unterschiedlicher Schulbildung bestehen, wurde ein Balkendiagramm erstellt (Abbildung 1.16). Das Diagramm stellt die mittlere Testleistung für jede Bildungsgruppe grafisch dar. Die Gruppe von Testpersonen, für die keine Angaben über die Schulbildung vorlagen, wurde nicht in die Analysen einbezogen.

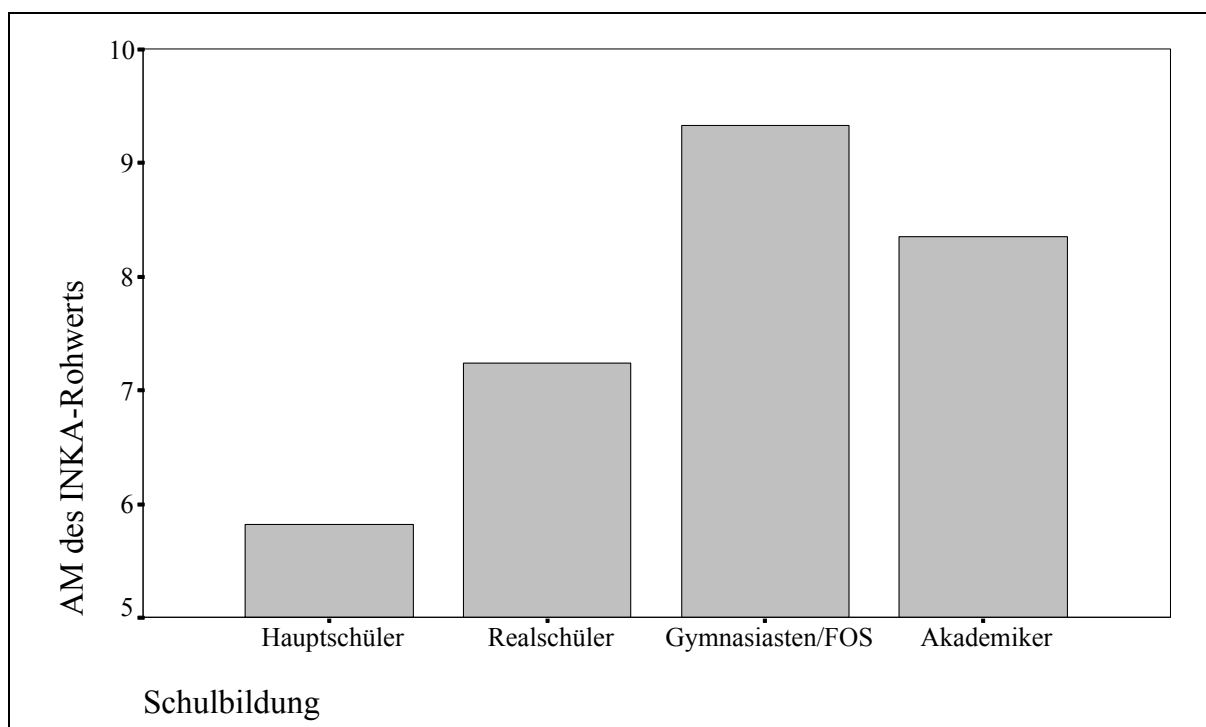


Abb. 1.16: Balkendiagramm der mittleren Testleistungen im INKA bezogen auf die Variable ‚Schulbildung‘

Die Bearbeitungsleistung im INKA weist deutliche Unterschiede zwischen den vier Bildungsgruppen auf. Der Leistungsabfall der Akademiker gegenüber den Gymnasiasten und

Fachoberschülern könnte ein weiterer Beleg dafür sein, dass kulturspezifisches Wissen und die Beherrschung von Kulturtechniken keinen Einfluss auf die Testleistung im INKA haben. Tabelle 1.26 enthält die dem Balkendiagramm zugrunde liegenden Daten.

Schulbildung	n	AM	s	SE des AM	95%-KI für das AM
Hauptschüler	1712	5,8236	3,2624	,0788	5,6690 - 5,9782
Realschüler	3794	7,2359	3,1672	,0514	7,1351 - 7,3367
Gymnasiasten/FOS	2754	9,3348	3,1315	,0597	9,2178 - 9,4518
Akademiker	293	8,3549	3,3254	,1943	7,9726 - 8,7373
gesamt	8553	7,6674	3,4333	,0371	7,5946 - 7,7401

Tab. 1.26: Deskriptive Statistiken des INKA-Rohwerts für die Schulbildungsgruppen

Zur Überprüfung, ob die Unterschiede der arithmetischen Mittel der INKA-Testleistungen statistische Signifikanz besitzen, soll eine einfaktorielles Varianzanalyse durchgeführt werden. Der Levene-Test auf Gleichheit der Gruppenvarianzen ergibt einen F-Wert von $F = 6,4636$ ($df_1 = 3$, $df_2 = 8549$). Dieser Wert besitzt eine 2-seitige Wahrscheinlichkeit von $p = ,000$. Die Nullhypothese, dass alle Gruppenvarianzen in der Grundgesamtheit gleich seien, muss zurückgewiesen werden. Statt einer Varianzanalyse wird zur statistischen Überprüfung der Leistungsunterschiede der Kruskal-Wallis-Test durchgeführt.

Schulbildung	n	mittlerer Rang
Hauptschüler	1712	2977,46
Realschüler	3794	3945,05
Gymnasiasten/FOS	2754	5487,87
Akademiker	293	4786,23
gesamt	8553	

Tab. 1.27: Mittlere Ränge der INKA-Testleistungen für die Schulbildungsgruppen

Für die mittleren Ränge der Testleistungen (Tabelle 1.27) ergibt sich ein χ^2 -Wert von 1217,5458 (df = 3). Dieser Wert besitzt eine Wahrscheinlichkeit von $p = ,000$. Die Nullhypothese muss zurückgewiesen werden: Die mittleren Ränge der Testleistung im INKA sind, bezogen auf die Bildungsgruppen, in der Grundgesamtheit nicht gleich.

Da die bisherigen Untersuchungsergebnisse zur Validität des INKA (siehe Teil 1, Kapitel 6.3.2) zeigen, dass die Testleistung nicht durch die Schulbildung beeinflusst wird, müssen die Leistungsunterschiede zwischen den Bildungsgruppen anders begründet sein.

6.4.4 Zusammenhang zwischen Schulbildung und Alter

Wie eben dargestellt, zeigt sich ein Abfall der mittleren Testleistung bei Akademikern ggü. Gymnasiasten und Fachoberschülern. Falls die Schulbildung einen positiven Einfluss auf die Testleistungen im INKA hätte, dürfte sich ein derartiger Leistungsabfall nicht ergeben.

Zur Überprüfung, ob der Unterschied zwischen den mittleren Testleistungen der beiden Bildungsgruppen Signifikanz besitzt, wird zunächst ein t-Test berechnet (Tabelle 1.28). Der Levene-Test auf Varianzgleichheit ergibt einen F-Wert von $F = 1,761$. Dieser Wert hat eine 2-seitige Wahrscheinlichkeit von $p = ,185$. Die Nullhypothese, dass die beiden Bildungsgruppen aus Populationen mit gleichen Varianzen stammen, darf beibehalten werden.

Gruppenstatistiken

Schulbildung	n	AM	s	Standardfehler des AM
Gymnasiasten/FOS	2754	9,3348	3,132	,060
Akademiker	293	8,3549	3,325	,194

Test bei unabhängigen Stichproben mit gleicher Varianz

t-Wert	df	Sig. (2-seitig)	Differenz der AM	SE der Diff.	95%-KI der Diff.
5,06	3045	,000	,9798	,194	,600; 1,360

Tab. 1.28: Ergebnis des t-Tests mit der abhängigen Variablen ‚INKA-Testwert‘ und der Gruppierungsvariablen ‚Bildung: Gymnasiasten/Fachoberschüler – Akademiker‘

Der ermittelte t-Wert besitzt eine 2-seitige Wahrscheinlichkeit von $p = ,000$. Die Nullhypothese, dass beide Bildungsgruppen aus Populationen mit gleichen Mittelwerten stammen, muss verworfen werden. Der Leistungsabfall bei Akademikern gegenüber Gymnasiasten und Fachoberschülern ist hoch signifikant.

In Teil 1, Kapitel 6.4.2 wurde aufgezeigt, dass die Testleistung im INKA vom 13ten bis 18ten Lebensjahr ansteigt, sich auf diesem Niveau bis zu einem Alter von 37 Jahren hält und anschließend kontinuierlich bis zum 63sten Lebensjahr wieder abfällt. Dies begründet die Vermutung, dass der Leistungsunterschied zwischen Akademikern und Gymnasiasten/Fachoberschülern auf das Alter der Testpersonen zurückzuführen sein könnte.

Schulbildung	n	Alter	
		AM	s
Hauptschüler	1711	22,7	11,1
Realschüler	3793	19,6	9,4
Gymnasiasten/FOS	2753	20,5	6,8
Akademiker	293	39,9	10,2
gesamt	8550	21,2	9,8

Tab. 1.29: Durchschnittliches Alter der Testpersonen bezogen auf die Schulbildung

Tabelle 1.29 zeigt, dass die Akademiker in der vorliegenden Stichprobe mit Abstand am ältesten sind und sich die anderen drei Bildungsgruppen, bezogen auf das durchschnittliche Alter, relativ wenig unterscheiden.

Zur Überprüfung, ob der Altersunterschied zwischen Akademikern und Gymnasiasten/Fachoberschülern Signifikanz besitzt, wird ein t-Test berechnet (Tabelle 1.30). Der Levene-Test auf Varianzgleichheit ergibt einen F-Wert von $F = 188,328$. Dieser Wert hat eine 2-seitige Wahrscheinlichkeit von $p = ,000$. Die Nullhypothese, dass die beiden Bildungsgruppen aus Populationen mit gleichen Varianzen stammen, muss verworfen und ein t-Test für unabhängige Stichproben mit ungleicher Varianz berechnet werden.

Gruppenstatistiken

Bildung	n	AM	s	Standardfehler des AM
Gymnasiasten/FOS	2753	20,4519	6,788	,129
Akademiker	293	39,9317	10,218	,597

Test bei unabhängigen Stichproben mit ungleicher Varianz

t-Wert	df	Sig. (2-seitig)	Differenz der AM	SE der Diff.	95%-KI der Diff.
-31,98	320,00	,000	-19,4799	,611	-20,682; -18,278

Tab. 1.30: Ergebnis des t-Tests mit der abhängigen Variablen ‚Alter‘ und der Gruppierungsvariablen ‚Bildung: Gymnasiasten/Fachoberschüler – Akademiker‘

Der ermittelte t-Wert besitzt eine 2-seitige Wahrscheinlichkeit von $p = ,000$. Die Nullhypothese, dass beide Bildungsgruppen aus Populationen mit gleichen Mittelwerten stammen, muss verworfen werden. Der Altersunterschied zwischen Akademikern und Gymnasiasten/Fachoberschülern ist hoch signifikant.

Der Leistungsabfall bei der Bearbeitung des INKA bei Akademikern gegenüber Gymnasiasten und Fachoberschülern ist offensichtlich auf den Altersunterschied zwischen den Personen dieser beiden Bildungsgruppen zurückzuführen. Dieses Ergebnis ist kongruent mit dem altersbezogenem Verlauf der Testleistung (siehe Teil 1, Kapitel 6.4.2).

Die ansteigende Testleistung zwischen Personen mit Hauptschulabschluss, Realschulabschluss und Abitur bzw. Fachhochschulreife ist offenbar nicht durch das Alter der Testpersonen begründet (siehe Tabelle 1.29 und Abbildung 1.15). Die bisherigen Untersuchungsergebnisse weisen auch darauf hin, dass die Testleistung im INKA unabhängig von der schulischen Ausbildung ist (siehe Teil 1, Kapitel 6.3.2.1 und Kapitel 6.3.2.2). Die ansteigende Testleistung bei den drei Schulbildungsgruppen muss auf einem weiteren Faktor beruhen. Es besteht hierbei die Vermutung, dass im Wege einer Fremd- oder Selbstselektion der Schullaufbahn, die besuchte Schulform mit dem individuellen Niveau der kognitiven Leistungsfähigkeit in Zusammenhang steht.

Um diese Vermutung zu Überprüfen, wurde die Stichprobe von Personen herangezogen, die INKA zusammen mit den SPM bearbeitet hatten (siehe Teil 1, Kapitel 6.3.2.3). Die SPM sind ein sprachfreies, kulturunabhängiges Verfahren zur Messung der fluiden Intelligenz. Falls sich für die Testleistung in den SPM zwischen den vier Schulbildungsgruppen ein ähnlicher Leistungsverlauf wie beim INKA zeigen sollte, wäre dies ein Hinweis auf die Gültigkeit der Annahme, dass die besuchte Schulform mit der kulturunabhängigen kognitiven Leistungsfähigkeit in Zusammenhang steht.

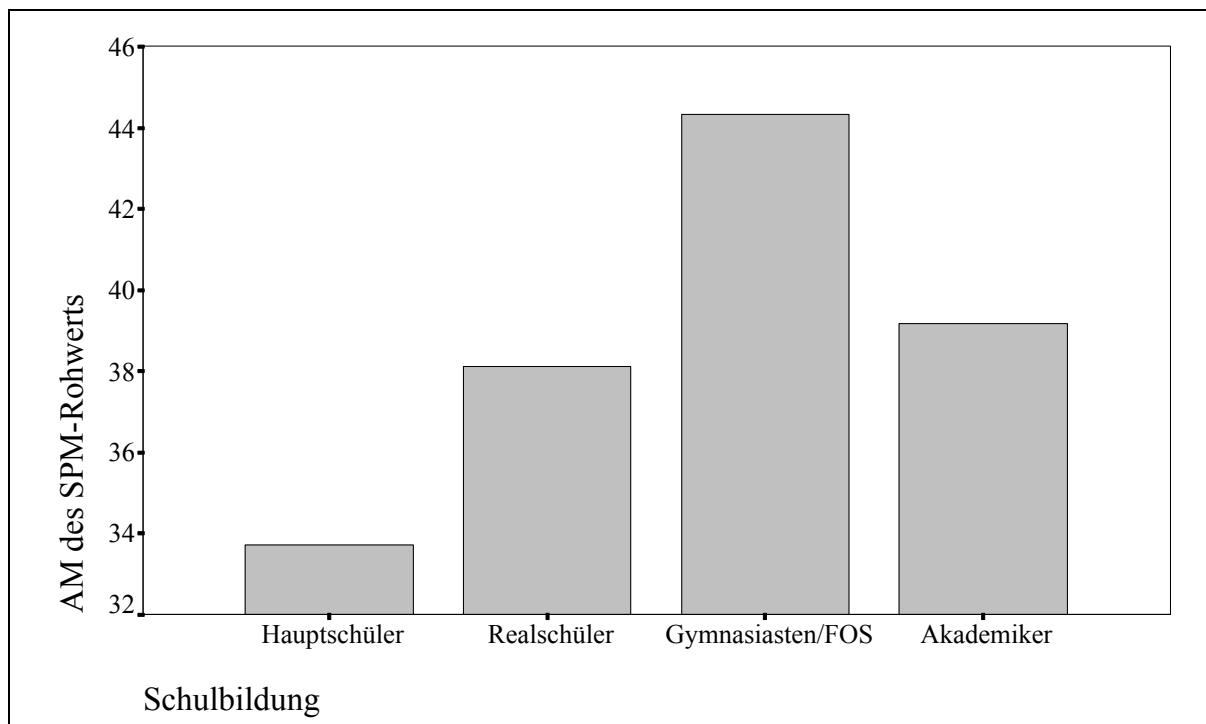


Abb. 1.17: Balkendiagramm der mittleren Testleistungen in den SPM bezogen auf die Variable ‚Schulbildung‘

Das Balkendiagramm in Abbildung 1.17 zeigt einen ähnlichen Verlauf der Testleistung in den SPM, bezogen auf die Schulbildung, wie beim INKA.

Die Frage, ob die Leistungsunterschiede in der Bearbeitung der SPM zwischen Personen mit unterschiedlicher Schulbildung Signifikanz besitzen, soll mit Hilfe einer einfaktoriellen Varianzanalyse geklärt werden.

Tabelle 1.31 enthält die deskriptiven Statistiken des SPM-Rohwerts für die vier Bildungsgruppen:

Schulbildung	n	AM	s	SE des AM	95%-KI für das AM
Hauptschüler	26	33,7308	3,6612	,7180	32,2520; 35,2096
Realschüler	134	38,1194	4,9113	,4243	37,2802; 38,9586
Gymnasiasten/FOS	221	44,3394	4,0550	,2728	43,8018; 44,8769
Akademiker	6	39,1667	6,2743	2,5615	32,5823; 45,7510
gesamt	387	41,3928	5,6377	,2866	40,8293; 41,9562

Tab. 1.31: Deskriptive Statistiken des SPM-Rohwerts für die vier Bildungsgruppen

Der Levene-Test auf Gleichheit der Gruppenvarianzen ergibt einen F-Wert von $F = 1,0697$ ($df_1 = 3$, $df_2 = 383$). Dieser Wert besitzt eine 2-seitige Wahrscheinlichkeit von $p = ,362$. Die Nullhypothese, dass alle Gruppenvarianzen in der Grundgesamtheit gleich seien, darf beibehalten und es darf eine Varianzanalyse berechnet werden. Das Ergebnis enthält Tabelle 1.32.

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
zwischen den Gruppen	4910,7140	3	1636,9047	85,2093	,0000
innerhalb der Gruppen	7357,5858	383	19,2104		
Gesamt	12268,2997	386			

Tab. 1.32: Ergebnis der einfaktoriellen Varianzanalyse mit der abhängigen Variablen ‚SPM-Testwert‘ und der Faktorvariablen ‚Schulbildung‘

Der für den Gruppenvergleich ermittelte F-Wert von $F = 85,2093$ besitzt eine Auftretenswahrscheinlichkeit von $p = ,000$. Die Nullhypothese, dass die Mittelwerte der SPM-Testleistungen der vier Gruppen in der Grundgesamtheit identisch sind, muss zurückgewiesen werden.

Um zu klären, zwischen welchen Gruppen signifikante Leistungsunterschiede bestehen, wurden Einzelvergleiche mit Hilfe des Scheffé-Tests durchgeführt. Das Ergebnis zeigt Tabelle 1.33.

	Hauptschüler	Realschüler	Akademiker
Hauptschüler			
Realschüler	*		
Gymnasiasten/FOS	*	*	*
Akademiker			

Tab. 1.33: Ergebnis der Einzelvergleiche mit Hilfe des Scheffé-Tests

Der Scheffé-Test ergibt, dass zwischen den Gruppen ‚Hauptschüler‘ und ‚Realschüler‘ und zwischen der Gruppe ‚Gymnasiasten/FOS‘ und den restlichen drei Gruppen jeweils ein signifikanter Unterschied in den SPM-Testleistungen auf dem 5%-Niveau besteht.

Das Ergebnis der Varianzanalyse bestätigt die Vermutung, dass die Unterschiede in der Bearbeitungsleistung beim INKA mit der fluiden Intelligenz der Testpersonen in Zusammenhang stehen.

6.4.5 Leistungsunterschiede zwischen Personen unterschiedlicher Händigkeit

Für INKA existiert eine Linkshänderversion des Testbogens mit einem veränderten grafischen Aufgabendesign (siehe Teil 1, Kapitel 4.2). Das spezielle Design soll eine mögliche Benachteiligung von Linkshändern bei der Aufgabenbearbeitung ausschließen. Die Erfahrungen bei der Testdurchführung zeigen, dass linkshändige Personen das Angebot des Linkshänder-Testbogens positiv auf- und annehmen.

Mit dem Begriff ‚Händigkeit‘ wird das Phänomen beschrieben, dass eine Person durchwegs dieselbe Hand für Aufgaben verwendet, bei denen Hand- und Fingerfertigkeit erforderlich sind und nur eine Hand benutzt werden muss, z. B. beim Schreiben, Ballwerfen oder

Hämmern (Coren, 2002). Die Angaben über den Anteil der Linkshänder an der Gesamtbevölkerung schwanken etwa um 10% (Cohen, 2002; Geschwind & Crabtree, 2003; Sattler, 2000) und zeigen keine signifikanten Differenzen zwischen unterschiedlichen Kulturen (Geschwind & Crabtree, 2003). Die Gründe für die Schwankungen bei den Angaben sind vielfältig. Sie liegen z. B. in den eingesetzten Messmethoden (Selbsteinschätzung, Befragung, Beobachtung, Test) oder in der Operationalisierung der Linkshändigkeit, z. B. die Frage, zu welcher Gruppe umgeschulte Linkshänder gezählt werden (Coren, 2002).

Zur Überprüfung der Nullhypothese, dass zwischen den Mittelwerten der Testleistungen von Personen unterschiedlicher Händigkeit in der Grundgesamtheit kein Unterschied besteht, soll ein t-Test durchgeführt werden. Grundlage für diese und die folgenden Berechnungen ist die Stichprobe der faktorenanalytischen Konstruktvalidierung (siehe Teil 1, Kapitel 6.3.1). Von den 615 Testpersonen benutzten 48 (7,8%) die Linkshänderversion, 567 (92,2%) die Rechtshänderversion. Der Linkshänderanteil in der vorliegenden Stichprobe entspricht also in etwa ihrem Anteil in der Gesamtbevölkerung, zumal man davon ausgehen darf, dass auch umgeschulte Linkshänder die Rechtshänderversion des INKA-Testbogens verwendet haben.

Der Levene-Test auf Varianzgleichheit ergab einen F-Wert von $F = ,000$. Dieser Wert hat eine 2-seitige Wahrscheinlichkeit von $p = ,997$. Die Nullhypothese, dass Verwender der Links- und Verwender der Rechtshänderversion des Tests aus Populationen mit gleichen Varianzen stammen, darf beibehalten werden.

Gruppenstatistiken

Händigkeit	n	AM	s	Standardfehler des AM
Linkshänder	48	8,3958	3,841	,554
Rechtshänder	567	8,3792	3,677	,154

Test bei unabhängigen Stichproben mit gleicher Varianz

t-Wert	df	Sig. (2-seitig)	Differenz der AM	SE der Diff.	95%-KI der Diff.
,03	613	,976	,0166	,555	-1,073; 1,106

Tab. 1.34: Ergebnis des t-Tests mit der abhängigen Variablen ‚INKA-Testwert‘ und der Gruppierungsvariablen ‚Händigkeit‘

Der ermittelte t-Wert (Tabelle 1.34) besitzt eine 2-seitige Wahrscheinlichkeit von $p = ,976$. Die Nullhypothese, dass Verwender der Links- und Verwender der Rechtshänderversion des Tests aus Populationen mit gleichen Mittelwerten stammen, darf beibehalten werden.

Zur Überprüfung, ob sich die Mittelwerte der Leistungen in den einzelnen Aufgaben zwischen der Rechtshänder- und der Linkshänderversion unterscheiden, wurde für jede der 18 Aufgaben ein t-Test berechnet. Zur Überprüfung der Gleichheit der Varianzen diente der Levene-Test. Bei einer Wahrscheinlichkeit des resultierenden F-Werts von $p \geq ,05$ wurde die Nullhypothese, dass die Verwender der Links- und die Verwender der Rechtshänderversion aus Populationen mit gleichen Varianzen stammen, beibehalten. Entsprechend wurde anschließend entweder ein t-Test für unabhängige Stichproben mit gleicher oder mit ungleicher Varianz berechnet.

Aufgabe	AM Rechtshänder	AM Linkshänder	t-Wert	Sig. (2-seitig)	Varianzgleichheit
1	,9206	,8958	-,60	,547	ja
2	,7090	,8125	1,72	,090	nein
3	,8183	,8542	,62	,535	ja
4	,6614	,7500	1,34	,186	nein
5	,4303	,5000	,93	,351	ja
6	,6437	,6875	,61	,543	ja
7	,6473	,6250	-,31	,757	ja
8	,4444	,4375	-,09	,926	ja
9	,6772	,6042	-1,03	,301	ja
10	,3986	,3125	-1,22	,228	nein
11	,5026	,4792	-,31	,755	ja
12	,4638	,4792	,20	,838	ja
13	,4109	,3542	-,77	,443	ja
14	,3263	,3333	,10	,920	ja
15	,1517	,1250	-,50	,619	ja
16	,0847	,0625	-,53	,594	ja
17	,0653	,0417	-,64	,520	ja
18	,0176	,0417	,81	,422	nein

Tab. 1.35: Ergebnisse der t-Tests mit den abhängigen Variablen ‚INKA-Aufgabenwert‘ und der Gruppierungsvariablen ‚Händigkeit‘

Tabelle 1.35 zeigt, dass die Unterschiede der durchschnittlichen Leistungen zwischen der Rechtshänder- und der Linkshänderversion bei allen 18 Aufgaben keine Signifikanz besitzen.

Die Bearbeitungsleistungen in der Rechtshänder- und der Linkhänderversion des INKA unterscheiden sich nicht. Dieses Ergebnis gilt sowohl für den Gesamttest als auch für jede einzelne der 18 Aufgaben.

6.5 Skalierung

Aufgrund der nachgewiesenen Rasch-Skalierung des INKA ist gewährleistet, dass die Testwerte adäquate Abbildungen der Ausprägung der latenten Persönlichkeitseigenschaft darstellen.

6.6 Unverfälschbarkeit

Wenn man bei einem Leistungstest von Verfälschbarkeit spricht, muss in erster Linie in Erwägung gezogen werden, inwieweit eine „Leistung“ (richtige Antwort) von der Testperson gezeigt werden kann, ohne eigentlich eine Leistung erbracht zu haben. Verfälschbarkeit sollte hier also darauf bezogen werden, wie wahrscheinlich es ist, allein durch Raten oder Zufall zu einer richtigen Lösung zu kommen. Wie bereits in Teil 1, Kapitel 4 ausführlich dargestellt, sind die Aufgaben im INKA so gestaltet, dass die Wahrscheinlichkeit für eine Testperson gegen Null läuft, durch zufälliges Notieren von (Lösungs-)Buchstaben zu einer richtigen Aufgabenlösung zu kommen. Insofern lassen sich die Testergebnisse als nicht verfälschbar bezeichnen, zumal, wie Kubinger (2003a, S. 432) zur Verfälschbarkeit von Leistungstests schreibt, „der Fall, eine Tp (Testperson, Anm. d. Verf.) erbringt absichtlich schlechte Leistungen, zwar möglich, aber kaum realistisch ist“.

6.7 Zumutbarkeit

Unter Zumutbarkeit wird verstanden „das Ausmaß, in dem ein Test (absolut und relativ zu dem aus der Anwendung des Verfahrens resultierenden Nutzen) die getestete Person in zeitlicher, psychischer (insbesondere ‚energetisch‘-motivational und emotional) sowie körperlicher Hinsicht beansprucht“ (Testkuratorium, 1986, S. 359).

INKA ist ein Verfahren, das mit einer reinen Testdauer von 15 Minuten die Testpersonen zeitlich nur gering beansprucht. Man darf davon ausgehen, dass die Aufgabenbearbeitung in psychischer Hinsicht aufgrund der Komplexität und ansteigenden Schwierigkeiten der Aufgaben von den Testpersonen als nicht monoton empfunden wird. Ein Leistungstest wie INKA wirkt sich am ehesten als beanspruchend, weil Kräfte zehrend aus, wenn die Bearbeitung einer Aufgabe als schwierig empfunden wird. Da die ansteigenden Schwierigkeiten der Aufgaben stark streuen, haben selbst leistungsschwache Testpersonen eine gute Chance, wenigstens einige Aufgaben komplett zu bearbeiten. Insofern darf man davon ausgehen, dass die Bearbeitung des INKA nicht ungewöhnlich häufig Versagensängste auslösen wird. In körperlicher Hinsicht ist lediglich für die Testbearbeitung nötig, dass die Testperson zumindest 15 Minuten an einem Tisch sitzt und eine begrenzte Menge an Buchstaben notiert. INKA darf als zumutbar gelten. Das bestätigen auch Beobachtungen des Verhaltens der Testpersonen während der Testdurchführungen.

6.8 Fairness

Die Fairness eines Tests beschreibt, inwiefern Testpersonen nicht aufgrund einer bestimmten Gruppenzugehörigkeit systematisch diskriminiert werden.

Um die Instruktion des INKA verstehen zu können, sollten die Testpersonen ausreichende Kenntnisse der deutschen Sprache besitzen. Eine mangelhafte Beherrschung der deutschen Sprache führt jedoch nicht unmittelbar dazu, dass Testpersonen bei der Testbearbeitung benachteiligt sind. Der Testleiter muss Fragen zur Aufgabenbearbeitung klären und darf erst dann die Bearbeitung der Aufgaben einleiten, wenn von den Testpersonen verstanden wurde, was genau zu tun ist. Da die Suchreihen und die Buchstabenvorgaben aus lateinischen Konsonanten bestehen, sollten die Testpersonen zumindest das lateinische Alphabet beherrschen. Um eine mögliche Benachteiligung von Linkshändern bei der Testbearbeitung zu vermeiden, wurde für diese Personengruppe ein eigenes grafisches Aufgabendesign entwickelt (siehe Teil 1, Kapitel 4.2 und Kapitel 6.4.5). INKA kann, sofern die Testperson das lateinische Alphabet beherrscht und die deutsche Sprache ausreichend versteht, als fairer Test gelten.

7 Durchführung des Tests

INKA ist sowohl in Einzel- als auch in Gruppenuntersuchungen einsetzbar. Werden Linkshänder getestet, so sollten diese Personen Linkshänder-Testbögen erhalten. Wegen seiner relativ kurzen Durchführungsdauer eignet sich das Verfahren gut für die Verwendung im Rahmen von Testbatterien.

7.1 Instruieren der Testpersonen

Zunächst müssen die Testpersonen auf der ersten Seite des Testbogens Angaben zu ihrer Person machen. Im Einzelnen werden abgefragt: Name, Vorname, Alter in Jahren, Geschlecht und Schulbildung. Bei Geschlecht und Schulbildung sind entsprechende Auswahlmöglichkeiten anzukreuzen. Die Schulbildung ist der Schulabschluss, der aktuell angestrebt wird oder bereits erzielt wurde. Sie gliedert sich auf in Hauptschule, mittlerer Bildungsweg (Realschulabschluss, der Abschluss einer Berufsfachschule und der Abschluss der 10. Klasse des Gymnasiums), Abitur/Fachhochschulreife und Hochschulabschluss. Die personenbezogenen Angaben werden zur korrekten Zuordnung der Ergebnisse zu den Personen und zur Standardisierung der Testergebnisse benötigt.

Die Testpersonen sollen anschließend die schriftliche Instruktion auf der Titelseite des Testbogens durchlesen. Sie beschreibt schrittweise das Vorgehen bei der Aufgabenbearbeitung. Danach erhalten die Testpersonen die Möglichkeit, Verständnisfragen zu stellen. Verständnisfragen sollen vom Testleiter in größtmöglicher Anlehnung an die schriftliche Instruktion beantwortet werden. Wenn sichergestellt ist, dass die Testpersonen die Art und Weise der Aufgabenbearbeitung verstanden haben, sollen sie zu Übungszwecken die zweite Aufgabe im Aufgabenbeispiel selbständig bearbeiten. Nach ihrer Bearbeitung hat der Testleiter die richtige Lösung zu benennen. Er nennt weiterhin die Anzahl der im Test zu bearbeitenden Aufgaben sowie die dafür zur Verfügung stehende Zeit. Abschließend muss der Testleiter deutlich darauf hinweisen, dass bis auf das Eintragen der Lösungsbuchstaben nichts auf den Antwortbogen geschrieben werden darf und auch keine Markierungen in den Suchreihen angebracht werden dürfen. Der Testleiter muss auch deutlich darauf hinweisen, dass eine Aufgabe nur dann als richtig gelöst gilt, wenn alle umgewandelten Vorgaben in der

entsprechenden Suchreihe entdeckt und alle Lösungsbuchstaben in der richtigen Reihenfolge notiert wurden.

Werden bei einer Gruppentestung neben Rechtshänder-Testbögen auch Linkshänder-Testbögen eingesetzt, so sollte der Testleiter, um Irritationen bei den Testpersonen zu vermeiden, unmittelbar nach dem Durchlesen der Instruktion kurz auf den Unterschied zwischen Rechts- und Linkshänderversion hinweisen (das Vertauschen der Spalten ‚Vorgaben‘ und ‚Buchstaben‘).

7.2 Auswertung der Ergebnisse

Bearbeitete Testbögen werden mit Hilfe einer Lösungsschablone ausgewertet. Die Lösungsschablone enthält für jede Aufgabe die korrekte Lösungsbuchstabenreihe. Die Schablone wird an den Rand des Testbogens angelegt, an dem sich die Spalte für die Eintragung der Lösungsbuchstaben befindet. Die Lösungen der Testperson werden mit den Lösungen auf der Schablone verglichen und für jede richtig gelöste Aufgabe ein Punkt vergeben. Die Summe der erzielten Punkte bildet den Rohwert. Dieser wird anschließend mit Hilfe von Normtabellen standardisiert.

7.3 Interpretation der Ergebnisse

Mit Hilfe der Untersuchungen zur Validität konnte aufgezeigt werden, dass das Konstrukt ‚selektive visuelle Aufmerksamkeit‘ als Messgegenstand des INKA Gültigkeit besitzt. Das Verfahren erfasst - aufgabenbedingt - zudem kurzzeitige Merkfähigkeit. INKA folgt inhaltlich den Kapazitätsmodellen der Aufmerksamkeit. Die grundlegende Annahme dieser Modelle ist, dass die Ressourcen der bewussten kognitiven Informationsverarbeitung begrenzt sind. Je schwieriger eine Aufgabe, desto mehr Ressourcen müssen ihr über Aufmerksamkeitsprozesse zugeteilt werden. Bei Aufgaben, die die Ressourcen übersteigen, kommt es zu einer Verschlechterung der Leistung: die Aufgabenbearbeitung verlangsamt sich oder es steigt die Anzahl der Fehler bei der Aufgabenbearbeitung. Beide Effekte können bei der Bearbeitung des INKA festgestellt werden.

Die Aufgaben des INKA sind aufgrund ihrer Rasch-Skalierung homogen und voneinander lokal stochastisch unabhängig (siehe Teil 1, Kapitel 5.2.3). INKA liefert folglich eine erschöpfende Statistik: Die Anzahl der gelösten Testaufgaben charakterisiert erschöpfend die Ausprägung der zugrunde liegenden Leistungsmerkmale.

Die Bedeutung der individuellen Testleistung einer Person (ihr Testrohwert) ergibt sich durch ihre Lokalisierung innerhalb der entsprechenden Normstichprobe. Die Standardisierung des Testwertes muss, getrennt für weibliche und männliche Personen, altersbezogen erfolgen (siehe Teil 1, Kapitel 6.4.1 und Kapitel 6.4.2).

Teil 2: Entwicklung eines Tests zur Erfassung der kristallinen verbalen Intelligenz, Inventar Lexikalischer Kompetenz (ILKO)

1 Ausgangspunkt der Testentwicklung

Mit dem Inventar Lexikalischer Kompetenz wird ein neues ökonomisches Testverfahren vorgestellt, das zur Messung der kristallinen Intelligenz sensu Cattell (1963) bei Personen mit höherer schulischer Bildung dient. Da das Verfahren Langzeitgedächtnisinhalte mit Hilfe des aktiven Wortschatzes überprüft, sollte es auch Hinweise auf die Sprachkompetenz der Testpersonen geben.

Bei der Entwicklung eines Tests, der Gedächtnisinhalte erfasst, besteht das Problem, Aufgaben zu finden, die repräsentativ für das zu überprüfende Wissensgebiet sind: Keine der Testpersonen darf auf Grund ihrer Herkunft oder Bildung bei der Bearbeitung des Tests benachteiligt sein. Die Repräsentativität der Zielwörter des ILKO für den allgemeinen deutschen Wortschatz wurde dadurch sichergestellt, indem sie der Liste der 30.000 in den Textkorpora des Instituts für Deutsche Sprache, Mannheim (IDS) am häufigsten vorkommenden Wortformen entnommen wurden.

Die Korpora geschriebener Gegenwartssprache des IDS bilden mit rund zwei Milliarden Wörtern (Stand 2002) die weltweit größte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit. Sie werden nach Kriterien zusammengestellt, die gewährleisten sollen, dass die Korpora ein authentisches Dokument des Gebrauchs der deutschen Sprache darstellen (Institut für Deutsche Sprache: Ausbau und Pflege der Korpora geschriebener Gegenwartssprache. <http://www.ids-mannheim.de/kt/projekte/korpora>, Aufruf am 25.11.2003).

2 Theoretische Fundierung

Zur Definition und theoretischen Einordnung des Messgegenstands des Inventars Lexikalischer Kompetenz werden zunächst die hierfür relevanten Modelle der Intelligenz und der Sprachkompetenz vorgestellt.

Die für die angewandte Psychologie bedeutsamsten Intelligenztheorien beruhen auf der Analyse der Bearbeitung von Aufgaben, die Intelligenzleistungen direkt erfassen sollen. Zu ihnen gehören neben den wichtigen psychometrischen Theorien auch kognitive Theorien und biologische Theorien (Li & Schmiedek, 2001; Sternberg, 2004; Sternberg & Kaufman, 2002).

2.1 Psychometrische Intelligenzmodelle

Zu Beginn des letzten Jahrhunderts gelang es dem Franzosen Alfred Binet mit Hilfe von standardisierten Intelligenztests zwischen geistig zurückgebliebenen und geistig normalen Schulkindern zu unterscheiden. Seit diesem Zeitpunkt spielt der psychometrische Ansatz, die individuellen Unterschiede der kognitiven Kompetenz unmittelbar zu quantifizieren, in der Intelligenzforschung die maßgebende Rolle (Li & Schmiedek, 2001; Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg & Urbina, 1996; Resing, 2005; Schonemann 2005; Wasserman & Tulsky, 2005). Atkinson (2001, S. 7600 ff.) schreibt hierzu: „Psychometric tests have dominated intelligence testing for a century. ... Although test tasks are selected pragmatically, they cluster in remarkably similar ways across tests and studies, giving insight into the structure of intelligence.“ Daniel (1997, S. 1043) führt aus: „The fact that the model (psychometric-ability model, Anm. d. Verf.) is based on decades of empirical findings gives it a type of robustness that more theoretically driven models do not enjoy to the same degree.“ Auch Neisser et al. (1996, S. 77) weisen darauf hin: „... the dominant psychometric approach ... has not only inspired the most research and attracted the most attention (up to this time) but is by far the most widely used in practical settings.“

Psychometrische Theorien versuchen die Struktur der Intelligenz zu erklären. Sie basieren auf Daten, die durch Tests gewonnen werden, die kognitive Fähigkeiten mit unterschiedlichen Aufgaben erfassen. Hierzu gehören z. B. die Vervollständigung von Zahlenreihen, das Ergänzen geometrischer Strukturen, die Komplettierung sprachlicher Analogien oder die

mentale Rotation von Objekten. Neben Unterschieden in der methodischen Aufarbeitung der gewonnenen Daten, unterscheiden sich die psychometrischen Theorien vor allem darin, ob sie sich eher zu einem Generalfaktor der Intelligenz, g , oder zu einer multidimensionalen Sicht hin orientieren (Li & Schmiedek, 2001; Resing, 2005).

2.1.1 Die Anfänge

Die erste der grundlegenden psychometrischen Theorien wurde von dem britischen Psychologen Charles E. Spearman entwickelt, die ‚Two-Factor Theory‘ der Intelligenz. In einem Aufsatz im *American Journal of Psychology* argumentierte Spearman (1904), dass den individuellen Unterschieden bei der Bearbeitung von Aufgaben, deren Lösung Intelligenzleistungen erfordern, zwei Faktoren zugrunde liegen. Spearman nannte den ersten Faktor den Generalfaktor g . g liegt der Bearbeitungsleistung bei allen Aufgaben in einem mehr oder weniger starken Ausmaß zugrunde. Der zweite Faktor, s , ist spezifisch für die Art der zu bearbeitenden Aufgabe und bezieht sich damit auf die hierfür erforderliche spezifische kognitive Operation, z. B. Objekte benennen oder Rechenaufgaben lösen. Nach Spearman existieren viele voneinander unabhängige s -Faktoren (Li & Schmiedek, 2001; Resing 2005; Sternberg, 2004; Schonemann, 2005; Wasserman & Tulsky, 2005). Spearmans Intelligenztheorie beruhte auf Annahmen, die von ihm mit Hilfe von Datensätzen, die Leistungen in Intelligenzaufgaben enthielten, korrelationsstatistisch überprüft wurden. Er gab seinem Aufsatz von 1904 daher auch den Titel „General intelligence, objectively determined and measured.“

Der amerikanische Psychologe L. L. Thurstone postulierte in den 1930er Jahren die Theorie der ‚Primary Mental Abilities‘. Grundlage seiner Theorie waren die faktorenanalytischen Auswertungen von Testergebnissen, die er mittels umfangreicher Testbatterien gewonnen hatte. Nach seiner Auffassung tragen die ‚Primary Mental Abilities‘ in unterschiedlichen Anteilen gemeinsam zur Erklärung der Leistung bei der Bearbeitung von Intelligenzaufgaben bei. Evidenz für einen Generalfaktor g konnte er, nach eigenen Aussagen, in seinen Datensätzen nicht finden. Thurstone argumentierte, dass die kognitive Leistungsfähigkeit einer Person mit einem Profil unterschiedlicher mentaler Fähigkeiten beschrieben werden sollte und nicht hauptsächlich mit einem einzigen Generalmaß. Die sieben von ihm ermittelten ‚Primary Mental Abilities‘ sind: spatial visualization, perceptual speed, numerical

facility, verbal comprehension, associative memory, word fluency und reasoning (Resing, 2005; Sternberg, 2004; Schonemann, 2005; Wasserman & Tulsky, 2005). Seine Annahme der Orthogonalität der sieben Faktoren konnte empirisch nicht gestützt werden. Spätere statistische Analysen seiner Daten, die u. a. von Spearman mit verbesserten faktorenanalytischen Techniken durchgeführt wurden, ergaben, dass die positiven Zusammenhänge zwischen den ‚Primary Mental Abilities‘ wiederum auf einen Faktor höherer Ordnung, g , zurückzuführen waren (Carroll, 1997; Resing, 2005; Li & Schmedek, 2001; Wasserman & Tulsky, 2005).

Raymond B. Cattell, ein früherer Assistent Spearmans, baute auf den Arbeiten von Thurstone und Spearman auf und entwickelte in den frühen 1940er Jahren seine ‚ g - g_c Theory‘ der Intelligenz. Cattell ersetzte den Spearman’schen Generalfaktor der Intelligenz g durch die beiden Generalfaktoren ‚fluide Intelligenz‘ (auch G_f statt g_f) und ‚kristalline Intelligenz‘ (auch G_c statt g_c ; Alfonso, Flanagan & Radwan, 2005; Amelang & Bartussek, 1985; Colom & Andrés-Pueyo, 2000; Horn & Blankson, 2005; Resing, 2005; Sternberg, 2004). G_f repräsentiert die Fähigkeit zum logischen Denken, insbesondere wenn eine Anpassung an neue Situationen erforderlich ist und hierbei der Rückgriff auf erworbenes Wissen nur geringen Nutzen hat. G_c repräsentiert die Ansammlung von deklarativem und prozeduralem Wissen über die gesamte Lebensspanne hinweg und seine Anwendung. Nach Cattell ist G_f der wichtigere der beiden Generalfaktoren, da er auch die Aneignung des Wissens determiniert (Wasserman & Tulsky, 2005). Auf Grundlage der beiden Konzepte wurden spezielle Messinstrumente entwickelt: G_f wird mit Hilfe von Aufgaben erfasst, deren Lösung logisches Denken erfordert, z. B. Beziehungen erkennen, Auswirkungen sehen und Schlussfolgerungen ziehen. Hierfür wird Aufgabenmaterial verwendet, das entweder neuartig oder allen Personen gleich vertraut ist. G_c wird mit Tests gemessen, die die Breite und die Tiefe des Wissens über die Sprache, die Konzepte und die Kenntnisse der vorherrschenden Kultur erfassen (Horn & Blankson, 2005).

2.1.2 Die Modelle von Cattell-Horn und Carroll

John L. Horn, ein Schüler von Raymond B. Cattell, ergänzte ab Mitte der 1960er bis Anfang der 1990er Jahre das G_f - G_c -Modell um acht weitere Faktoren, so dass es in seiner Endform die folgenden zehn Fähigkeiten zweiter Ordnung umfasste (siehe auch Abbildung 2.1):

Gf (fluid intelligence), Gc (crystallized intelligence), Gsm (short term memory), Gv (visual processing), Ga (auditory processing), Glr (long term retrieval), Gs (processing speed), Gt (reaction time/decision speed), Grw (reading/writing ability) und Gq (quantitative ability). Dieses Modell ist unter dem Namen ‚Cattell-Horn Gf-Gc Theory‘ bekannt (Alfonso, Flanagan & Radwan, 2005; McGrew, 2005; Wasserman & Tulsy, 2005). Jeder der zehn Faktoren wurde empirisch aus unterschiedlichen kognitiven Fähigkeiten abgeleitet, die mittels psychometrischer Tests operational definiert und gemessen wurden. Die fluide Intelligenz Gf und die kristalline Intelligenz Gc stellen die beiden Hauptfaktoren des Modells dar (Stankov, 2000; Sternberg, 2004).

John B. Carroll veröffentlichte 1993 unter dem Titel „Human cognitive abilities: A survey of factor-analytic studies“ eine monumentale Reanalyse von Ergebnissen der psychometrischen Intelligenzforschung. Er untersuchte mit Hilfe faktorenanalytischer Techniken 461 Datensätze, die zwischen 1927 und 1987 an mehr als 130.000 Personen mit Hilfe von verschiedensten Aufgaben gewonnen wurden, mit denen typischer Weise die kognitive Leistungsfähigkeit gemessen wird. Seine resultierende ‚Three-Stratum Theory of Cognitive Abilities‘ integriert die Erkenntnisse der verschiedenen psychometrischen Theorien in einer hierarchischen Struktur (Alfonso, Flanagan & Radwan, 2005; Carroll, 2005; Daniel, 1997; Sternberg, 2004). Carroll (2005, S. 71) schreibt hierzu: „The theory was intended to constitute a provisional statement about the enumeration, identification, and structuring of the total range of cognitive abilities known or discovered thus far. In this way it was expected to replace, expand, or supplement previous theories of the structure of cognitive abilities, ...“. Carroll zeigte auf, dass drei Schichten (strata), die in ihrer Breite und Universalität variieren, die Struktur der kognitiven Fähigkeiten abbilden: Stratum I (narrow) besteht aus 65 messbaren Basisfähigkeiten, wie z. B. muttersprachlicher Wortschatz, Grundrechenfertigkeit oder Tonhöhendifferenzierung. Aus den Basisfähigkeiten leiten sich die acht Fähigkeiten des Stratum II (broad) ab. Die Stratum II-Faktoren entsprechen im Wesentlichen denen des erweiterten Gf-Gc Modells von Cattell und Horn. Stratum III (general) repräsentiert einen Generalfaktor der Intelligenz, der mit Spearman's g vergleichbar ist (Alfonso, Flanagan & Radwan, 2005; Atkinson, 2001; Carroll, 2005; Kamphaus, Winsor, Rowe & Kim, 2005; Li & Schmiedek, 2001; Resing, 2005). Carrolls Generalfaktor hat inhaltlich deutliche Bezüge zur fluiden Intelligenz: Er repräsentiert mit Schwerpunkt Faktoren, die Tests enthalten, die das logische Denken erfassen (Wasserman & Tulsy, 2005).

Sternberg und Kaufman (2002, S. 591) resümieren zu Carrolls Arbeit: „Although Carroll does not break much new ground, in that many of the abilities in his model have been mentioned in other theories, he does masterfully integrate a large and diverse factor-analytic literature, thereby giving great authority to his model.“ Es dürfte nicht zuletzt Carrolls Arbeit zu verdanken sein, dass Daniel (1997, S. 1039) in seiner Übersicht zum Stand und zu den aktuellen Trends in der Intelligenzmessung festhält: „The long standing argument over general intelligence versus multiple abilities has given way to a broad acceptance of a hierarchical model in which abilities are nested under a higher order general factor, each level having substantial amount of explanatory power“. Auch Hunt (2000, S. 124) schreibt: “In his masterful review of literature, Carroll (1993) pointed out that the second level abilities are themselves correlated. This correlation can be interpreted as evidence for a third level, pervasive intelligence *g*“. Daniel (1997, S. 1043) ergänzt: „This* multifactor model provides a common frame of reference for test analysis and interpretation and for many avenues of research. It is an invaluable tool for establishing the construct validity of scores on psychometric-ability tests and for shedding light on the meaning of scores derived from other types of tests“ (*Carrolls, Anm. d. Verf.).

Carrolls Three-Stratum Theory und die letzte Version der Horn-Cattell Gf-Gc Theory sind bemerkenswert ähnliche Modelle. Der hauptsächliche Unterschied liegt darin, dass das Modell Carrolls eine höchste Ebene enthält, die *g* repräsentiert. Abbildung 2.1 (nach Alfonso, Flanagan & Radwan, 2005, S. 187) stellt die Gf-Gc Theory von Cattell und Horn und die Three-Stratum Theory von Carroll gegenüber. Sie beschränkt sich auf die Strata I und II. Beschreibungen der wenigen inhaltlichen Unterschiede der Fähigkeiten zweiter Ordnung der beiden Modelle finden sich z. B. bei McGrew (2005) und Alfonso, Flanagan & Radwan (2005).

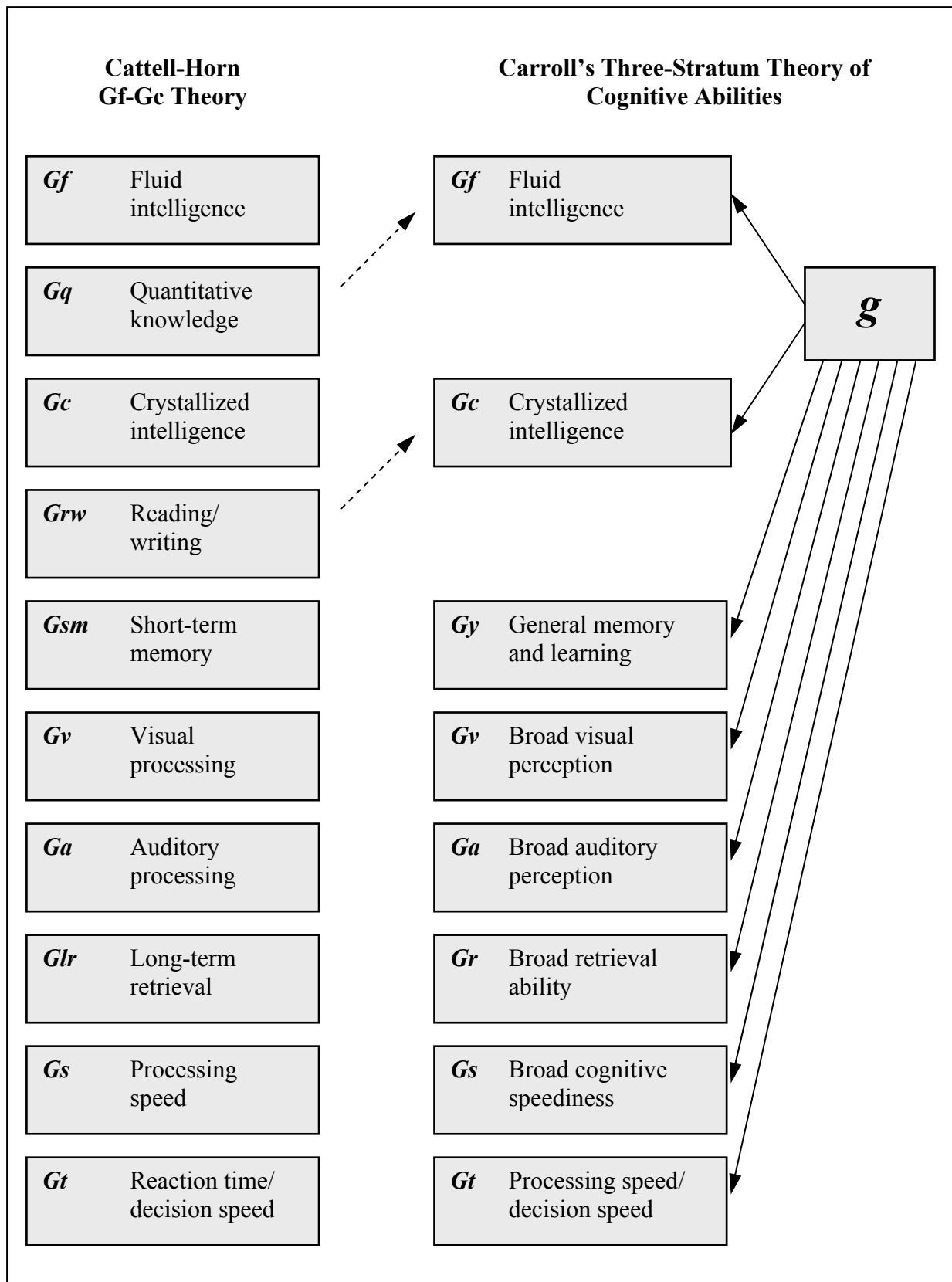


Abb. 2.1: Vergleich der Gf-Gc Intelligenztheorie von Cattell-Horn mit der Three-Stratum Intelligenztheorie von Carroll

Aufgrund seines Messgegenstandes ‚Wortschatz‘ ist das Inventar Lexikalischer Kompetenz im Rahmen der Cattell-Horn Gf-Gc Theory und der Three-Stratum Theory of Cognitive Abilities von Carroll eindeutig als Test der kristallinen Intelligenz einzuordnen:

- Cattell-Horn Gf-Gc Theory: Gc – V (verbal comprehension)
- Three-Stratum Theory of Cognitive Abilities von Carroll: Gc – VL (lexical knowledge)

2.2 Modelle der Sprachkompetenz

Obgleich der Begriff ‚Sprachkompetenz‘ in verschiedensten Zusammenhängen, in denen es um die Beschreibung bzw. Erfassung von personenbezogenen Merkmalen geht, mit einer großen Selbstverständlichkeit Verwendung findet (z. B. bei der Personalauswahl oder im Bildungsbereich), existieren in der Wissenschaft nur wenige umfassende Definitionsversuche. Die wenigen expliziten Definitionen variieren überdies inhaltlich deutlich. Das mag zum einen darin begründet sein, dass der Begriff in unterschiedlichen Disziplinen genutzt wird (Linguistik, Pädagogik, Psychologie, Soziologie, Ethnologie, Theaterwissenschaft, usw.) und dass diese Disziplinen, je nach Interessenschwerpunkt, bestimmte Aspekte der Sprachkompetenz betonen. Zum anderen setzt der (kompetente) Umgang mit Sprache eine derart große Vielzahl von menschlichen Fähigkeiten und Fertigkeiten voraus, dass deren vollständige Beschreibung in jedem Falle nur interdisziplinär erfolgen kann.

Um einen Überblick über das zu geben, was unter einem umfassenderen Begriff der Sprachkompetenz verstanden werden kann, werden im Folgenden die Modelle von Nodari (2002) und Coseriu (1988) und der ‚Gemeinsame europäische Referenzrahmen für Sprachen: lernen, lehren, beurteilen‘ (Trim, North und Coste, 2001) vorgestellt.

2.2.1 Sprachkompetenz nach Nodari

Claudio Nodari (2002) teilt das Konstrukt Sprachkompetenz in vier Dimensionen auf: sprachliche, soziolinguistische, sprachlogische und strategische Kompetenz (siehe Abbildung 2.2).

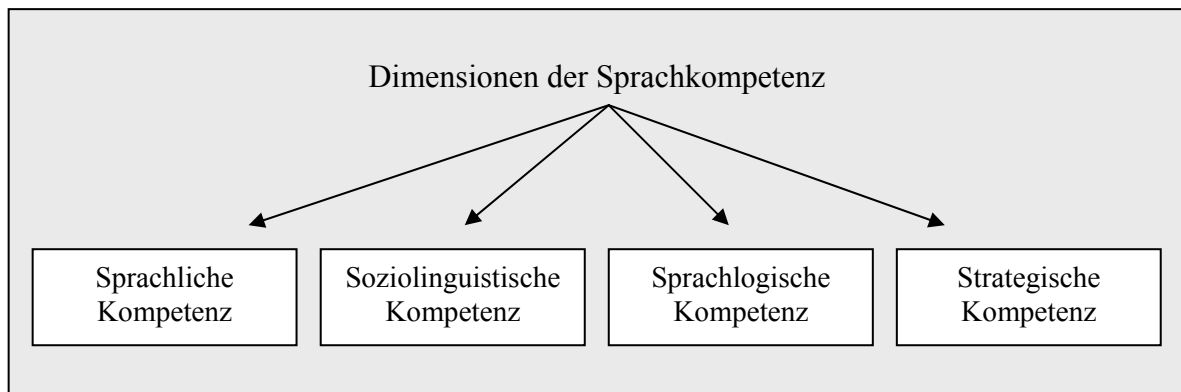


Abb. 2.2: Dimensionen der Sprachkompetenz nach Nodari

Die sprachliche Kompetenz entspricht im Modell von Nodari der umgangssprachlichen Kompetenz, d. h. dem Wissen und dem routinemäßigen Können in einer Sprache. Sie lässt sich in sechs Bereiche einteilen: Hörverstehen und Leseverstehen (rezeptive Sprachleistungen), Sprechen und Schreiben (produktive Sprachleistungen), Wortschatz und Grammatik (sprachliches Wissen). Nach Nodari genügt sprachliche Kompetenz allein nicht, um Sprache erfolgreich einsetzen zu können.

Die soziolinguistische Kompetenz umfasst das Wissen und die Anwendung von kulturspezifischen Normen, die festlegen, wie man sich sprachlich gegenüber bestimmten Personen und in bestimmten Situationen adäquat verhält. Diese kulturspezifischen Normen werden i. d. R. im sozialen Umgang erlernt und von den muttersprachlichen Sprachverwendern als selbstverständlich betrachtet bzw. gar nicht bewusst wahrgenommen. Besitzt eine Person in mehreren Sprachen eine hohe soziolinguistische Kompetenz, so besitzt sie für die Kulturen, aus denen die Sprachen stammen, eine hohe interkulturelle Kompetenz.

Die sprachlogische Kompetenz umfasst die Fähigkeit sich kohärent und nachvollziehbar in einer Sprache auszudrücken und komplexe Sprache zu verstehen. Im Gegensatz zu den ersten beiden Dimensionen, die für verschiedene Sprachen inhaltlich unterschiedlich ausfallen, handelt es sich bei der sprachlogischen Kompetenz um eine Basisfähigkeit, die weitgehend von einer Sprache in eine andere übertragen werden kann. Notwendige Voraussetzung für die Entwicklung und die Entfaltung der sprachlogischen Kompetenz ist das Vorhandensein der sprachlichen und der soziolinguistischen Kompetenz.

Die strategische Kompetenz umfasst die Fähigkeit, Probleme der sprachlichen Verständigung zu erkennen, sie anzugehen und sie zu lösen. Sie ist ebenfalls weitgehend sprachunabhängig und kann übereinzelsprachlich genutzt werden.

Der einzelsprachliche Wortschatz einer Person ist im Modell von Nodari eine Komponente der Dimension ‚Sprachliche Kompetenz‘. Sie ist eine notwendige aber nicht hinreichende Voraussetzung für sprachkompetentes Verhalten.

2.2.2 Sprachkompetenz nach Coseriu

Die Sprachkompetenz in ihrer Gesamtheit ist im Modell von Eugenio Coseriu (1988) einer der beiden konstituierenden Teile der allgemeinen menschlichen Ausdrucksfähigkeit (siehe Abbildung 2.3). Der zweite Teil ist die Fähigkeit zu sprachbegleitenden Tätigkeiten, die die Verwendung von Sprache ergänzen, z. B. Mimik und Gestik.



Abb. 2.3: Einteilung der Sprachkompetenz nach Coseriu

Coseriu unterteilt die Sprachkompetenz ‚in ihrer Gesamtheit‘ in die Bereiche physisch-psychische Sprachkompetenz und kulturelle Sprachkompetenz. Die physisch-psychische Sprachkompetenz umfasst die körperlichen und geistigen Voraussetzungen für die Produktion und Rezeption von sprachlichen Zeichen. Die kulturelle Sprachkompetenz definiert die Verwendung von Sprache als „eine universelle allgemein-menschliche Tätigkeit, die jeweils von individuellen Sprechern als Vertretern von Sprachgemeinschaften mit gemeinschaftlichen Traditionen des Sprechkönnens individuell in bestimmten Situationen realisiert wird“ (Coseriu, 1988, S. 70). Sie beinhaltet die Ebenen allgemein-sprachliche Kompetenz und einzelsprachliche Kompetenz sowie die individuelle Ebene Text-Kompetenz für geschriebene Sprache bzw. Diskurs-Kompetenz für gesprochene Sprache.

Den drei Ebenen der kulturellen Sprachkompetenz sind unterschiedliche Formen von Wissen zugeordnet: das elokutionelle, das idiomatische und das expressive Wissen.

Das elokutionelle Wissen liegt der allgemein-sprachlichen Kompetenz zu Grunde. Es ist übereinzelsprachlich, d. h. es gehört zu jedem Sprechen, gleichgültig in welcher Sprache, und umfasst die allgemeine Kenntnis der Sachen, wie sie normalerweise sind, und des normalen Verhaltens. Das elokutionelle Wissen erlaubt u. a. Gesagtes als kohärent oder inkohärent zum „Normalen“ zu beurteilen, Gesagtes zu interpretieren und Nichtgesagtes, das als normal oder zu erwartend vorausgesetzt wird, mitzuverstehen.

Das idiomatische Wissen liegt der einzelsprachlichen Kompetenz zu Grunde. Es enthält einerseits „Verfahren der Kombination, denen als Regeln explizierbare Normen innewohnen“ und andererseits „vorgegebene Elemente, die kombiniert werden“ (Coseriu, 1988, S. 248). Das idiomatische Wissen erlaubt, eine bestimmte Einzelsprache oder eine Varietät einer Einzelsprache zu verstehen und sich in ihr ausdrücken zu können.

Das expressive Wissen liegt der Text- oder Diskurs-Kompetenz zu Grunde. Die kulturelle Tätigkeit des Sprechens wird jeweils von individuellen Sprechern in einzelnen Situationen ausgeführt. Das expressive Wissen bezieht sich auf die allgemeinen Determinationen des Sprechens, wie Sprecher, Adressat, Gegenstand, Situation und die jeweiligen Normen, die diese Determinationen betreffen. Es erlaubt dem Sprecher über den Inhalt der sprachlichen Botschaft, ihre Strukturierung, ihre Angemessenheit und soziale Korrektheit, mit dem Gesagten Stellungnahmen, Meinungen oder Absichten adäquat auszudrücken.

Im Modell von Coseriu ist der einzelsprachliche Wortschatz Bestandteil des idiomatischen Wissens. Auch hier ist er eine notwendige aber nicht hinreichende Voraussetzung für sprachkompetentes Verhalten.

2.2.3 Sprachkompetenz gemäß dem europäischen Referenzrahmen

Der umfangreichste Ansatz, das Konstrukt ‚Sprachkompetenz‘ inhaltlich differenziert zu erfassen und messbar zu machen, stellt der ‚Gemeinsame europäische Referenzrahmen für Sprachen: lernen, lehren, beurteilen‘ dar, der im Auftrag des Bildungsausschusses ‚Sprachenlernen für europäische Bürger‘ des Rates für kulturelle Zusammenarbeit des Europarates in Straßburg erstellt wurde (Trim, North & Coste, 2001; im Folgenden vereinfacht als ‚europäischer Referenzrahmen‘ bezeichnet).

Der europäische Referenzrahmen wurde in einem umfangreichen Prozess wissenschaftlicher Forschung und Beratung entwickelt. Seine übergeordneten Ziele sind die Verbesserung der Kommunikation und des gegenseitigen Verständnisses unter den Europäern und damit die Förderung der persönlichen Mobilität, des Austausches von Ideen und der Zusammenarbeit. Der Referenzrahmen beschreibt handlungsorientiert und umfassend, was Lernende tun müssen, um eine Sprache für kommunikative Zwecke nutzen zu können und um in der Lage zu sein, erfolgreich zu handeln. Er definiert international vergleichbare sprachliche Kompetenzniveaus, enthält objektive Kriterien für die Beschreibung von Sprachkompetenzen, liefert die Grundlage für eindeutige Beschreibungen von Zielen und Inhalten von Maßnahmen zum Spracherwerb und zur Überprüfung von Sprachfertigkeiten und ermöglicht und fördert die Anerkennung von eindeutigen sprachbezogenen Qualifikationen. Er ist damit von besonderem Interesse für alle Personengruppen und Institutionen, die mit dem Lehren und Überprüfen von Sprachfertigkeiten betraut sind.

Das Kapitel 5 des europäischen Referenzrahmens kategorisiert detailliert die allgemeinen und kommunikativen Kompetenzen eines Sprachverwenders, die er im Laufe früherer Erfahrungen entwickeln und als Persönlichkeitsmerkmale besitzen muss, um die in kommunikativen Situationen gestellten Aufgaben bewältigen und die erforderlichen Aktivitäten ausführen zu können. Auf die obersten vier Ebenen beschränkt, sind das im Einzelnen (siehe Abbildung 2.4):

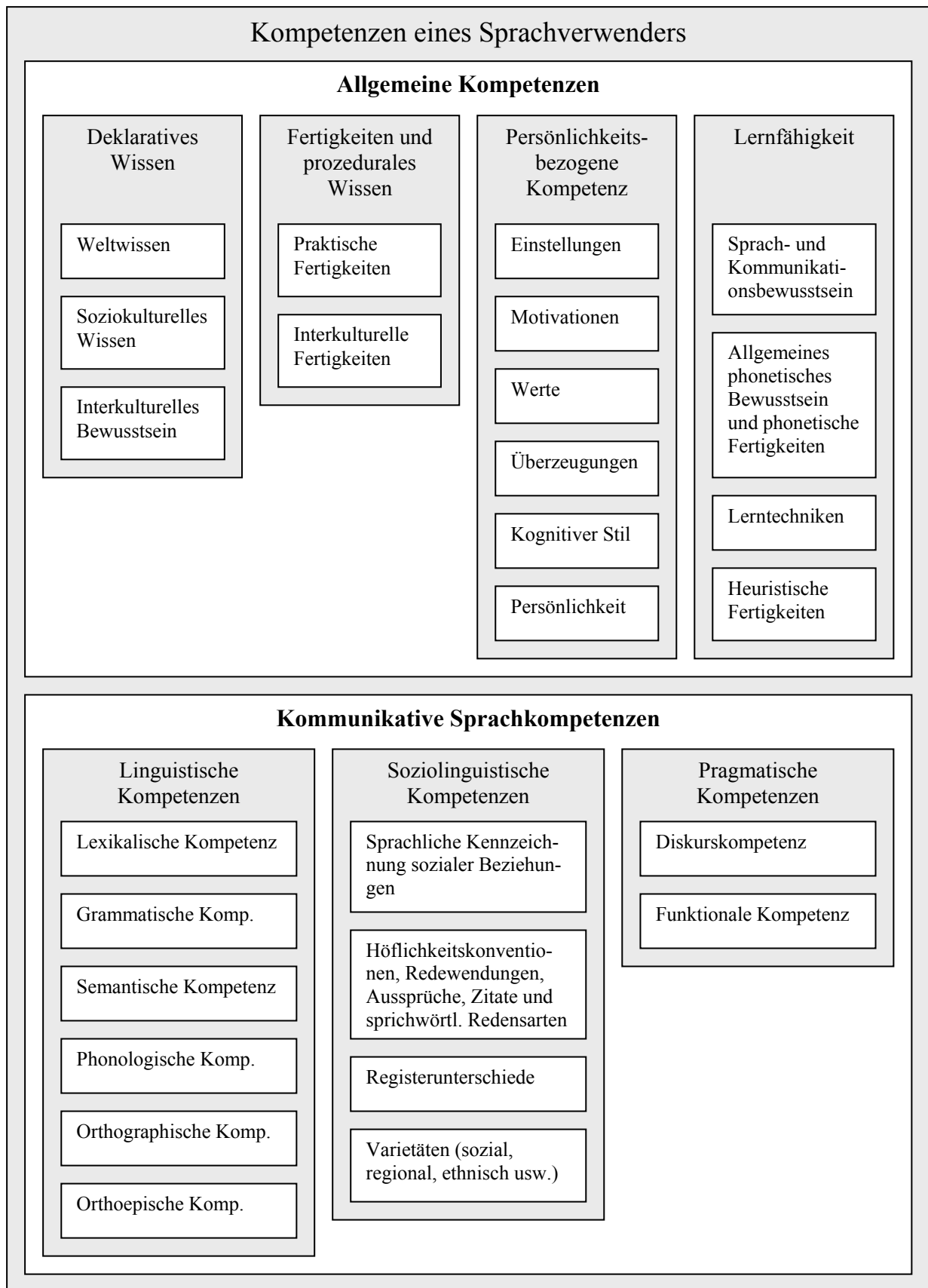


Abb. 2.4: Kompetenzen eines Sprachverwenders gemäß des ‚Gemeinsamen europäischen Referenzrahmens für Sprachen: lernen, lehren, beurteilen‘

Der einzelsprachliche Wortschatz ist ein Bestandteil der lexikalischen Kompetenz, die ihrerseits ein Element der linguistischen Kompetenz darstellt. Die lexikalische Kompetenz umfasst die Kenntnis des Vokabulars einer Sprache, das aus lexikalischen und aus grammatischen Elementen besteht, sowie die Fähigkeit, es zu verwenden. Bei den lexikalischen Elementen unterscheidet man zwischen festen Wendungen, die aus mehreren Wörtern bestehen und jeweils als ein Ganzes gelernt und verwendet werden, z. B. Satzformeln (Nett, Sie kennen zu lernen.) und idiomatische Wendungen (Er hat den Löffel abgegeben.), und Einzelwörtern. Einzelwörter sind frei stehende einzelne Wörter (Lexeme). Manche Lexeme können verschiedene Bedeutungen haben (Polysemie; z. B. Bank, ein Geldinstitut oder eine Objekt zum Sitzen). Lexeme können offenen Wortklassen angehören und Substantiv, Verb, Adjektiv, Adverb sein; diese können aber auch ihrerseits geschlossene lexikalische Gruppen bilden (z. B. die Wochentage, die Monate eines Jahres, Gewichte, Maße, usw.).

Gemäß dem europäischen Referenzrahmen überprüft das Inventar Lexikalischer Kompetenz die Kenntnis von Einzelwörtern der deutschen Sprache, die eine eindeutige Bedeutung besitzen und der Wortklasse der Substantive angehören. Auch hier ist die Kenntnis der Einzelwörter eine notwendige aber nicht hinreichende Voraussetzung für sprachkompetentes Verhalten.

Fazit

Der Wortschatztest ILKO erfasst Elemente des deutschen Wortschatzes, die repräsentativ für den Alltagsgebrauch der deutschen Sprache sind. In den dargestellten Modellen der Sprachkompetenz ist der einzelsprachliche Wortschatz nur ein Bestandteil unter vielen, wenn auch ein wichtiger. Mit der bloßen Kenntnis des Wortschatzes einer Person lassen sich keinerlei Aussagen über ihre Sprachkompetenz, noch nicht einmal über die Qualität ihrer Sprachproduktion oder -rezeption treffen, selbst wenn sie außerhalb eines sozialen Kontextes isoliert stattfinden.

Wegen der vielfältigen Aspekte des Konstrukts ‚Sprachkompetenz‘ wäre es nur mit Hilfe umfangreicher Untersuchungen möglich, zu überprüfen, ob das Testergebnis im Inventar Lexikalischer Kompetenz ein valider Indikator der individuellen Sprachkompetenz ist. Aus

diesem Grund wird im Rahmen der vorliegenden Arbeit zunächst versucht, das Instrument für den Einsatz im Bereich der Intelligenzmessung empirisch abzusichern.

3 Beschreibung des Tests

Das Inventar Lexikalischer Kompetenz besteht aus insgesamt 18 Aufgaben, die in ihrer Schwierigkeit ansteigen. Alle Aufgaben müssen in einem vorgegebenen Zeitraum von 9 Minuten nacheinander bearbeitet werden. Für die Bearbeitung der einzelnen Aufgaben gibt es keine zusätzlichen Zeitbegrenzungen. Trotz der Begrenzung der Gesamtbearbeitungszeit kann ILKO als Power-Test gelten: Die Erfahrung mit der Aufgabenbearbeitung durch die Testpersonen zeigt, dass die Antworten spontan erfolgen. Die Bearbeitungszeit für eine Aufgabe kann mit maximal 30 Sekunden veranschlagt werden. Die Gesamttestzeit von 9 Minuten beinhaltet im Regelfall also einen Zeitüberschuss. Diesen Zeitüberschuss können die Testpersonen mit der Überprüfung ihrer Antworten ausfüllen. Die Sichtung der bearbeiteten Testbögen zeigt jedoch, dass einmal gegebene Antworten i. d. R. nicht korrigiert werden.

Jede Aufgabe besteht aus einer eindeutigen inhaltlichen Beschreibung eines Substantivs der deutschen Sprache. Die Testperson soll die Beschreibung lesen und das beschriebene Wort in die darunter liegende Antwortzeile schreiben. Als Beispiel wird in Abbildung 2.5 die erste Aufgabe des ILKO dargestellt.

Wort, das mit einem anderen oder mehreren Wörtern derselben Sprache bedeutungsgleich ist.

Antwort: **Synonym**

Abb. 2.5: Die erste Aufgabe des Inventars Lexikalischer Kompetenz

4 Entwicklung der Testsaufgaben

Bei der Entwicklung eines Tests, der Wörter als Inhalte des Langzeitgedächtnisses abfragt, ergibt sich das Problem der Repräsentativität der Wörter für den allgemeinen Sprachgebrauch. Keine Person der Zielpopulation darf aufgrund z. B. ihrer Herkunft oder Ausbildung gegenüber anderen Testpersonen bei der Bearbeitung des Tests bevorteilt oder benachteiligt sein. Statt den nahe liegenden aber wenig eleganten Ansatz einer Zufallsauswahl von Wörtern aus einem Wörterbuch zu verfolgen (vgl. z. B. den Wortschatztest - aktiv und passiv von Ibrahimovic & Bulheller, 2005), wurden zur Entwicklung des ILKO die Textkorpora des Instituts für Deutsche Sprache, Mannheim (IDS) herangezogen.

4.1 Korpora geschriebener Gegenwartssprache

Die Korpora geschriebener Gegenwartssprache des Instituts für Deutsche Sprache, Mannheim sind die weltweit größte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit. Das IDS begann Mitte der 1960er Jahre mit ihrem Aufbau. Die Korpora umfassten im Jahre 2002 über 1,9 Milliarden Textwörter (das entspricht gerundet 4,8 Millionen Buchseiten, wenn man durchschnittlich 400 Wörter pro Seite zugrunde legt). Sie enthalten belletristische, wissenschaftliche und populärwissenschaftliche Texte, eine große Zahl von Zeitungstexten sowie eine breite Palette weiterer Textarten und werden kontinuierlich weiterentwickelt.

Die Korpora geschriebener Gegenwartssprache sollen den tatsächlichen Gebrauch der deutschen Sprache dokumentieren und diese Dokumentation stetig fortschreiben. Künstliche Texte kommen als mögliche Quellen nicht und Webseiten nur bedingt in Frage, da sie nur einen sehr speziellen Ausschnitt der Sprache darstellen. Die Aufgabe des Projekts ist demnach die Beschaffung von geeigneten, möglichst elektronischen Vorlagen, die ein authentisches Dokument des Gebrauchs der deutschen Sprache darstellen. Die Korpora dienen in erster Linie als empirische Basis für die linguistische Forschung (Institut für Deutsche Sprache: Ausbau und Pflege der Korpora geschriebener Gegenwartssprache. <http://www.ids-mannheim.de/kt/projekte/korpora>, Aufruf am 25.11.2003).

4.2 Auswahl der Zielwörter

Die Zielwörter, die der Konstruktion der Aufgaben des ILKO zugrunde lagen, wurden der Liste der 30.000 in den Textkorpora am häufigsten vorkommenden Wortformen entnommen. Die Liste enthält die Wortformen nach relativer Häufigkeit absteigend sortiert. Sie wurde von der Arbeitsgruppe für Korpustechnologie des IDS am 24. September 2003 erstellt. Der Zugriff auf die Liste erfolgte am 25. November 2003 unter <http://www.ids-mannheim.de/kt/30000wordforms.dat>.

Abbildung 2.6 enthält die Kriterien, die der Auswahl der Zielwörter aus der Liste zugrunde lagen.

1. Substantive aus den Rangplätzen 20.001 bis 30.000.
2. Substantive, die aus nur einem Wortstamm bestehen.

Abb. 2.6: Kriterien für die Auswahl der Zielwörter des ILKO aus der Liste der 30.000 häufigsten Wortformen der Textkorpora des IDS

Die Festlegung des ersten Kriteriums erfolgte, um den Schwierigkeitsgrad der Zielwörter vorab einzugrenzen. Es ist evident, dass Substantive, die sehr häufig in Texten vorkommen, die für den deutschen Sprachgebrauch repräsentativ sind, einen sehr hohen Bekanntheitsgrad bei Personen besitzen, deren Muttersprache Deutsch ist. (Z. B. lautet das in den Textkorpora am häufigsten vorkommende Substantiv auf Rangplatz 56 ‚Uhr‘.) Bei der Anwendung des ersten Kriteriums wurde sichergestellt, dass die ausgewählten Substantive nicht in einer anderen Wortform bereits im Bereich der Rangplätze 1 bis 20.000 vorkamen. Die Festlegung des zweiten Kriteriums fand statt, um die Einheitlichkeit bei der Benennung der Zielwörter bereits im Vorfeld zu erhöhen. Substantive, die aus mehreren Wortstämmen zusammengesetzt sind, können eine ganze Reihe unterschiedlicher bedeutungsgleicher Formen annehmen (z. B. Jubelfest, Jubelfeier, Freudenfest, Freudenfeier).

Unter Anwendung der beiden Auswahlkriterien und einer ersten subjektiven Überprüfung der ausgewählten Wörter hinsichtlich der Existenz von Synonymen resultierte ein Pool von 126 Zielwörtern. Im nächsten Schritt wurden die inhaltlichen Bedeutungen der Zielwörter genau beschrieben.

4.3 Beschreibung der inhaltlichen Bedeutungen der Zielwörter

Die Beschreibungen der inhaltlichen Bedeutungen der Zielwörter wurden folgenden Standardwerken zur deutschen Sprache entnommen:

- BROCKHAUS, Der Brockhaus in einem Band (2003)
- DUDEN, Deutsches Universalwörterbuch (2003a)
- DUDEN, Band 5: Das große Fremdwörterbuch (2003b)

Zielwörter, deren Inhaltsbeschreibungen folgende Charakteristika aufwiesen, wurden von der weiteren Testentwicklung ausgeschlossen:

1. Wörter, zu deren Beschreibung ihr eigener Wortstamm Verwendung fand.

Beispiel: Jemand, der bestimmte Erscheinungen analysiert; jemand, der die Analytik anwendet und beherrscht (Zielwort: Analytiker).

2. Beschreibungen, deren Verständnis ein hohes Abstraktionsvermögen erfordert.

Beispiel: Geometrischer Körper, dessen Oberfläche von einer in einer Spitze endenden, gleichmäßig gekrümmten Fläche über einer kreisförmigen oder elliptischen Grundfläche gebildet wird (Zielwort: Kegel).

Der verbleibende Pool von Aufgaben umfasste die inhaltlichen Beschreibungen von 80 Zielwörtern. Sie sind in Abbildung 2.7 dargestellt. Ihre Inhaltsbeschreibungen wurden zur ersten Testvorform zusammengestellt.

1. Kollegium	21. Abstinenz	41. Relation	61. Synonym
2. Ballast	22. Kartell	42. Pauschale	62. Rabatt
3. Chiffre	23. Detektiv	43. Omen	63. Obduktion
4. Trilogie	24. Erosion	44. Aquarium	64. Exkursion
5. Dekade	25. Kasette	45. Dressur	65. Märtyrer
6. Populismus	26. Blamage	46. Grotteske	66. Rezeption
7. Spurt	27. Rivalität	47. Relikt	67. Enklave
8. Rabatte	28. Matinee	48. Boutique	68. Notar
9. Abonnement	29. Epidemie	49. Statur	69. Alibi
10. Provision	30. Ombudsmann	50. Paradox	70. Internat
11. Manege	31. Parabel	51. Strudel	71. Monolog
12. Pessimismus	32. Karosserie	52. Relevanz	72. Duett
13. Prolog	33. Primat	53. Vitrine	73. Biografie
14. Skrupel	34. Dementi	54. Trance	74. Menü
15. Befruchtung	35. Opportunismus	55. Pseudonym	75. Patriotismus
16. Gemäuer	36. Mäzen	56. Unkraut	76. Baracke
17. Diskretion	37. Tilgung	57. Kittel	77. Allergie
18. Mosaik	38. Akquisition	58. Mobiliar	78. Saldo
19. Jargon	39. Inserat	59. Collage	79. Gage
20. Kantor	40. Diagnostik	60. Zirkel	80. Klonen

Abb. 2.7: Zielwörter der ersten Testvorform des ILKO

5 Konstruktion der Skala

Das Inventar Lexikalischer Kompetenz wurde als selbst instruierender Test entworfen. Auf der Titelseite müssen die Probanden zunächst einige Angaben zu ihrer Person machen. Die in Abbildung 2.8 dargestellten Angaben wurden im Rahmen der Testentwicklung erhoben. Sie dienen zur Analyse der Skala und zur Entwicklung der Normen.

Name/Code: _____ **Alter** (in Jahren): _____ **Geschlecht:** männlich weiblich

Muttersprache: Deutsch andere: _____

1. Fremdsprache:

Englisch Französisch Latein Spanisch Russisch Italienisch _____

2. Fremdsprache:

Englisch Französisch Latein Spanisch Russisch Italienisch _____

Schulbildung: Hauptschulabschluss Realschulabschluss Fachhochschulreife Abitur

Hochschule-Fachgebiet:

<input type="radio"/> Ingenieurwissenschaften	<input type="radio"/> Mathematik, Naturwissenschaften
<input type="radio"/> Agrar- und Forstwissenschaften	<input type="radio"/> Medizin, Gesundheitswesen
<input type="radio"/> Rechts- und Wirtschaftswissenschaften	<input type="radio"/> Gesellschafts- und Sozialwissenschaften
<input type="radio"/> Sprach- und Kulturwiss., Kunst und Gestaltung	

Lehramtsstudium: ja nein

Abb. 2.8: Personenbezogene Daten, die im Rahmen der Entwicklung des ILKO erhoben wurden

Die Erfassung des Namens bzw. Codes diente zur Zuordnung der Testergebnisse zu den Probanden und zu den Ergebnissen anderer diagnostischer Verfahren, die im Rahmen von Testbatterien gemeinsam mit dem ILKO bearbeitet wurden. Die Erfassung der Kontrollvariablen Alter, Geschlecht, Sprachkenntnisse und Hochschule-Fachgebiet sollte eine Überprüfung ermöglichen, ob diese personenbezogenen Merkmale in einem systematischen Zusammenhang mit dem Wortschatz der Testpersonen stehen. Die Aufgliederung der Hochschulfachgebiete wurde übernommen von der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, Bonn (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung: Studienfächer.

<http://www.studienwahl.de/fmg.htm>, Aufruf am 06.02.2004). Unter 1. und 2. Fremdsprache wurden die sechs vom Statistischen Bundesamt für die Schuljahre 2000/01 bis 2002/03 am häufigsten gelehrt Fremdsprachen an den allgemein bildenden Schulen in der Bundesrepublik Deutschland aufgeführt (Statistisches Bundesamt Deutschland: Bildung, Wissenschaft und Kultur. <http://www.destatis.de/basis/d/biwiki/schultab15.php>, Aufruf am 06.02.2004; siehe Tabelle 2.1).

Sprachen	Schuljahr		
	2000/01	2001/02	2002/03
Englisch	6 509,1	6 584,8	6 755,4
Französisch	1 616,6	1 610,1	1 644,0
Latein	618,6	627,1	654,0
Spanisch	114,8	130,3	151,7
Russisch	162,4	151,8	145,3
Italienisch	34,6	34,7	38,6
Griechisch	13,2	12,8	13,3
Türkisch	11,4	11,7	12,2
sonstige	38,6	43,3	47,6

(Einheit: 1000)

Tab. 2.1: Schüler/innen mit fremdsprachlichem Unterricht an den allgemein bildenden Schulen der Schuljahre 2000/01, 2001/02 und 2002/03 in der Bundesrepublik Deutschland

Die Angaben zu den Sprachkenntnissen und zu Hochschule-Fachgebiet wurden ab der ersten Datenerhebung zur Selektion der Items nach Schwierigkeiten und Synonymen erfasst (siehe Teil 2, Kapitel 5.1 und Kapitel 5.2.).

Die Erhebung der Sprachkenntnisse der Testpersonen sollte, wie bereits erwähnt, die Frage klären helfen, ob der mutter- bzw. fremdsprachliche Wortschatz einen signifikanten Einfluss auf die Verfügbarkeit der Zielwörter hat. Eine Überprüfung der Herkunft der Zielwörter ergab, dass fast alle Wörter lateinischen Ursprungs sind. Dieser Sachverhalt begründete die Vermutung, dass Personen mit Lateinkenntnissen oder Kenntnissen einer anderen romanischen Sprache (z. B. Französisch, Italienisch oder Spanisch) bei der Testbearbeitung im Vorteil sein könnten.

Eine Übersetzung der (deutschen) Zielwörter ins Französische und Englische, die beiden am häufigsten gelehrt Fremdsprachen an den allgemein bildenden Schulen in der Bundesrepublik Deutschland (siehe Tabelle 2.1), zeigte, dass 16 der 22 Wörter, die nach der

ersten Itemselektion verblieben waren (siehe Teil 2, Kapitel 5.2), in allen drei Sprachen denselben Wortstamm besitzen (Tabelle 2.2).

Wort	(Herkunft)	Englisch	Französisch
Collage	(fr.)	collage	collage
Synonym	(gr.-lat.)	synonym	synonyme
Gage	(germ.-fr.)	fee	cachet
Karosserie	(gall.-lat.-it.-fr.)	bodywork	carrosserie
Trilogie	(gr.)	trilogy	trilogie
Erosion	(lat.)	erosion	érosion
Prolog	(gr.-lat.)	prologue	prologue
Kartell	(ägypt.-gr.-lat.-it.-fr.)	cartel	cartel
Abstinenz	(lat.)	abstinence	abstinence
Saldo	(lat.-it.)	balance	solde
Trance	(lat.-fr.-engl.)	trance	transe
Provision	(lat.-it.)	commission	provision
Relation	(lat.-fr.)	relation	relation
Dekade	(gr.-lat.-fr.)	decade	décade
Omen	(lat.-fr.)	omen	omineux
Matinee	(lat.-fr.)	matinee	(matinée - Morgen, Vormittag)
Ballast	(ahd.-dän.-schwed.)	ballast	lest
Rabatte	(lat.-fr.-niederl.)	border	rabate, plate-bande
Pseudonym	(gr.)	pseudonym	pseudonyme
Relevanz	(lat.)	relevance	importance
Enklave	(lat.-fr.)	enclave	enclave
Opportunismus	(lat.-fr.)	opportunism	opportunisme

ägypt. – ägyptisch, ahd. – althochdeutsch, dän. – dänisch, engl. – englisch, fr. – französisch, gall. – gallisch, germ. – germanisch, gr. – griechisch, it. – italienisch, lat. – lateinisch, niederl. – niederländisch, schwed. – schwedisch

Tab. 2.2: Herkunft und Sprachverwandtschaft der Zielwörter des ILKO

Die Überprüfung der Herkunft der Wörter erfolgte mit Hilfe des DUDEN, Herkunftswörterbuch der deutschen Sprache (1989). Die Übersetzungen ins Englische bzw. Französische erfolgten mit Hilfe des COLLINS, Deutsch-Englisch (1991) bzw. des PONS, Deutsch-Französisch (1985).

5.1 Itemselektion nach Synonymen

Die Beschreibungen der 80 Zielwörter (siehe Abbildung 2.7), im Folgenden auch ‚Aufgaben‘ oder ‚Items‘ genannt, wurden im Dezember 2003 einer ersten Stichprobe von 37 Studenten einer baden-württembergischen Verwaltungsfachhochschule vorgelegt. Diese erste Datenerhebung sollte lediglich Hinweise auf die Existenz von Synonymen der Zielwörter liefern. Bei der Testbearbeitung wurde den Probanden kein Zeitlimit vorgegeben, so dass sie alle Aufgaben bearbeiten konnten.

Nach Maßgabe dessen, dass nur Wortbeschreibungen in die Endversion des Tests aufgenommen werden sollten, bei denen nur ein einziges Wort die richtige Antwort darstellt, wurden nach der ersten Datensammlung 35 Items aus der Testvorform eliminiert. Die Überprüfung der inhaltlichen Bedeutung der gültigen Alternativantworten (Synonyme), die von den Testpersonen gefunden wurden, erfolgte ebenfalls mit Hilfe der im Teil 2, Kapitel 4.3 aufgeführten Standardwerke zur deutschen Sprache.

Gesamtheit der Lehrerinnen und Lehrer einer Schule.	
<i>gesuchte Antwort:</i>	Kollegium
<i>weitere korrekte Antwort:</i>	Lehrkörper
Öffnung einer Leiche zur Feststellung der Todesursache.	
<i>gesuchte Antwort:</i>	Obduktion
<i>weitere korrekte Antwort:</i>	Autopsie

Abb. 2.9: Beispiele für Items, zu denen korrekte Alternativantworten gefunden wurden

5.2 Itemselektion nach Schwierigkeit

In einem zweiten Schritt wurde im Februar 2004 die Stichprobe zur Aufgabenanalyse um 23 Schüler der 13 Klasse eines hessischen Gymnasiums auf insgesamt 60 Personen erweitert. Bei vier der vorgegebenen 45 Items konnten weitere korrekte Alternativantworten identifiziert

werden, so dass auch diese Items aus dem Itempool entfernt wurden. Anschließend erfolgte die Berechnung der Schwierigkeiten (relative Lösungshäufigkeit in Prozent) der verbliebenen 41 Items. Das Ergebnis gibt Tabelle 2.3 wieder:

Ballast21,7%	Karosserie..... 55,0%	Collage*70,0%
Chiffre88,3%	Opportunismus ... 10,0%	Zirkel*100,0%
Trilogie.....55,0%	Akquisition 3,3%	Synonym* 68,3%
Dekade26,7%	Relation 28,3%	Rezeption* 76,7%
Populismus3,3%	Pauschale 3,3%	Enklave* 10,0%
Rabatte20,0%	Omen 23,3%	Alibi*96,7%
Provision30,0%	Aquarium..... 95,0%	Monolog*90,0%
Pessimismus88,3%	Groteske 1,7%	Duett* 81,7%
Prolog.....48,3%	Boutique 95,0%	Menü* 91,7%
Mosaik.....83,3%	Relevanz 11,7%	Allergie*96,7%
Abstinenz41,7%	Vitrine..... 91,4%	Saldo* 38,3%
Kartell46,7%	Trance..... 31,7%	Gage* 65,0%
Erosion53,3%	Unkraut..... 88,3%	Klonen*96,7%
Matinee23,3%	Pseudonym* 16,7%	

(10% > p > 70% grau unterlegt)

Tab. 2.3: Schwierigkeiten der Items der Testvorform mit 41 Aufgaben

Nach der zweiten Datenerhebung wurden alle Items eliminiert, die entweder sehr leicht ($p > 70\%$) oder sehr schwer ($p < 10\%$) waren (in Tabelle 2.3 grau unterlegt). Die Entscheidung, auch relativ schwere Items beizubehalten, wurde durch folgendes Untersuchungsergebnis begründet: Die Hälfte der Personen einer Gruppe von 10 Mitarbeitern des gehobenen und höheren Verwaltungsdienstes aus Baden-Württemberg, denen eine Auswahl von 16 Items mittlerer Schwierigkeit ($20\% \leq p \leq 60\%$) im Februar 2004 vorgelegt wurde, löste 70 bis 100 Prozent der Aufgaben. Dieses Ergebnis korrespondiert mit der Veränderung der Schwierigkeitsindizes nach der Bearbeitung des Tests durch Gruppen von Universitätsstudenten (siehe Teil 2, Kapitel 5.3 und Kapitel 6.4).

Exkurs: Für die Wörter aus dem Bereich der Rangplätze 20.001 bis 25.000 der Textkorpora (in Tabelle 2.3 mit * gekennzeichnet) ergibt sich eine durchschnittliche Schwierigkeit von $p = 71,3\%$. Die Wörter aus dem Bereich der Rangplätze 25.001 bis 30.000 (ohne Kennzeichnung) weisen hingegen eine durchschnittliche Schwierigkeit von $P = 43,3\%$ auf. Dieser deutliche Unterschied ist ein Beleg dafür, dass die Auftretenshäufigkeit von Wörtern im (schriftlichen) Sprachgebrauch in Zusammenhang steht mit ihrer allgemeinen Bekanntheit.

Die 22 verbliebenen Zielwörter wurden abschließend mit Hilfe des DUDEN, Band 8: Die sinn- und sachverwandten Wörter (1997) noch einmal auf die Existenz von Synonymen untersucht. Die Überprüfung ergab, dass die Eindeutigkeit der Aufgabenlösungen gegeben ist.

5.3 Itemselektion nach Trennschärfe

Zur weiteren Item- und Skalenanalyse wurden im Juni und Juli 2004 zusätzliche Untersuchungsstichproben erhoben. Es handelte sich hierbei um insgesamt 127 Studenten der Studienfächer Germanistik, Rechtswissenschaften, Wirtschaftswissenschaften und Architektur dreier hessischer Universtäten. Für die 22 Items des Tests wurden die Schwierigkeiten und die part-whole-korrigierten Trennschärfen berechnet. Für die Gesamtskala wurde als Maß der inneren Konsistenz Cronbachs Alpha ermittelt. Als Statistikprogramm zur Durchführung dieser und aller folgenden Berechnungen diente SPSS® für Windows™, Version 6.0.1. Die Gesamtstichprobe war folgendermaßen soziodemografisch aufgebaut (Tabelle 2.4):

	männlich	weiblich	gesamt
Geschlecht	47	80	127
Alter (in Jahren)	AM = 24,2 s = 4,6 Min.: 20 Max.: 47	AM = 24,3 s = 3,9 Min.: 19 Max.: 42	AM = 24,3 s = 4,2 Min.: 19 Max.: 47

Tab. 2.4: Soziodemografische Daten der Stichprobe zur Aufgabenanalyse

Aufgabe	Schwierigkeit	Trennschärfe	Aufgabe	Schwierigkeit	Trennschärfe
Collage	,77	,07	Provision	,42	,21
Synonym	,86	,23	Relation	,37	,33
Gage	,75	,18	Dekade	,55	,44
Karosserie	,44	,29	Omen	,37	,37
Trilogie	,60	,21	Matinee	,46	,46
Erosion	,59	,35	Ballast	,35	,39
Prolog	,59	,41	Rabatte	,20	,25
Kartell	,43	,30	Pseudonym	,49	,39
Abstinenz	,61	,18	Relevanz	,39	,42
Saldo	,41	,20	Enklave	,35	,51
Trance	,56	,24	Opportunismus	,35	,53

Tab. 2.5: Ergebnisse der Analyse der Aufgaben des ILKO nach der Klassischen Testtheorie

Nach schrittweisem Ausschluss der Items mit der jeweils niedrigsten Trennschärfe (Collage, Gage, Abstinenz, Trilogie) ergab sich folgende Endform der Skala (Tabelle 2.6):

Aufgabe	Schwierigkeit	Trennschärfe	Aufgabe	Schwierigkeit	Trennschärfe
Synonym	,86	,21	Provision	,42	,23
Erosion	,59	,40	Saldo	,41	,25
Prolog	,59	,37	Relevanz	,39	,43
Trance	,56	,23	Omen	,37	,33
Dekade	,55	,43	Relation	,37	,32
Pseudonym	,49	,43	Enklave	,35	,52
Matinee	,46	,43	Opportunismus	,35	,52
Karosserie	,44	,29	Ballast	,35	,41
Kartell	,43	,32	Rabatte	,20	,25

Tab. 2.6: Kennwerte der Aufgaben der Endform des ILKO

Die Schwierigkeiten der Items liegen in einem Bereich von $,20 \leq p \leq ,59$ (mit Ausnahme des leichtesten Items). Die mittlere Schwierigkeit beträgt $p = ,45$. 15 der 18 Items der Testendform besitzen eine ausreichende Trennschärfe von $r_{it} \geq ,25$ (vgl. Moosbrugger & Hartig, 2003). Der Reliabilitätskoeffizient Cronbachs Alpha beträgt $r = ,78$.

5.4 Semantische Struktur der Skala

Nach Collins und Quillian (1969) sind die Inhalte des semantischen Langzeitgedächtnisses in einer hierarchischen Netzwerkstruktur von Kategorien (Bedeutungseinheiten, wie Tier, Vogel, Kanarienvogel) organisiert. Mit jeder Kategorie sind die für sie zutreffenden Eigenschaften verbunden (atmet, kann fliegen, ist gelb). Eine Eigenschaft, die für eine Kategorie zutrifft, trifft auch (mit begründeten Ausnahmen) auf alle hierarchisch tiefer liegenden Kategorien zu. Die Relationen zwischen den Kategorien (z. B. ‚ist ein‘) wird als Verknüpfung bezeichnet. Kategorien, die einen hohen Bedeutungszusammenhang aufweisen, sind nahe beieinander gespeichert. Der Abruf von relevanten semantischen Gedächtnisinhalten erfolgt durch die sich ausbreitende Aktivierung von Kategorien (Richardson, 1996b; Hahn & Heit, 2004; Lockhart, 2004). Wurde eine Kategorie aktiviert, so vergrößert sich die Wahrscheinlichkeit der nachfolgenden Aktivierung von Kategorien, die ihr semantisch nahe liegen (Lucas, 2003; Roediger III & Karpicke, 2005). Dieser Effekt wird als Priming bezeichnet und begründet die Annahme, dass für die Items des ILKO mit Hilfe einer explorativen Faktorenanalyse eine semantische Struktur aufgezeigt werden kann.

Um zu überprüfen, ob die Daten, die mit dem ILKO gewonnen wurden, eine Struktur aufzeigen, in der die 18 Wörter der Testendform nach ihrer semantischen Bedeutung angeordnet sind, soll eine explorative Faktorenanalyse mit den Daten der Stichprobe zur Trennschärferechnung (siehe Teil 2, Kapitel 5.3) durchgeführt werden.

Vorab ist zu klären, ob sich die vorliegenden Daten für eine Faktorenanalyse eignen. Hierzu werden der Bartlett-Test auf Sphärizität durchgeführt, die Anti-Image-Korrelationsmatrix erstellt, die MSA-Werte (Measures of Sampling Adequacy) gesichtet und das Kaiser-Meyer-Olkin-Maß berechnet (siehe hierzu auch Teil 1, Kapitel 6.3.1).

Die Korrelationen der 18 Variablen bewegen sich in einem Bereich von $r = ,00052$ bis $r = ,42323$. Um auszuschließen, dass in der Grundgesamtheit keinerlei Zusammenhang zwischen den Variablen besteht und sich die berechneten Korrelationen nur zufällig in der Stichprobe ergeben haben, wurde der Bartlett-Test auf Sphärizität durchgeführt. Für die vorliegenden Korrelationskoeffizienten ergibt sich ein Chi-Quadrat-Wert von $\chi^2 = 409,64987$, für den eine Wahrscheinlichkeit von $p = ,000$ ausgewiesen wird. Die Nullhypothese, alle Korrelationen zwischen den 18 Variablen seien in der Grundgesamtheit gleich Null, kann zurückgewiesen werden.

	Synonym	Erosion	Prolog	Trance	Dekade	Pseudonym	Matinee	Karosserie	Kartell
Synonym	,63111								
Erosion	-,18321	,78849							
Prolog	-,04553	,08157	,83951						
Trance	-,02319	-,15088	,12489	,74366					
Dekade	,07347	-,00856	-,15852	-,04682	,80309				
Pseudonym	,03602	-,11679	-,01444	-,04841	,03864	,72988			
Matinee	-,03321	-,08139	-,11142	-,02671	-,12259	,20893	,77975		
Karosserie	-,04958	-,03988	-,06895	,09295	-,14104	-,16330	-,04534	,73515	
Kartell	-,09580	-,08676	,00724	,08067	,07378	,06329	,03080	,07226	,70613
Provision	,00984	,04382	,05148	-,00630	-,02668	-,06912	-,09574	-,02651	-,16429
Saldo	,08496	,01781	-,02052	-,05567	-,08622	,07116	,05131	-,01020	-,26502
Relevanz	,00233	,05673	-,13821	-,02937	,14256	-,32446	-,21206	,02994	-,01259
Omen	-,08252	,01167	-,07115	,00648	-,12771	-,04559	-,06280	,02023	,08257
Relation	-,15504	-,03741	,00080	-,03083	-,19497	-,18607	,04608	,12317	,01881
Enklave	-,14694	-,02085	-,06988	-,08599	-,11583	-,21046	-,11016	-,09884	-,14534
Opportun.	-,05184	-,08031	-,04685	-,07850	-,10221	-,14143	-,18235	-,01206	-,15097
Ballast	,16314	-,14120	-,13417	-,14643	-,09148	-,01504	,00302	,00958	-,18264
Rabatte	,14731	-,23394	-,06645	,07757	-,06384	,02560	-,18032	-,30370	-,06723

	Provision	Saldo	Relevanz	Omen	Relation	Enklave	Opportun.	Ballast	Rabatte
Provision	,66320								
Saldo	-,06049	,72853							
Relevanz	-,11511	-,12023	,74540						
Omen	-,09130	,05885	-,07687	,84550					
Relation	,04041	,04539	-,09213	-,18903	,73910				
Enklave	-,21694	-,09919	,10195	-,06508	,14968	,80772			
Opportun.	,14287	-,05948	-,12657	-,02040	-,19495	-,09027	,84029		
Ballast	-,02438	-,01236	-,01462	-,01732	-,06828	-,08073	-,01925	,82470	
Rabatte	,18297	-,00586	-,08794	,01910	,01947	-,02128	,12014	-,05262	,65333

Tab. 2.7: Anti-Image-Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen

Die Sichtung der Anti-Image-Korrelationsmatrix in Tabelle 2.7 zeigt, dass die partiellen Korrelationskoeffizienten im Allgemeinen nahe Null liegen. Die MSA-Werte, die sich in der Hauptdiagonalen der Matrix befinden, liegen in einem Bereich zwischen ,63111 und ,84550 und sind als „mäßig“ bis „recht gut“ zu beurteilen. Das Kaiser-Mayer-Olkin-Maß hat einen Wert von ,76507. Obgleich die Eignung des Faktorenmodells zusammenfassend nur als „mittelprächtigt“ bezeichnet werden kann, wurde dennoch die Faktorenanalyse durchgeführt, da sie im Rahmen der Skalenanalyse lediglich einem heuristischen Zweck dient (Bewertungen zitiert nach Brosius, 1998, S. 647).

Zur Verrechnung der Daten wurde eine Hauptkomponentenanalyse mit anschließender orthogonaler Varimax-Rotation durchgeführt. Das Kaiser-Guttman-Kriterium diente zur Bestimmung der Anzahl der zu extrahierenden Faktoren.

Faktor	Eigenwert	Varianzaufklärung	VA kumuliert
1	3,93441	21,9%	21,9%
2	1,66175	9,2%	31,1%
3	1,53058	8,5%	39,6%
4	1,20138	6,7%	46,3%
5	1,08575	6,0%	52,3%
6	1,03290	5,7%	58,0%
7	,97361	5,4%	63,4%
8	,86957	4,8%	68,3%
9	,80062	4,4%	72,7%
10	,74838	4,2%	76,9%
11	,67622	3,8%	80,6%
12	,61564	3,4%	84,1%
13	,56613	3,1%	87,2%
14	,53253	3,0%	90,2%
15	,51180	2,8%	93,0%
16	,46198	2,6%	95,6%
17	,41690	2,3%	97,9%
18	,37986	2,1%	100,0%

Tab. 2.8: Ergebnis der Hauptkomponentenanalyse der Skala

Variable	Kommunalität	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6
Synonym	,70126	,07615	,04773	,06643	-,00934	,13078	,81950
Erosion	,66159	-,01875	,05868	,10466	,45627	,59651	,28780
Prolog	,49338	,59799	,11463	,19398	,24220	-,10577	-,12315
Trance	,48373	,01673	,09742	,08016	-,14146	,66439	,07815
Dekade	,63016	,73645	,07378	-,07208	,16413	,22252	-,02666
Pseudonym	,68463	,02823	,08025	,78694	,13482	,14092	,14171
Matinee	,42853	,50465	,17261	,06581	,35613	,06691	,09181
Karosserie	,60412	,16993	,06661	,12137	,72497	-,13529	,11038
Kartell	,55763	-,04112	,69701	,02643	,07907	,25029	,02275
Provision	,58330	,12689	,64660	,16376	-,15043	-,22642	,21999
Saldo	,51048	,04363	,63893	,04847	,01007	,14529	-,27710
Relevanz	,69110	,14672	,16670	,78946	,09714	-,00020	-,09540
Omen	,52160	,60555	-,04968	,26354	-,14009	,00406	,25168
Relation	,60467	,38662	-,21472	,45364	-,21895	,37108	,13290
Enklave	,58777	,27999	,57014	,11412	,23560	,11269	,32107
Opportunismus	,49675	,33837	,15388	,41696	,03533	,41130	,11958
Ballast	,52361	,28249	,30240	,07594	,13706	,50063	-,27781
Rabatte	,68245	,12626	-,03754	,02817	,78923	,09678	-,17906

(höchste Ladungen grau unterlegt)

Tab. 2.9: Kommunalitäten der Variablen und Ladungen der Variablen auf den rotierten Faktoren

Gemäß dem Kaiser-Guttman-Kriterium ergab sich eine 6-Faktoren-Lösung. Die Faktoren wurden orthogonal varimax-rotiert. Bei der Sichtung der höchsten Ladungen der einzelnen Variablen (Tabelle 2.9) fällt auf, dass sich die Wörter, wie vermutet, nach semantischen Bezügen auf den Faktoren gruppieren. Als inhaltliche Interpretation des semantischen Gehalts der Faktoren wird vorgeschlagen:

Faktor 1: ‚Zeit‘

Alle Wörter haben einen zeitlichen Bezug: *Prolog* – Vorrede, Vorwort. *Dekade* – Zeitraum. *Matinee* – am Vormittag stattfindend, *Omen* – Vorzeichen, Vorbedeutung.

Faktor 2: ‚Ökonomie‘

Alle Wörter haben einen Bezug zu ökonomischen Sachverhalten: *Kartell*, *Provision*, *Saldo*, *Enklave* (Staatsgebiet auch im Sinne einer wirtschaftlichen Einheit).

Faktor 3: ‚Zusammenhang‘

Alle Wörter drücken einen Zusammenhang aus: *Pseudonym* – Deckname, der einer Person zugehörig ist, *Relevanz* – Bedeutsamkeit, die einem Sachverhalt zugehörig ist, *Relation* – Zusammenhang zwischen zwei Begriffen, Dingen, usw.

Faktor 4: ‚Gestalt‘

Beide Wörter beziehen sich auf eine äußere Gestalt: *Karosserie* – Gestalt eines Kraftwagens, *Rabatte* – Gestalt einer Weg- oder Flächenbegrenzung.

Faktor 5: ‚Kraft‘

Alle Wörter beinhalten einen Bezug zu Kräften: *Erosion* – Wirkung von Kräften der Natur, *Trance* – Wirkung von Kräften des Unterbewusstseins, *Ballast* – Wirkung von Gewichtskräften.

Faktor 6: ‚Synonym‘

Dieser Faktor entspricht inhaltlich dem Begriff, der ihn konstituiert: bedeutungsgleiches Wort.

Der Begriff ‚Opportunismus‘ wurde zur Interpretation nicht herangezogen, da er auf den Faktoren 3 und 5 gleich hoch lädt.

6 Gütekriterien

Nach Kubinger (1996) und dem Testkuratorium (1986) sollen folgende Gütekriterien für das Inventar Lexikalischer Kompetenz überprüft werden: Objektivität, Reliabilität, Validität, Normierung, Nützlichkeit, Unverfälschbarkeit, Zumutbarkeit und Fairness.

6.1 Objektivität

Das Inventar Lexikalischer Kompetenz ist ein voll standardisiertes Verfahren. Sowohl die Durchführung des Tests als auch die Auswertung und Interpretation der Testergebnisse sind genau festgelegt und für den Testleiter in einer Handanweisung beschrieben. Befolgt der Testleiter die Vorschriften der Handanweisung, so kann das Testergebnis als unabhängig von seinem Verhalten bezeichnet werden.

6.2 Reliabilität

Der Reliabilitätskoeffizient Cronbachs Alpha beträgt für die Skala $r = ,77$ (siehe Teil 2, Kapitel 6.4). Dieser Wert kann als akzeptabel bezeichnet werden, zumal sich der Standardfehler der Reliabilitätsschätzung bei der gegebenen Stichprobengröße von $n = 581$ im Bereich von maximal $\pm ,04$ bewegt (Lienert & Raatz, 1994; Bartram & Lindley, 1994).

6.3 Validität

Zur Beurteilung der Validität der Inventars Lexikalischer Kompetenz sollen zunächst die Inhaltsvalidität untersucht und anschließend die Ergebnisse einer faktorenanalytischen Konstruktvalidierung, die mit allen neun Untertests des Intelligenz-Struktur-Tests 70 (IST 70; Amthauer, 1973) durchgeführt wurde, dargestellt werden. Weitere Belege für die Validität des Verfahrens werden in Teil 3 der vorliegenden Arbeit dargestellt.

6.3.1 Inhaltsvalidität

Ein Test ist inhaltlich valide, wenn seine Elemente so beschaffen sind, dass sie den zu erfassenden Messgegenstand repräsentieren (Lienert & Raatz, 1994). Als Verfahren zur Erfassung des Wortschatzes der deutschen Sprache ist die Inhaltsvalidität des Inventars Lexikalischer Kompetenz gewährleistet. Grund hierfür ist die Art der Generierung seiner Items: Die Repräsentativität der Zielwörter für den allgemeinen deutschen Wortschatz wurde sichergestellt, indem sie der Liste der 30.000 in den Textkorpora des Instituts für Deutsche

Sprache, Mannheim am häufigsten vorkommenden Wortformen entnommen wurden. Die Korpora ihrerseits wurden nach Kriterien zusammengestellt, die gewährleisten sollen, dass sie ein authentisches Dokument des Gebrauchs der deutschen Sprache darstellen. Ein weiterer Hinweis auf die Inhaltsvalidität des ILKO ergibt sich aus der inneren Struktur der Skala, die mittels einer explorativen Faktorenanalyse aufgezeigt wurde (siehe Teil 2, Kapitel 5.4): Die Zielwörter sind nach semantischen Gemeinsamkeiten auf den Faktoren lokalisiert.

6.3.2 Konstruktvalidität

Für eine erste Überprüfung der Konstruktvalidität wurde das Inventar Lexikalischer Kompetenz zusammen mit dem Intelligenz-Struktur-Test 70 einer Gruppe von insgesamt 73 Personen zur Bearbeitung vorgelegt. Die Daten wurden im Rahmen zweier eignungsdiagnostischer Untersuchungen der Berufsfeuerwehr einer hessischen Großstadt im Oktober 2004 und Oktober 2005 erhoben. Die Testpersonen waren Bewerber für den gehobenen Dienst und verfügten alle über einen Hochschulabschluss. 62 der Personen (84,9%) hatten ein ingenieurwissenschaftliches Studium, weitere 3 (4,1%) ein mathematisch-naturwissenschaftliches Studium absolviert.

Die Stichprobe baute sich folgendermaßen soziodemografisch auf:

	männlich	weiblich	gesamt
Geschlecht	68	5	73
Schulbildung			
Abitur	43	4	47
Fachhochschulreife	25	1	26
Alter (in Jahren)	AM = 28,0 s = 3,0 Min.: 23 Max.: 35	AM = 28,8 s = 2,2 Min.: 26 Max.: 31	AM = 28,1 s = 3,0 Min.: 23 Max.: 35

Tab. 2.10: Soziodemografische Daten der Stichprobe der Faktorenanalyse zur Konstruktvalidierung

In die Untersuchung gingen die in Abbildung 2.10 dargestellten Testrohwerte ein.

Intelligenz-Struktur-Test 70	Inventar Lexikalischer Kompetenz
IST-SE Satzergänzung IST-WA Wortauswahl IST-AN Analogien IST-GE Gemeinsamkeiten IST-ME Merkaufgaben IST-RA Rechenaufgaben IST-ZR Zahlenreihen IST-FA Figurenauswahl IST-WÜ Würfelaufgaben (jeweils Anzahl der richtig bearbeiteten Aufgaben)	ILKO Anzahl der richtig bearbeiteten Aufgaben

Abb. 2.10: Testwerte, die in die erste Untersuchung der Konstruktvalidität des ILKO Eingang fanden

Die Variablen ILKO und alle 9 Untertests des IST 70 sollen einer explorativen Faktorenanalyse unterzogen werden.

Zunächst muss geklärt werden, ob sich die vorliegenden Daten für eine Faktorenanalyse eignen. Nach Brosius (1998) wurde hierfür die Multikollinearität der Variablen überprüft, der Bartlett-Test auf Sphärizität berechnet, die Anti-Image-Korrelationsmatrix gesichtet und das Kaiser-Mayer-Olkin-Maß berechnet.

	ILKO	IST-SE	IST-WA	IST-AN	IST-GE	IST-ME	IST-RA	IST-ZR	IST-FA	IST-WÜ
ILKO	1,00000									
IST-SE	,28053	1,00000								
IST-WA	,10299	,07953	1,00000							
IST-AN	,36725	,37880	,26770	1,00000						
IST-GE	,21617	,35199	,33022	,49436	1,00000					
IST-ME	,14114	,08422	,25885	,35240	,30420	1,00000				
IST-RA	,16176	,32427	,13252	,27963	,32184	,38300	1,00000			
IST-ZR	-,03655	,10739	,18407	,28671	,16313	,06353	,35670	1,00000		
IST-FA	,07592	,14159	,31641	,34075	,26375	,27412	,27643	,13264	1,00000	
IST-WÜ	-,11315	-,01481	,03676	-,11029	,04254	,12780	,28598	,06469	,24683	1,00000

Tab. 2.11: Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen

Die für die vorliegende Stichprobe ausgewiesenen Korrelationskoeffizienten streuen zwischen $r = -,01481$ und $r = ,49436$. Um auszuschließen, dass sich in der Stichprobe Korrelationen nur zufällig ergeben haben, wurde der Bartlett-Test auf Sphärizität durchgeführt. Er testet die Hypothese, dass alle Korrelationskoeffizienten zwischen den Variablen in der Grundgesamtheit den Wert Null haben. Die Testgröße ist ein Chi-Quadrat-Wert. Für die vorliegenden Korrelationskoeffizienten ergibt sich ein Wert von $\chi^2 = 134,31956$, für den eine Wahrscheinlichkeit von $p = ,000$ ausgewiesen wird. Die Nullhypothese, alle Korrelationen zwischen den 10 Variablen seien in der Grundgesamtheit gleich Null, kann zurückgewiesen werden.

	ILKO	IST-SE	IST-WA	IST-AN	IST-GE	IST-ME	IST-RA	IST-ZR	IST-FA	IST-WÜ
ILKO	,69340									
IST-SE	-,13059	,72517								
IST-WA	-,04425	,04343	,73293							
IST-AN	-,27064	-,20716	-,00157	,69528						
IST-GE	,00140	-,17959	-,21680	-,29298	,81343					
IST-ME	,01021	,16144	-,14666	-,23779	-,07506	,68976				
IST-RA	-,11213	-,24904	,09266	,06763	-,11882	-,31914	,65855			
IST-ZR	,18333	,06907	-,14745	-,27499	,04058	,17823	-,34962	,51226		
IST-FA	,05652	,00360	-,22104	-,24321	-,00992	-,05403	-,09326	,04948	,74683	
IST-WÜ	,08498	,03426	,02461	,20874	-,03273	-,04810	-,24663	,00011	-,24134	,52257

Tab. 2.12: Anti-Image-Korrelationsmatrix der Variablen, die in die Faktorenanalyse einfließen sollen

Die Sichtung der Anti-Image-Korrelationsmatrix (Tabelle 2.12) zeigt, dass die Werte überwiegend nahe Null liegen.

Ein zusammenfassendes Maß für die Eignung des Faktorenmodells, in das die partiellen Korrelationskoeffizienten einfließen, ist das Kaiser-Mayer-Olkin-Maß (KMO). Im vorliegenden Fall ist der KMO-Wert = 0,69231. Nach Kaiser sind Werte ab 0,6 bis unter 0,7 als „mäßig“ zu beurteilen (diese und die folgenden Beurteilungen zitiert nach Brosius, 1998, S. 647). Abschließend sollen die MSA-Werte, die sich in der Hauptdiagonalen der Matrix befinden, betrachtet werden. Sie entsprechen inhaltlich dem KMO-Maß, beziehen sich aber jeweils nur auf eine Variable. Werte ab 0,8 bis unter 0,9 gelten als „recht gut“, ab 0,7 bis unter 0,8 als „mittelprächtigt“ ab 0,6 bis unter 0,7 als „mäßig“ und ab 0,5 bis unter 0,6 als „schlecht“. Unter 0,5 sind die Werte „unakzeptabel“.

Die Untersuchungen zeigen, dass die 10 Variablen zur Berechnung einer Faktorenanalyse geeignet sind. Als Methode zur Verrechnung der Daten diene eine Hauptkomponentenanalyse mit anschließender orthogonaler Varimax-Rotation. Zur Bestimmung der Anzahl der bedeutsamen Faktoren diene das Kaiser-Guttman-Kriterium. Tabelle 2.13 und Tabelle 2.14 enthalten die Ergebnisse. Auf Grund der kleinen Stichprobengröße ist die Stabilität der gewonnenen Faktorenstruktur als eher gering zu bewerten (Bortz, 1993).

Faktor	Eigenwert	Varianzaufklärung	VA kumuliert
1	3,01462	30,1	30,1
2	1,41695	14,2	44,3
3	1,06235	10,6	54,9
4	1,00382	10,0	65,0
5	,80329	8,0	73,0
6	,70833	7,1	80,1
7	,66088	6,6	86,7
8	,55912	5,6	92,3
9	,45110	4,5	96,8
10	,31953	3,2	100,0

Tab. 2.13: Ergebnis der Hauptkomponentenanalyse

Variable	Kommunalität	Ladung			
		Faktor 1	Faktor 2	Faktor 3	Faktor 4
ILKO	,59778	,71298	,10673	-,11271	-,25561
IST-SE	,62498	,76130	-,05386	,12054	,16722
IST-WA	,67709	-,04786	,79266	-,14373	,16073
IST-AN	,68637	,61141	,48546	-,12248	,24876
IST-GE	,52301	,49453	,49880	,04335	,16662
IST-ME	,50948	,21897	,58774	,32047	-,11568
IST-RA	,71642	,41367	,15528	,61205	,38286
IST-ZR	,87691	,04309	,11241	,05583	,92699
IST-FA	,52621	,06885	,63405	,34407	,03249
IST-WÜ	,75949	-,17305	,07494	,85061	-,01983

(höchste Ladungen grau unterlegt)

Tab. 2.14: Kommunalitäten der Variablen und Ladungen der Variablen auf den rotierten Faktoren

Wenn man die kognitiven Anforderungen und Prozesse in Erwägung zieht, die bei der Bearbeitung der einzelnen Tests maßgebend sind, und nicht die Art des Testmaterials (verbal, numerisch, räumlich), bieten sich folgende inhaltliche Interpretationen der rotierten Faktoren an:

Faktor 1: ‚Sprachgebundene kristalline Intelligenz‘

ILKO, IST 70-Analogien und IST 70-Satzergänzung repräsentieren primäre mentale Fähigkeiten, die in der erweiterten Gf-Gc Theorie von Cattell-Horn (siehe Teil 2, Kapitel 2.1.2) gemeinsam mit anderen primären mentalen Fähigkeiten die Fähigkeit zweiter Ordnung Gc (kristalline Intelligenz) konstituieren (Horn & Blankson, 2005, S. 44 f):

ILKO und IST 70-Analogien: „V – Verbal comprehension: Demonstrate understanding of words, sentences, paragraphs“

IST 70-Satzergänzung: „Vi – General information: Indicate understanding of a wide range of information“

Faktor 2: ‚Fluide Intelligenz und kurzzeitiges Memorieren‘

IST 70-Wortauswahl und IST 70-Figurenauswahl repräsentieren innerhalb der erweiterten Gf-Gc Theorie von Cattell-Horn primäre mentale Fähigkeiten, die gemeinsam mit anderen primären mentalen Fähigkeiten die Fähigkeit zweiter Ordnung Gf (fluide Intelligenz) konstituieren (Horn & Blankson, 2005, S. 44 f):

IST 70-Wortauswahl: „CMR – Semantic relations: Demonstrate awareness of relationships among pieces of information“

IST 70-Figurenauswahl: „CFR – Figural relations: Solve problems of relationships among figures“

Bei beiden Untertests muss aus mehreren vorgegebenen Elementen auf ein Gemeinsames/Übergeordnetes geschlossen werden. Um derartige Aufgaben lösen zu können, bedarf es der kurzzeitigen Merkfähigkeit: Sie hat auf Faktor 2 ebenfalls ihre höchste Ladung.

Faktor 3: ‚Sukzessives Problemlösen‘

Faktor 3 lässt sich mit Hilfe der erweiterten Gf-Gc-Theorie inhaltlich nicht sinnvoll interpretieren: Der IST 70-Untertest Rechenaufgaben entspricht der Primärfähigkeit „CMS – Algebraic reasoning: Find solutions for problems that can be framed algebraically“. Diese ist der Fähigkeit zweiter Ordnung „Gq – Quantitative mathematical abilities“ zugeordnet. Der IST 70-Untertest Würfelaufgaben findet sich unter der Fähigkeit zweiter Ordnung „Gv – Visualization and spatial orientation abilities“ als Primärfähigkeit „Vi – Visualization: Mentally manipulate forms to ‚see‘ how they would look under altered conditions“ (alle Zitate nach Horn & Blankson, 2005, S. 44 f).

Um eine mögliche Gemeinsamkeit bei der Aufgabenbearbeitung der beiden Untertests zu finden, sei die jeweilige Aufgabenstellung kurz beschrieben: Der IST 70-Untertest Rechenaufgaben besteht aus mathematischen Dreisatzaufgaben in Textform. Beim IST 70 Untertest Würfelaufgaben müssen vorgegebene Würfel, auf deren Seitenflächen unterschiedliche Zeichen abgebildet sind, in drei Dimensionen mental rotiert werden, um entscheiden zu können, welchem von 5 vorgegebenen Würfeln sie gleichen.

Die Gemeinsamkeit bei der Bearbeitung der Untertests könnte darin liegen, dass bei beiden Aufgabenstellungen zuerst Zwischenergebnisse ermittelt werden müssen, um zur Aufgabenlösung zu gelangen. Als Bezeichnung des Faktors wird daher - unter Vorbehalt - ‚Sukzessives Problemlösen‘ vorgeschlagen.

Faktor 4: ‚Zahlenreihen‘

Dieser Faktor entspricht inhaltlich der Variablen, die ihn konstituiert. Amthauer (1973, S. 39) beschreibt den IST 70-Subtets Zahlenreihen inhaltlich als „Theoretisch-rechnerisches Denken“.

Der IST 70-Untertest Gemeinsamkeiten wurde zur Interpretation nicht herangezogen, da er auf den Faktoren 1 und 2 etwa gleich hoch lädt.

Fazit

Die explorative Faktorenanalyse mit den neun Untertests des IST 70 zeigt, dass das Inventar Lexikalischer Kompetenz sprachgebundene kristalline Intelligenz erfasst.

6.4 Normierung

Die Stichprobe des ILKO umfasste zum Zeitpunkt des Abschlusses der vorliegenden Arbeit insgesamt $n = 581$ Personen. Sie wurde im Zeitraum Dezember 2003 bis April 2006 an Universitäten in Hessen, Rheinland-Pfalz und Sachsen-Anhalt, einer Verwaltungsfachhochschule in Baden-Württemberg und im Rahmen eines Personalauswahlverfahrens einer hessischen Berufsfeuerwehr gewonnen. Die Testpersonen hatten alle Abitur ($n = 544$; 93,6%) bzw. Fachhochschulreife ($n = 37$; 6,4%). Sie waren Studenten unterschiedlicher Fachrichtungen ($n = 478$; 82,3%) oder hatten bereits einen Hochschulabschluss ($n = 103$; 17,7%).

In die Testentwicklungs- und Normstichprobe gingen nur die Ergebnisse von den Testpersonen ein, deren Muttersprache Deutsch war. Der Grund hierfür lag darin, dass die Testleistungen bei allen Personen, die aus dem nicht-deutschsprachigen Ausland kamen und eine andere Muttersprache als Deutsch angegeben hatten, sehr schlecht ausfielen (siehe Tabelle 2.15).

Nation	Geschlecht	Alter in Jahren	ILKO-Rohwert
Bulgarien	w	20	1
Bulgarien	w	22	5
China	w	23	0
China	w	24	1
China	w	29	0
Israel	w	23	3
Japan	w	k. A.	1
Russland	w	20	3
Russland	w	24	5
Russland	w	25	4
Russland	w	25	5
Russland	m	27	2
Spanien	w	23	4
Spanien	w	25	3
Spanien	w	27	5
Spanien	w	31	3
Spanien	w	33	0
Türkei	w	30	6
Türkei	w	48	1

Tab. 2.15: Soziodemografische Daten und Testergebnisse der Testpersonen mit einer anderen Muttersprache als Deutsch

Von den 18 Aufgaben des ILKO wurden von den Personen mit einer anderen Muttersprache als Deutsch ($n = 19$) im Durchschnitt $AM = 2,74$ Aufgaben richtig gelöst ($s = 1,97$). Die maximale Anzahl richtiger Lösungen lag bei 6, die minimale Anzahl bei 0. Dieses äußerst schlechte Ergebnis führte dazu, dass für den Einsatz des Verfahrens eine weitere Bedingung aufgestellt wurde: ILKO darf nur bei Personen angewendet werden, deren Muttersprache Deutsch ist.

Die (vorläufige) Normstichprobe des ILKO baut sich soziodemografisch wie folgt auf:

	männlich	weiblich	gesamt
Geschlecht	257 (44,2%)	324 (55,8%)	581 (100%)
Schulbildung			
Fachhochschulreife	31 (12,1%)	6 (1,9%)	37 (6,4%)
Abitur	226 (87,9%)	318 (98,1%)	544 (93,6%)
Alter (in Jahren)	AM = 25,3 s = 4,7 Min.: 19 Max.: 61	AM = 23,6 s = 3,9 Min.: 19 Max.: 53	AM = 24,4 s = 4,4 Min.: 19 Max.: 61

Tab. 2.16: Soziodemografische Daten der Personen der Normstichprobe des ILKO

In der Normstichprobe beträgt der Mittelwert des Testrohweres $AM = 7,64$, die Standardabweichung $s = 3,48$. Der maximale Testrohwerwert ist 17, der minimale 0.

Der zweiseitige Kolmogorov-Smirnov-Test auf Normalverteilung der Rohwerte ergibt ein K-S-Z von 2,3589. Dieser Wert besitzt eine Wahrscheinlichkeit von $p = ,000$. Die Nullhypothese, dass die Testrohwerwerte normalverteilt sind, muss damit zurückgewiesen werden. Trotz dieser geringen Irrtumswahrscheinlichkeit ist es dennoch möglich, dass die Werte in der Grundgesamtheit annähernd normalverteilt sind (Brosius, 1998). Um einen Eindruck zu gewinnen, ob die empirische Verteilung der Normalverteilung ähnelt, wurde ein Histogramm mit Normalverteilungskurve erstellt (siehe Abbildung 2.11).

Das Histogramm der Rohwerteverteilung zeigt tendenziell eine Normalverteilung der Werte. Gegenüber der Standardnormalverteilung ist die gefundene Verteilung flacher (Exzess = $-,585$) und aufgrund der relativ großen Anzahl an leichten Aufgaben (siehe Tabelle 2.17) rechtsschief (Schiefe = $,225$).

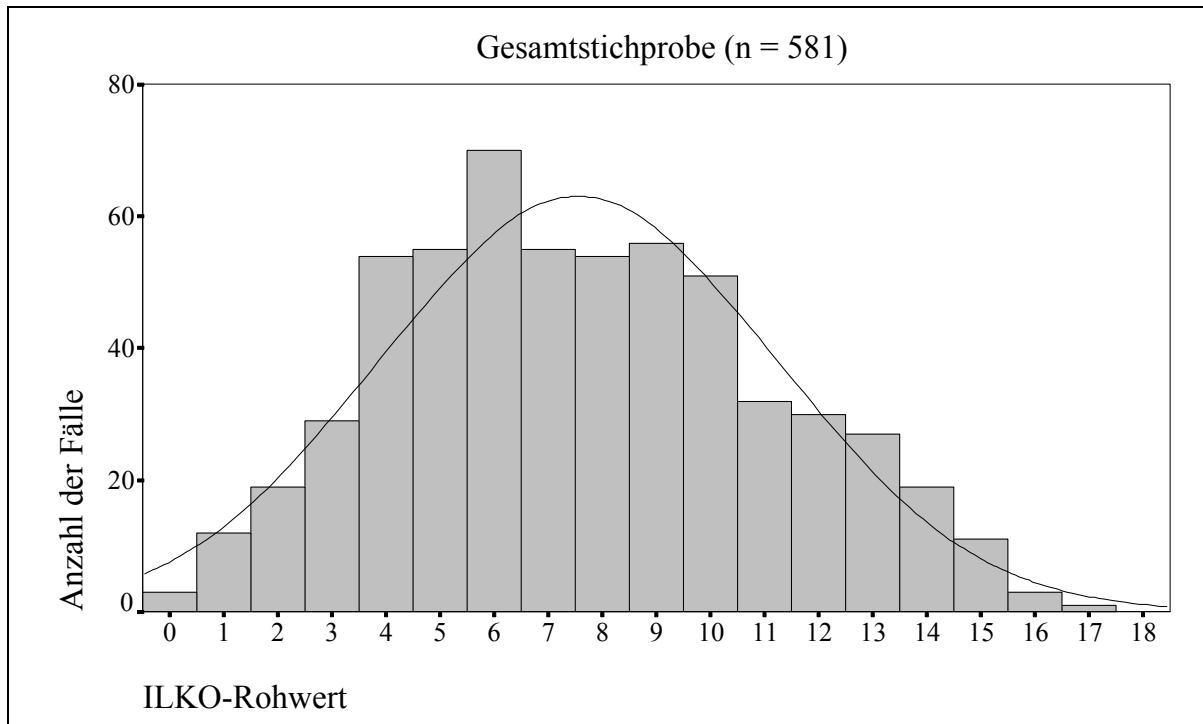


Abb. 2.11: Histogramm der Rohwertverteilung in der Normstichprobe des ILKO

Mit Hilfe der Daten der Normstichprobe wurden noch einmal die Schwierigkeiten und Trennschärfen der Aufgaben und die Reliabilität der Skala berechnet. Die Ergebnisse enthält Tabelle 2.17.

Aufgabe	Schwierigkeit	Trennschärfe	Aufgabe	Schwierigkeit	Trennschärfe
Synonym	,91	,14	Dekade	,72	,37
Erosion	,88	,29	Matinee	,65	,38
Provision	,83	,32	Relation	,63	,42
Kartell	,82	,34	Omen	,60	,38
Prolog	,82	,30	Enklave	,54	,38
Karosserie	,80	,34	Ballast	,52	,31
Saldo	,79	,38	Opportunismus	,44	,48
Trance	,77	,26	Relevanz	,43	,36
Pseudonym	,73	,34	Rabatte	,39	,31

Tab. 2.17: Kennwerte der Aufgaben des ILKO, n = 581

Die Schwierigkeiten der Items liegen für die Normstichprobe in einem Bereich von $,39 \leq p \leq ,91$. Die mittlere Schwierigkeit beträgt $p = ,68$. 17 der 18 Items besitzen eine ausreichende Trennschärfe von $r_{it} \geq ,25$ (vgl. Moosbrugger & Hartig, 2003). Der Reliabilitätskoeffizient Cronbachs Alpha beträgt $r = ,77$.

Zur Klärung der Frage, ob sich die Testergebnisse zwischen Teilgruppen der Testpersonen signifikant voneinander unterscheiden und damit teilgruppenspezifische Normen erstellt werden müssen, wurden Mittelwertsvergleiche berechnet. Die Gruppierungsvariablen waren das Geschlecht, das Alter, die in der Schule erlernten Fremdsprachen und das Studienfachgebiet der Testpersonen.

6.4.1 Leistungsunterschiede zwischen Personen unterschiedlichen Geschlechts

Zur Überprüfung, ob signifikante Mittelwertsunterschiede in der Bearbeitungsleistung im ILKO zwischen weiblichen und männlichen Testpersonen bestehen, wurde ein t-Test berechnet. Der Levene-Test auf Varianzgleichheit ergab einen F-Wert von $F = ,038$. Dieser Wert hat eine 2-seitige Wahrscheinlichkeit von $p = ,845$. Die Nullhypothese, dass die weiblichen und die männlichen Testpersonen aus Populationen mit gleichen Varianzen stammen, darf beibehalten werden.

Gruppenstatistiken

Geschlecht	n	AM	s	Standardfehler des AM
weiblich	324	7,0278	3,451	,192
männlich	257	8,4086	3,358	,209

Test bei unabhängigen Stichproben mit gleicher Varianz

t-Wert	df	Sig. (2-seitig)	Differenz der AM	SE der Diff.	95%-KI der Diff.
-4,85	579	,000	-1,3808	,285	-1,940; -,821

Tab. 2.18: Ergebnis des t-Tests mit der abhängigen Variablen ‚ILKO-Testwert‘ und der Gruppierungsvariablen ‚Geschlecht‘

Der ermittelte t-Wert ist hoch signifikant. Die Nullhypothese, dass weibliche und männliche Testpersonen aus Populationen mit gleichen Mittelwerten stammen, muss verworfen werden. Männliche Testpersonen schneiden signifikant besser ab als weibliche Testpersonen.

6.4.2 Leistungsunterschiede zwischen Personen unterschiedlichen Alters

Zur Klärung der Frage, ob Leistungsunterschiede bei der Bearbeitung des ILKO zwischen Personen unterschiedlichen Alters bestehen, wurde zunächst ein Balkendiagramm erstellt, das die mittlere Testleistung für jede Altersstufe grafisch darstellt (siehe Abbildung 2.12).

Wie zu erkennen ist, steigt die Leistungsfähigkeit der Testpersonen mit zunehmenden Alter tendenziell an. Dieses Ergebnis ist kongruent mit den Erkenntnissen des altersbezogenen Verlaufs der kristallinen Intelligenz sensu Cattell (Horn & Cattell, 1967; Horn & Blankson, 2005). Allerdings sind insbesondere die Kategorien von 31 bis 61 Jahren nur gering besetzt (siehe Tabelle 2.19), so dass das Ergebnis in diesem Bereich als instabil anzusehen ist.

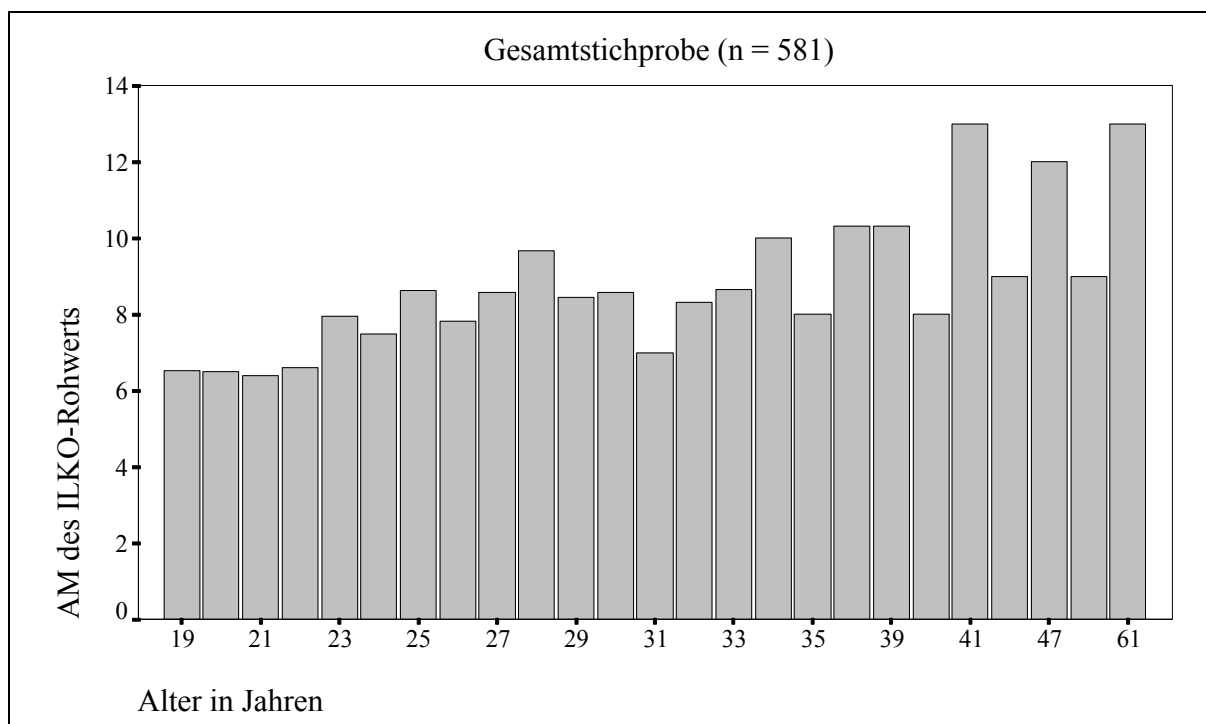


Abb. 2.12: Balkendiagramm der mittleren ILKO-Testwerte bezogen auf die Variable ‚Alter‘

Alter	19	20	21	22	23	24	25	26	27	28	29	30	31
n	17	67	52	60	98	71	50	34	34	19	20	19	10
Alter	32	33	34	35	38	39	40	41	46	47	53	61	
n	3	3	1	6	3	3	1	1	1	1	1	1	

Tab. 2.19: *Kategorienhäufigkeiten der Altersverteilung*

Zur Überprüfung, ob die Mittelwertsunterschiede in der Bearbeitungsleistung zwischen Testpersonen unterschiedlichen Alters Signifikanz besitzen, wurde die Stichprobe am Median der Alterskategorien (31 Jahre) geteilt und anschließend für die beiden resultierenden Altersgruppen ein t-Test für unabhängige Stichproben berechnet.

Der Levene-Test auf Varianzgleichheit ergab einen F-Wert von $F = ,273$. Dieser Wert hat eine 2-seitige Wahrscheinlichkeit von $p = ,602$. Die Nullhypothese, dass die jüngeren und die älteren Testpersonen aus Populationen mit gleichen Varianzen stammen, darf beibehalten werden.

Gruppenstatistiken

Alter	n	AM	s	Standardfehler des AM
19 bis 31 Jahre	551	7,5572	3,469	,148
32 bis 61 Jahre	25	9,4000	3,266	,653

Test bei unabhängigen Stichproben mit gleicher Varianz

t-Wert	df	Sig. (2-seitig)	Differenz der AM	SE der Diff.	95%-KI der Diff.
2,60	574	,009	1,8428	,708	,453; 3,233

Tab. 2.20: *Ergebnis des t-Tests mit der abhängigen Variablen ‚ILKO-Testwert‘ und der Gruppierungsvariablen ‚Alter‘*

Der ermittelte t-Wert ist hoch signifikant. Die Nullhypothese, dass ältere und jüngere Testpersonen aus Populationen mit gleichen Mittelwerten stammen, muss verworfen werden. Ältere Testpersonen schneiden signifikant besser ab als jüngere Testpersonen.

6.4.3 Leistungsunterschiede zwischen Personen mit unterschiedlichen Schulfremdsprachen

Wie bereits in Teil 2, Kapitel 5 dargestellt, wurden die Kenntnisse der Schulfremdsprachen erhoben, um zu klären, ob die Verfügbarkeit fremdsprachlicher Wortschätze einen signifikanten Einfluss auf die Verfügbarkeit der deutschen Zielwörter hat. Die Übersetzung der (deutschen) Zielwörter ins Französische und Englische (siehe Tabelle 2.1) zeigt, dass 13 der 18 Zielwörter der Testendform in allen drei Sprachen denselben Wortstamm besitzen.

Die Frage, ob Leistungsunterschiede in der Bearbeitung des ILKO zwischen Personen mit unterschiedlichen Schulfremdsprachen-Kenntnissen bestehen, soll nachfolgend mit Hilfe einer einfaktoriellen Varianzanalyse geklärt werden.

Von denjenigen Testpersonen, die die Frage nach der ersten Schulfremdsprache beantworteten (n = 543), gaben die weitaus meisten (n = 510; 93,9%) Englisch als erste Fremdsprache an. Daher soll sich die Überprüfung der Mittelwertsunterschiede der Testleistung im ILKO auf die zweite Schulfremdsprache als Gruppierungsvariable bei Personen mit Englisch als erster Schulfremdsprache beschränken. Tabelle 2.21 gibt die Antwort dieser Personen auf die Frage nach ihrer zweiten Fremdsprache wieder:

Zweite Schulfremdsprache	n (%)	Zweite Schulfremdsprache	n (%)
Französisch	299 (58,6%)	Italienisch	1 (0,2%)
Latein	129 (25,3%)	Sonstige	8 (1,6%)
Spanisch	19 (3,7%)	keine Angabe	31 (6,1%)
Russisch	23 (4,5%)	gesamt	510 (100%)

Tab. 2.21: Zweite Schulfremdsprache der Personen, die Englisch als erste Schulfremdsprache angaben

Da nur eine Person Italienisch als zweite Fremdsprache angegeben hatte, wird diese Gruppe nicht in die einfaktorielle Varianzanalyse eingehen. Tabelle 2.22 enthält die deskriptiven Statistiken des ILKO-Rohwerts für die restlichen vier Gruppen:

Zweite Fremdsprache	n	AM	s	SE des AM	95%-KI für das AM
Französisch	299	7,7023	3,3098	,1914	7,3257; 8,0790
Latein	129	8,2558	3,7714	,3320	7,5988; 8,9128
Spanisch	19	5,1053	2,9795	,6835	3,6692; 6,5413
Russisch	23	7,8261	3,5885	,7483	6,2743; 9,3779
gesamt	470	7,7553	3,4846	,1607	7,4395; 8,0712

Tab. 2.22: Deskriptive Statistiken des ILKO-Rohwerts für die vier Fremdsprachengruppen

Der Levene-Test auf Gleichheit der Gruppenvarianzen ergibt einen F-Wert von $F = 1,3386$ ($df_1 = 3$, $df_2 = 466$). Dieser Wert besitzt eine 2-seitige Wahrscheinlichkeit von $p = ,261$. Die Nullhypothese, dass alle Gruppenvarianzen in der Grundgesamtheit gleich seien, darf beibehalten und es darf eine Varianzanalyse berechnet werden.

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
zwischen den Gruppen	166,7014	3	55,5671	4,6841	,0031
innerhalb der Gruppen	5528,1603	466	11,8630		
gesamt	5694,8617	469			

Tab. 2.23: Ergebnis der einfaktoriellen Varianzanalyse mit der abhängigen Variablen ‚ILKO-Testwert‘ und der Faktorvariablen ‚Zweite Fremdsprache‘

Der für den Gruppenvergleich ermittelte F-Wert von $F = 4,6841$ besitzt eine Auftretenswahrscheinlichkeit von $p = ,0031$. Die Nullhypothese, dass die Mittelwerte der

ILKO-Testleistungen der vier Gruppen in der Grundgesamtheit identisch sind, muss zurückgewiesen werden.

Um zu klären, zwischen welchen Gruppen signifikante Leistungsunterschiede bestehen, wurden Einzelvergleiche mit Hilfe des Scheffé-Tests durchgeführt. Das Ergebnis zeigt Tabelle 2.24.

	Latein	Spanisch	Russisch
Französisch		*	
Latein		*	
Spanisch			
Russisch			

Tab. 2.24: Ergebnis der Einzelvergleiche mit Hilfe des Scheffé-Tests

Der Scheffé-Test ergibt, dass zwischen den Gruppen ‚Spanisch‘ und ‚Französisch‘ und ‚Spanisch‘ und ‚Latein‘ jeweils ein signifikanter Unterschied auf dem 5%-Niveau in der Testleistung im ILKO besteht. Für die beiden am häufigsten an deutschen Schulen gelehrteten Zweitfremdsprachen, Französisch und Latein (siehe Tabelle 2.2), zeigt die ILKO-Testleistung jedoch keinen signifikanten Unterschied.

6.4.4 Leistungsunterschiede zwischen Personen unterschiedlicher Studienfachgebiete

Abschließend soll die Frage geklärt werden, ob Leistungsunterschiede in der Bearbeitung des ILKO zwischen Personen unterschiedlicher Studienfachgebiete bestehen. Hierzu soll ebenfalls eine Varianzanalyse berechnet werden.

Die Testpersonen der Normstichprobe waren Studenten folgender Studienfachgebiete (Tabelle 2.25):

Gruppe	Studienfachgebiet	n (%)
1	Ingenieurwissenschaften	92 (15,8%)
2	Mathematik, Naturwissenschaften	83 (14,3%)
3	Agrar- und Forstwissenschaften	3 (0,5%)
4	Medizin, Gesundheitswesen	8 (1,4%)
5	Rechts- und Wirtschaftswissenschaften	75 (12,9%)
6	Gesellschafts- und Sozialwissenschaften	165 (28,4%)
7	Sprach- und Kulturwissenschaften, Kunst und Gestaltung	72 (12,4%)
8	Verwaltungswissenschaften	75 (12,9%)
k. A.		8 (1,4%)
gesamt		581 (100%)

Tab. 2.25: Studienfachgebiete der Personen der Normstichprobe des ILKO

Aufgrund der geringen Kategorienbesetzung der Gruppen ‚Agrar- und Forstwissenschaften‘ und ‚Medizin, Gesundheitswesen‘ soll mit ihrer Ausnahme eine einfaktorielle Varianzanalyse berechnet werden.

Gruppe	n	AM	s	SE des AM	95%-KI für das AM
1	92	8,0761	3,3124	,3453	7,3901; 8,7621
2	83	7,5904	3,1893	,3501	6,8940; 8,2868
5	75	8,3200	3,6099	,4168	7,4894; 9,1506
6	165	7,0121	3,1507	,2453	6,5278; 7,4964
7	72	8,0833	4,1782	,4924	7,1015; 9,0652
8	75	7,6400	3,7476	,4327	6,7778; 8,5022
gesamt	562	7,6673	3,4904	,1472	7,3781; 7,9565

Tab. 2.26: Deskriptive Statistiken des ILKO-Rohwerts für die sechs Fachgebiete

Der Levene-Test auf Gleichheit der sechs Gruppenvarianzen ergab einen F-Wert von $F = 4,0039$ ($df_1 = 5$, $df_2 = 556$). Dieser Wert besitzt eine 2-seitige Wahrscheinlichkeit von $p = ,001$. Die Nullhypothese, dass alle Gruppenvarianzen in der Grundgesamtheit gleich

sein, muss zurückgewiesen werden. Statt einer Varianzanalyse muss zur statistischen Überprüfung der Mittelwertsunterschiede ein verteilungsfreies Verfahren, wie z. B. der Kruskal-Wallis-Test, verwendet werden (Bortz, 1993).

Das Kruskal-Wallis-H, das annähernd χ^2 -verteilt ist, überprüft die Nullhypothese, dass alle Stichproben derselben Grundgesamtheit entstammen. Hierzu werden zuerst alle Werte in eine gemeinsame Rangordnung gebracht. Anschließend wird mit einem χ^2 -Test überprüft, ob die durchschnittlichen Rangwerte in den einzelnen Stichproben gleich groß sind (Brosius, 1998).

Gruppe	n	mittlerer Rang	Gruppe	n	mittlerer Rang
1	92	300,73	6	165	251,96
2	83	278,84	7	72	297,32
5	75	311,07	8	75	281,08

Tab. 2.27: Mittlere Ränge der INKA-Testleistungen für die sechs Fachgebiete

Für die mittleren Ränge (siehe Tabelle 2.27) ergibt sich ein hinsichtlich verbundener Ränge korrigierter χ^2 -Wert von 10,0180 (df = 5). Dieser Wert besitzt eine Wahrscheinlichkeit von $p = ,0747$. Die Nullhypothese darf beibehalten werden: Die mittleren Ränge der Testleistung im INKA sind, bezogen auf die Studienfachgebiete, in der Grundgesamtheit gleich. Die Unterschiede in den Testergebnissen der Gruppen von Studenten unterschiedlicher Studienfachgebiete sind nicht signifikant.

6.5 Nützlichkeit

Mit dem Inventar Lexikalischer Kompetenz wird ein Verfahren vorgestellt, das über die Messung des aktiven Wortschatzes ein Maß der verbalen kristallinen Intelligenz liefert. Im Sinne der Nützlichkeit ist zu überprüfen, ob bereits Tests verfügbar sind, die denselben Messgegenstand haben bzw. dasselbe Konstrukt erfassen (Lienert & Raatz, 1994; Kubinger, 1995). Im Folgenden werden die für den deutschen Sprachraum publizierten, aktuell

erhältlichen Einzeltestverfahren zur Messung des Wortschatzes und der kristallinen bzw. verbalen Intelligenz bei Jugendlichen und Erwachsenen vorgestellt. Im Einzelnen sind das:

- Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B; Lehrl, 1995)

- Wortschatztest (WST; Schmidt & Metzler, 1992)

- Verbaler Kurz-Intelligenztest (VKI; Anger, Mertesdorf, Wegener & Wülfing, 1980)

- Wortschatztest aktiv und passiv - Deutsche Version des Mill Hill Vocabulary Tests von J.C. Raven (WST-ap; Ibrahimovic & Bulheller, 2005a)

- Peabody Picture Vocabulary Test - Deutschsprachige Fassung des PPVT-III für Jugendliche und Erwachsene von L. M. Dunn & L. M. Dunn (PPVT; Bulheller & Häcker, 2003)

Alle nachfolgenden Angaben (Seiten 130 bis 134) sind den Manualen der Verfahren entnommen.

Verfahren	Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B)
Autor(en), Erscheinungsjahr, -ort, Verlag	S. Lehrl, 3. überarbeitete Auflage, 1995, Balingen, Perimed-Spitta
Messgegenstand	Allgemeines (kristallines/prä-morbides) Intelligenzniveau, „... die Spuren der geistigen Auseinandersetzung eines Menschen mit seiner Umwelt, die sich als Fertigkeiten und Wissen niedergeschlagen, kristallisiert, haben“ (Lehrl, 1995, S. 16).
Anwendungs- bereiche	Erwachsene ab 20 Jahren im Bereich der Psychiatrie, psychiatrisch-klinischen Psychologie und Neurologie und Arbeits- und Sozialmedizin. <i>Explizite Einschränkung:</i> Nicht Anwendbar bei Personen, deren Muttersprache nicht Deutsch ist, bei bestimmten Formen der Legasthenie, bei Sehschwachen und Personen mit schweren seelisch-geistigen Störungen.
Theoretischer Hintergrund	Gf-Gc Theory von R. B. Cattell. Der MWT-B zielt darauf ab, den kristallinen Aspekt der Intelligenz zu erfassen. Er spricht im Wesentlichen zwei psychische Funktionen an: Bekanntes wieder zu erkennen und Bekanntes von Unbekanntem zu unterscheiden.
Operationalisierung/ Aufgabenprinzip	Ein umgangs-, bildungs- oder wissenschaftssprachlich bekanntes Wort, das sich mit vier sinnlosen Wörtern in einer Reihe (Wortzeile) befindet, soll durchgestrichen werden. Der Gesamttest besteht aus 37, in der Schwierigkeit steigenden Wortzeilen.
Entwicklung (Aufgaben, Skala)	Die MWT- Erstform wurde 1971 entwickelt und bestand aus den 42 richtigen Lösungen des HAWIE-Wortschatztests. Der MWT-B mit 37 Wortzeilen stellt eine später konstruierte, von den HAWIE-Lösungswörtern unabhängige Testform dar.
Durchführung	Einzel- und Gruppentestung, keine Zeitbegrenzung (Bearbeitungsdauer bei psychiatrisch unauffälligen Personen ca. 5 Minuten).
Psychometrische Eigenschaften	<i>Objektivität</i> Vollstandardisiertes Verfahren <i>Reliabilität</i> Altersstabilität der Ergebnisse ab dem 20. Lebensjahr. <i>Validität</i> Durchschnittlicher Korrelationskoeffizient (Median von 26 Untersuchungen) zwischen MWT und anderen globalen Intelligenztests von $r = ,71$. <i>Normierung</i> Gesamtnorm (repräsentative Zufallsstichprobe von 1.952 20-60jährigen Einwohnern der alten Bundesländer der Bundesrepublik Deutschland; Vergleichswerte: Prozentränge, IQ- und Z-Werte)

Verfahren	Wortschatztest (WST)
Autor(en), Erscheinungsjahr, -ort, Verlag	K.-H. Schmidt & P. Metzler, 1992, Weinheim, Beltz
Messgegenstand	Allgemeines (kristallines/prämorbides) Intelligenzniveau, allgemeines Sprachverständnis.
Anwendungs- bereiche	Jugendliche und Erwachsene ab 16 Jahren im Bereich der klinischen Psychologie, Psychiatrie und Neuropsychologie. <i>Explizite Einschränkung:</i> Nicht Anwendbar bei Personen, deren Muttersprache nicht Deutsch ist, und bei Personen mit hirnorganischen Erkrankungen.
Theoretischer Hintergrund	Gf-Gc Theory von R. B. Cattell. Der WST soll den kristallinen Aspekt der Intelligenz mittels einer Wiedererkennungslleistung erfassen.
Operationalisierung/ Aufgabenprinzip	Jede Aufgabe besteht aus einer Zeile von sechs Wörtern. Eines davon ist ein sinnhaftes, reales Wort, das im Rahmen der Aufgabenbearbeitung durchgestrichen werden soll. Die übrigen fünf Wörter sind sinnleere, künstliche, wortähnliche Konstruktionen. Der Gesamtttest besteht aus 42 in der Schwierigkeit steigenden Aufgaben.
Entwicklung (Aufgaben, Skala)	Eine Testvorform mit 79 Items, die dem Aufgabenprinzip des MWT von Lehrl (1974) folgen, wurde 573 Personen zur Bearbeitung vorgegeben. Die Skalenkonstruktion orientierte sich am logistischen Testmodell von Rasch. Die Endform des Tests mit 42 Items resultierte durch die schrittweise Entfernung von nicht-modellkonformen Items.
Durchführung	Einzel- und Gruppentestung, keine Zeitbegrenzung (Bearbeitungsdauer ca. 10 bis 15 Minuten)
Psychometrische Eigenschaften	<i>Objektivität</i> Vollstandardisiertes Verfahren <i>Reliabilität</i> Cronbachs Alpha: $r = ,94$ ($n = 573$) <i>Validität</i> Hochsignifikante positive Korrelationen ($p < ,01$) der Testergebnisse der Probanden der Normstichprobe ($n = 573$) mit dem erreichten Schul- bzw. Berufsbildungsabschluss ($r = ,60$ bzw. $r = ,63$) und Altersunabhängigkeit der Testergebnisse ($r = ,08$). <i>Normierung</i> Gesamtnorm (573 Personen im Alter von 16 bis 90 Jahren; Vergleichswerte: z-, Z- und IQ-Werte)

Verfahren	Verbaler Kurz-Intelligenztest (VKI)
Autor(en), Erscheinungsjahr, -ort, Verlag	H. Anger, F. Mertesdorf, R. Wegener & G. Wülfing, 1980, Weinheim, Beltz
Messgegenstand	Verbale Intelligenz, allgemeine Urteils- und Denkfähigkeit
Anwendungs- bereiche	Erwachsene ab 21 Jahren im Bereich der psychosozialen Beratung und Therapie, Erwachsenenbildung und Umfrageforschung. <i>Explizite Einschränkung:</i> Da Fremdwörter verwendet werden, ist das Testergebnis bildungsschichtabhängig.
Theoretischer Hintergrund	k. A.
Operationalisierung/ Aufgabenprinzip	Jede der Parallelförmigen A und B des Tests enthält 20 Wörter und 4 einfache Zeichnungen mit Darstellungen alltäglicher, leicht verständlicher Situationen. Zu jedem Wort muss dasjenige Bild gefunden werden, das am Besten dazu passt. In jedem Falle gibt es eine eindeutige Optimallösung.
Entwicklung (Aufgaben, Skala)	Verkürzte Version des Wort-Bild-Tests (WBT 10+) von Anger, Mertesdorf, Wegener und Wülfing (1971). Kriterium der Itemselektion war die Höhe der Korrelation der WBT 10+-Items mit dem Gesamtwert einer Kurzversion des Leistungsprüfungs-system (Horn, 1962).
Durchführung	Einzel- und Gruppentestung, keine Zeitbegrenzung (Bearbeitungsdauer ca. 5 Minuten)
Psychometrische Eigenschaften	<i>Objektivität</i> Vollstandardisiertes Verfahren <i>Reliabilität</i> Halbierungsreliabilität, getrennt nach Geschlecht und Testform, zwischen $r = ,83$ und $r = ,87$; durchschnittliche Korrelation der Rohwerte, berechnet für die 4 einzelnen Bilder, von $r = ,82$. <i>Validität</i> Korrelation des Ergebnisses im VKI A bzw. VKI B mit dem Gesamtwert und den Ergebnissen in den Untertests einer Kurzversion des Leistungsprüfungs-system (Horn, 1962). Gesamtwert: $r = ,72$ bzw. $r =$ $,71$; verbal comprehension: $r = ,78$ bzw. $r = ,78$; closure: $r = ,65$ bzw. $r = ,64$; word fluency: $r = ,61$ bzw. $r = ,60$; reasoning: $r = ,54$ bzw. $r =$ $,55$ (alle weiteren Korrelationen $< ,40$). <i>Normierung</i> Gesamtnorm (Eichstichprobe wurde 1979 im Rahmen einer repräsentativen Meinungsumfrage bei 1.090 Einwohnern der Bundesrepublik Deutschland ab 21 Jahren gewonnen; Vergleichswerte: Prozentrangbänder, mittlere T-Werte, T- und Stanine-Wertbänder).

Verfahren	Wortschatztest aktiv und passiv (WST-ap) Deutsche Version des Mill Hill Vocabulary Tests von J.C. Raven
Autor(en), Erscheinungsjahr, -ort, Verlag	N. Ibrahimovic & S. Bulheller, 2005, Frankfurt am Main, Harcourt Test Services
Messgegenstand	Gegenwärtiger Leistungsstand einer Person im Hinblick auf ihren Wortschatzumfang bzw. ihre Fähigkeit zu verbaler Kommunikation.
Anwendungsbereiche	Jugendliche und Erwachsene ab 14 Jahren im Bereich der Berufseignungsdiagnostik und Forschung.
Theoretischer Hintergrund	Die Mill Hill Vocabulary Scale (J. Raven, J. C. Raven & J. H. Court, 1998). Sie wird in historischen Zusammenhang gestellt mit bedeutenden Verfahren, die zur Erfassung des Wortschatzes dienen.
Operationalisierung/ Aufgabenprinzip	Der Aktive Wortschatz wird durch die „Definitionsform“ (freies Format) erfasst: Ein vorgegebenes Wort soll inhaltlich definiert werden. Der passive Wortschatz wird durch die „Wiedererkennungslleistung“ (multiple choice-Format) überprüft: Zu einem vorgegebenen Wort sollen aus 6 dazugehörigen Begriffen alle ausgewählt werden, die mit dem vorgegebenen Wort inhaltlich übereinstimmen.
Entwicklung (Aufgaben, Skala)	Testvorform: Zufallsauswahl von 194 Wörtern der deutschen Sprache aus dem DUDEN - Die deutsche Rechtschreibung. Nach Itemanalyse (n = 205) Reduktion des Itempools auf 60 Items und anschließende Aufteilung in zwei Teile á 30 Items, von denen ein Teil zur Erfassung des aktiven Wortschatzes und der andere zur Erfassung des passiven Wortschatzes dient (Form A; vice versa Form B).
Durchführung	Einzel- und Gruppentestung, Bearbeitungsdauer 23 Minuten
Psychometrische Eigenschaften	<i>Objektivität</i> Vollstandardisiertes Verfahren <i>Reliabilität</i> Innere Konsistenz (Split-Half, n = 508) Form A: r = ,93, Form B: r = ,91 <i>Validität</i> Erwartungskonforme Korrelationen mit Intelligenz-Struktur-Analyse (Fay, Trost & Gittler, 2001), Figure Reasoning Test (Booth & Horn, 2004), Standard Progressive Matrices – Plus (Raven, Styles & Raven, 1998), Rechtschreibtest (Bulheller & Häcker, 2002), Mathematiktest – Grundkenntnisse für Lehre und Beruf (Ibrahimovic & Bulheller, 2005b), Inventar Komplexer Aufmerksamkeit (Heyde, 2000), Frankfurter Aufmerksamkeits-Inventar (Moosbrugger & Oehlschlägel, 1996). Faktorenanalyse mit Intelligenz-Struktur-Analyse - S (Fay, Trost & Gittler, 2001), Mathematiktest und Frankfurter Aufmerksamkeits-Inventar: Der WST-ap lädt zusammen mit den ISA-S Untertests, die die verbale Fähigkeit erfassen, substantiell ($\alpha \geq ,6$) auf dem zweiten von insgesamt drei Faktoren. <i>Normierung</i> Form A (n = 2053): Prozentränge für Jugendliche (14 bis 19 Jahre) getrennt nach Realschule und Gymnasium und für Erwachsene (20 bis 60+ Jahre). Form B (n = 873): Prozentränge für Jugendliche (14 bis 19 Jahre) getrennt nach Realschule und Gymnasium.

Verfahren	Peabody Picture Vocabulary Test (PPVT) Deutschsprachige Fassung des PPVT-III für Jugendliche und Erwachsene von L. M. Dunn & L. M. Dunn
Autor(en), Erscheinungsjahr, -ort, Verlag	S. Bulheller & H. O. Häcker, 2003, Frankfurt am Main, Swets Test Services
Messgegenstand	Wortschatz und Verständnis der gesprochenen Sprache
Anwendungsbereiche	Jugendliche und Erwachsene von 14 bis 60 Jahren bei Fragestellungen zum Verständnis der gesprochenen Sprache, besonders im Bereich der beruflichen Aus- und Weiterbildung.
Theoretischer Hintergrund	k. A.
Operationalisierung/ Aufgabenprinzip	Zu einem Begriff, der vom Testleiter genannt wird, soll die Testperson aus vier einfachen Abbildungen, die auf einer Seite angeordnet sind, diejenige herausuchen, die zu dem Begriff passt.
Entwicklung (Aufgaben, Skala)	Nach Itemanalyse von jeweils 130 auf Deutsch übersetzten Items der Originalformen A und B des PPVT-III (L. M. Dunn & L. M. Dunn, 1997) Zusammenstellung der Endformen A und B mit jeweils 89 Items (Analysestichprobe Form A: n = 1261, Form B: n = 1266).
Durchführung	Einzel- und Gruppentestung, keine Zeitbegrenzung (Bearbeitungsdauer ca. 30 Minuten)
Psychometrische Eigenschaften	<i>Objektivität</i> Vollstandardisiertes Verfahren <i>Reliabilität</i> Cronbachs Alpha für beide Parallelförmungen jeweils $r = ,93$ <i>Validität</i> Erwartungskonforme hoch signifikante ($p \leq ,01$) Unterschiede zwischen den Bildungsgruppen Hauptschulabschluss, Realschulabschluss und Abitur/Fachabitur der deutschen Stichproben. Es werden weiterhin Validierungsuntersuchungen mit der englischsprachigen Originalversion PPVT-III und den Intelligenztests (Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991), Kaufman Adolescent and Adult Intelligence Test (KAIT; Kaufman & Kaufman, 1993) und Kaufman Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990) zusammenfassend zitiert, die einen hohen Zusammenhang mit Intelligenzleistungen, insbesondere mit der verbalen Intelligenz, aufzeigen. <i>Normierung</i> Jeweils für Form A und Form B Gesamtnorm (n = 1261 bzw. n = 1266), bildungsspezifische Normen (Hauptschule, Realschule, Gymnasium) und Altersnormen. (13/14, 15, 16, 17, ≥ 18 Jahre).

Die beschriebenen Verfahren nehmen folgende Geltungsbereiche für sich in Anspruch:

- MWT-B: allgemeines kristallines Intelligenzniveau
- WST: allgemeines kristallines Intelligenzniveau und allgemeines Sprachverständnis
- VKI: verbale Intelligenz und allgemeine Urteils- und Denkfähigkeit
- WST-ap: Wortschatzumfang bzw. die Fähigkeit zu verbaler Kommunikation
- PPVT: Wortschatz und Verständnis der gesprochenen Sprache

Der Wortschatztest ILKO misst die verbale kristalline Intelligenz. Der MWT-B, der WST, der WST-ap und der PPVT haben, zumindest teilweise, einen vergleichbaren expliziten Messanspruch. Die Verfahren arbeiten mit unterschiedlichen Aufgabenstellungen:

Beim MWT-B und WST besteht jede Aufgabe aus einer Reihe von Wörtern. Eines der Wörter ist ein sinnhaftes reales Wort. Die übrigen Wörter sind sinnleere wortähnliche Konstruktionen. Die Testperson soll das reale Wort entdecken und durchstreichen. Die kognitive Leistung besteht also in einem Wiedererkennen eines schriftlich dargebotenen realen Wortes, ohne dessen genaue Bedeutung kennen zu müssen. Beim WST ist daher zu bezweifeln, ob er mit einer derartigen Aufgabenstellung seinem zusätzlichen Anspruch gerecht werden kann, das „allgemeine Sprachverständnis“ zu erfassen: Mit Hilfe einer bloßen Wiedererkennungslleistung von schriftlich dargebotenen Wörtern ist sicherlich nicht gewährleistet, dass die Testpersonen auch den genauen semantischen Gehalt der Wörter kennen.

Beim WST-ap und dem PPVT müssen die Testpersonen den semantischen Gehalt der Wörter kennen, um die Aufgaben lösen zu können. Bei diesen beiden Verfahren ist jedoch die Repräsentativität der Wörter für den allgemeinen deutschen Sprachgebrauch nicht sichergestellt. (Dasselbe Problem existiert auch beim MWT-B und beim WST.)

Bezogen auf die für den deutschen Sprachraum publizierten Verfahren erfüllt das Inventar Lexikalischer Kompetenz das Gütekriterium der Nützlichkeit. Die Gründe liegen in erster Linie in seinem gesicherten Messanspruch (siehe Teil 2, Kapitel 2.1.2 und Kapitel 6.3) und der Repräsentativität seiner Zielwörter für den deutschen Sprachgebrauch (siehe Teil 2, Kapitel 4).

6.6 Unverfälschbarkeit

Bei einem Leistungstest beschreibt die Unverfälschbarkeit, inwieweit es ausgeschlossen ist, dass eine Testperson eine Leistung (richtige Antwort) zeigen kann, ohne eigentlich eine Leistung erbracht zu haben. Die Unverfälschbarkeit beschreibt also die Wahrscheinlichkeit, allein durch Raten oder Zufall zur richtigen Lösung zu kommen. Die Gesamtgröße des deutschen Wortschatzes wird, je nach Quelle und Zählweise (sie betrifft z. B. die Frage, ob zusammengesetzte Wörter mitgezählt werden), auf 300.000 bis 500.000 Wörter geschätzt (Duden: Umfang des deutschen Wortschatzes.

<http://www.duden.bifab.de/index2.html?service/newsletterarchiv/2000/000704.html>, Aufruf am 25.01.2006). Ein durchschnittlicher Sprecher verfügt über einen aktiven Wortschatz von etwa 8.000 bis 10.000 Wörtern (Wikipedia: Wortschatz.

<http://de.wikipedia.org/wiki/Wortschatz>, Aufruf am 28.09.2005). Da die Aufgaben des ILKO ein freies Antwortformat besitzen, läuft die Wahrscheinlichkeit, durch Raten oder Zufall eine richtige Antwort zu notieren, folglich gegen Null. Den kaum realistischen Fall ausgeschlossen, dass eine Testperson absichtlich eine schlechte Leistung zeigt, sind die Testergebnisse des ILKO als unverfälschbar zu bezeichnen.

6.7 Zumutbarkeit

Zumutbarkeit wird vom Testkuratorium (1986, S. 359) definiert als „das Ausmaß, in dem ein Test (absolut und relativ zu dem aus der Anwendung des Verfahrens resultierenden Nutzen) die getestete Person in zeitlicher, psychischer (insbesondere ‚energetisch‘-motivational und emotional) sowie körperlicher Hinsicht beansprucht“.

Mit einer Instruktionsdauer von etwa 1 bis 2 Minuten und einer Testdauer von 9 Minuten ist ILKO ein Verfahren, das die Testperson zeitlich nur gering beansprucht. Der hohe Aufforderungscharakter der Aufgabenstellung, die der Bearbeitung von Quizfragen entspricht, wirkt bei der Testbearbeitung motivationssteigernd und trifft nach allgemeiner Erfahrung bei den Testpersonen auf hohe Akzeptanz. Durch die Spannbreite der Schwierigkeiten der Aufgaben werden leistungsstarke Testpersonen nicht unterfordert. Leistungsschwache Personen haben hingegen gute Chancen, wenigstens einige Aufgaben zu lösen. Man darf folglich davon ausgehen, dass die Bearbeitung des ILKO kaum

Versagensängste auslösen wird. In körperlicher Hinsicht ist für die Testbearbeitung nötig, dass die Probanden etwa 11 Minuten an einem Tisch sitzen und neben den wenigen Angaben zur Person maximal 18 Wörter aufschreiben. Die Bearbeitung des ILKO kann insgesamt als zumutbar gelten.

6.8 Fairness

Wenn Testpersonen aufgrund einer bestimmten Gruppenzugehörigkeit bei der Bearbeitung eines Tests nicht systematisch diskriminiert werden, darf der Test als fair gelten.

Die Aufgaben des ILKO wurden für den Einsatz bei Personen mit Deutsch als Muttersprache und höherer schulischer Bildung, also Hochschulzugangsberechtigung, entwickelt und normiert. Etwaige Unterschiede der Testpersonen beim Verständnis der Aufgabenstellung werden durch das einfache Aufgabendesign und die einfache schriftliche Instruktion praktisch ausgeschlossen. Falls dennoch Verständnisschwierigkeiten bestehen sollten, ist der Testleiter angehalten, die Fragen zu klären. Erst wenn sichergestellt ist, dass eine Testperson die Instruktion richtig verstanden hat, darf sie die Aufgaben bearbeiten. Unterschiede in der Bearbeitungsleistung von Personen unterschiedlichen Geschlechts und Alters werden insofern berücksichtigt, als dass spezielle Normen zur Verfügung gestellt werden.

Sofern das Inventar Lexikalischer Kompetenz, wie vorgeschrieben, bei Personen mit Hochschulzugangsberechtigung, deren Muttersprache Deutsch ist, Anwendung findet und die Testrohwerte mit Hilfe von geschlechts- und altersbezogenen Normen standardisiert werden, darf es als fairer Test gelten.

7 Durchführung des Tests

Das Inventar Lexikalischer Kompetenz ist sowohl in Einzel- als auch in Gruppenuntersuchungen einsetzbar. Wegen seiner kurzen Durchführungsdauer eignet es sich gut für die Verwendung im Rahmen von Testbatterien.

7.1 Instruieren der Testpersonen

Die Testpersonen müssen zuerst auf der Vorderseite des Testbogens einige Angaben zu ihrer Person machen. Im Einzelnen werden abgefragt: Name, Alter (in Jahren) und Geschlecht. Bei Geschlecht sind entsprechende Auswahlmöglichkeiten anzukreuzen. Diese personenbezogenen Angaben werden zur korrekten Zuordnung der Ergebnisse zu den Personen und zur Standardisierung der Testergebnisse benötigt. Zusätzlich werden die Schulbildung (Hauptschulabschluss, Realschulabschluss, Fachhochschulreife, Abitur) und die Muttersprache erhoben. Beide Angaben dienen zur Absicherung, dass eine Testperson tatsächlich zur Gruppe derjenigen Personen gehört, auf die der Einsatz des Tests beschränkt ist.

Anschließend sollen die Testpersonen die schriftliche Instruktion auf der Vorderseite des Testbogens durchlesen. Sie beschreibt die Aufgabenstellung und enthält zwei Beispielaufgaben mit Lösungen. Die Testpersonen erhalten dann die Möglichkeit, dem Testleiter Verständnisfragen zu stellen. Die Fragen sollen vom Testleiter in größtmöglicher Anlehnung an die schriftliche Instruktion beantwortet werden. Wenn sichergestellt ist, dass die Testpersonen die Aufgabenstellung verstanden haben, nennt der Testleiter die Anzahl der im Test zu bearbeitenden Aufgaben, sowie die dafür zur Verfügung stehende Zeit. Mit der Aufforderung zum Wenden des Testbogens und zur Bearbeitung der auf der Rückseite stehenden Aufgaben, läuft die Testzeit. Nach 9 Minuten werden die Testpersonen aufgefordert, die Bearbeitung zu beenden.

7.2 Auswertung der Ergebnisse

Die bearbeiteten Testbögen werden mit Hilfe eines Lösungsblattes ausgewertet, das die korrekten Zielwörter enthält. Man vergleicht die Lösungen der Testperson mit den Zielwörtern auf dem Blatt und vergibt für jede richtig gelöste Aufgabe einen Punkt. Die Summe der erzielten Punkte (Rohwert) wird anschließend mit Hilfe von Normtabellen standardisiert. Orthografische Fehler werden bei der Vergabe der Rohwertpunkte nicht berücksichtigt: Entspricht das notierte Zielwort phonetisch dem Lösungswort, so wird die Antwort als richtig gewertet.

7.3 Interpretation der Ergebnisse

Das Inventar Lexikalischer Kompetenz ist ein Test, der die verbale kristalline Intelligenz erfasst. Seine Zielgruppe sind Personen mit höherer schulischer Bildung, deren Muttersprache Deutsch ist. Die Bedeutung der individuellen Testleistung einer Person (ihr Testrohwert) ergibt sich durch ihre Lokalisierung innerhalb der entsprechenden Normstichprobe. Die Standardisierung sollte, getrennt für weibliche und männliche Personen, altersbezogen erfolgen.

Teil 3: Modellgeleitete Konstruktvalidierung des Inventars Lexikalischer Kompetenz und des Inventars Komplexer Aufmerksamkeit

1 Einleitung

„A test ... should be designed ... with a strong theoretical foundation, and supported by considerable validity evidence in order to measure a particular construct or set of constructs (and only those constructs)“ konstatieren Kamphaus, Winsor, Rowe und Kim (2005, S. 31) mit Bezug auf die noch immer vorherrschende Praxis der unzureichenden theoretischen Fundierung von Leistungstests und der hieraus resultierenden, mehr oder weniger subjektiven Interpretation der Testergebnisse.

Die gängige Methode zur Konstruktvalidierung eines psychodiagnostischen Tests besteht im gleichzeitigen Einsatz von Verfahren, die explizit denselben Messanspruch wie der zu validierende Test besitzen, und von Verfahren, die einen anderen Messanspruch haben. Die auf diese Weise ermittelten Koeffizienten der konvergenten und diskriminanten Validität hängen jedoch unmittelbar von der psychometrischen Qualität der für die Validierung eingesetzten Verfahren ab. Dies gilt insbesondere für die Qualität der Absicherung ihrer eigenen Validität. Mehr noch: Bei der Beurteilung von Ergebnissen, die in Leistungstests erzielt werden, besteht ohne ein Verständnis der Prozesse, die den kognitiven Leistungen zu Grunde liegen, die Gefahr zu irreführenden bzw. falschen Schlüssen zu gelangen. Floyd (2005, S. 203 f.) schreibt hierzu: „... a focus on breath of factors and scores representing them largely obscures an understanding on the particulars of cognitive performance, such as strategy use and the activation of specific mental operations“.

Um die Qualität der Konstruktvalidierung des ILKO und des INKA zu verbessern, wird die „klassische“ Validierungsmethodik im Folgenden erweitert: Zunächst wird ein empirisch abgesichertes Modell gesucht, das es erlaubt, die kognitiven Prozesse, die bei der Bearbeitung der Aufgaben von Leistungstests beteiligt sind, zu identifizieren. Anschließend werden Tests ausgewählt, die vergleichbare oder andere Messansprüche besitzen, wie die beiden zu validierenden Verfahren. Schließlich werden die kognitiven Prozesse, die bei der Bearbeitung der einzelnen Verfahren ablaufen, mit Hilfe des Modells beschrieben und die

Verfahren auf dieser Basis miteinander verglichen. Die identifizierten modellkonformen Gemeinsamkeiten und Unterschiede der Tests - und nicht deren explizite Messansprüche - dienen dann als Grundlage für die Bildung von Hypothesen über Zusammenhänge zwischen den Verfahren. Die Hypothesen werden abschließend einer empirischen Überprüfung unterzogen.

2 Ein Gedächtnismodell zur Beschreibung und Einordnung der Tests

Bei der Beschreibung und Erklärung der höheren kognitiven Informationsverarbeitung gewinnt das Konstrukt ‚Arbeitsgedächtnis‘ seit den 1970er Jahren eine immer zentralere Bedeutung (McGrew, 2005; Richardson, Engle, Hasher, Logie, Stoltzfus, & Zacks, 1996; Stoltzfus, Hasher & Zacks, 1996). Das Arbeitsgedächtnis ist die Exekutive für komplexe kognitive Operationen, wie z. B. das logische Denken, das Problemlösen, die Aufmerksamkeitssteuerung oder die Koordinierung der Speicher- und Verarbeitungssysteme (Della Salla & Logie, 2002). Es lässt sich am besten als das Zusammenwirken vielfältiger kognitiver Funktionen vorstellen, von denen jede einzelne eine spezielle Rolle beim temporären Memorieren bzw. bei der unmittelbaren Verarbeitung von Informationen spielt (Logie, 1996). Dieses Zusammenwirken ist grundlegend für kognitive Prozesse, die als kontrolliert, aufmerksam und neuartig bezeichnet werden können (Waltz, 2003). Es besteht die allgemeine Annahme, dass die Kapazität des Arbeitsgedächtnisses begrenzt ist. Es kann nur eine begrenzte Menge an Informationen speichern und simultan verarbeiten und eine begrenzte Anzahl derartiger Prozesse koordinieren (Süß, 2003).

Andrade (2001, S. 3) bezeichnet das Konstrukt ‚Arbeitsgedächtnis‘ als „cornerstone of cognitive psychology“ und führt aus: „Understanding how we temporarily store and process information is fundamental to understanding almost all other aspects of cognition“. McGrew (2005, S. 164 ff.) schreibt zur Bedeutung des Konstrukts: „Collectively, the MW*-criterion studies suggest that MW may be a significant causal factor working *behind* the scenes when complex cognitive performance is required“ und zitiert Kyllonen (1996): „... the remarkable finding is the consistency with which the working memory capacity factor has proven to be the central factor in cognition abilities ... working memory capacity is more highly related to performance on other cognitive tests, and is more highly related to learning, both short-term and long-term, than is any other cognitive factor“ (*MW steht für memory working bzw.

working memory). Engle (1996) konkretisiert die Bedeutung des Arbeitsgedächtnisses für die Erklärung der allgemeinen kognitiven Leistungsfähigkeit bei der Bewältigung von Alltagsanforderungen. Er listet auf, welche Leistungen bis dato erfolgreich durch die Arbeitsgedächtniskapazität vorhersagt werden konnten: das Lese- und Sprachverständnis, das Erlernen des Buchstabierens, das Ausführen von komplexen Anweisungen, das Vokabellernen, das effektive Notizenmachen, das ‚gute‘ Schreiben, das korrekte Verständnis mehrdeutiger Sprachinhalte und das Erlernen von komplexen funktionalen Sachverhalten.

Zur Messung der Arbeitsgedächtniskapazität werden i. d. R. so genannte Doppelaufgaben (dual tasks; siehe auch Sanders, 2004) verwendet. Doppelaufgaben beinhalten sowohl Komponenten der Informationsverarbeitung als auch der Informationsspeicherung. Die bekannteste und valideste ist die Lesespannenaufgabe: Versuchspersonen werden aufgefordert, nacheinander eine Reihe von nicht zusammenhängenden Sätzen zu lesen, jeweils ihren Inhalt zu beurteilen (Verarbeitungskomponente) und sich das letzte Wort jedes Satzes zu merken (Speicherungskomponente). Das Maß für die Kapazität des Arbeitsgedächtnisses ist die maximale Anzahl von Wörtern, die korrekt in ihrer Auftretensfolge erinnert werden (Andrade, 2001; Richardson, 1996b; Towse & Houston-Price, 2001). Es gibt deutliche Hinweise darauf, dass mit der Lesespannenaufgabe, auf Grund ihrer Struktur, die generelle Kapazität des Arbeitsgedächtnisses gemessen wird und nicht nur diejenige, die aufgabenspezifisch auf Sprachproduktion bezogen ist (Logie, 1996).

John T. E. Richardson (1996b, S. 147 f.) fasst in dem von ihm herausgegebenen Buch „Working Memory and Human Cognition“, zu dem eine ganze Reihe renommierter Kognitionswissenschaftler beigetragen haben, den Stand der Erkenntnisse über das Arbeitsgedächtnis zusammen:

- Es ist ein komplexes System, das für die temporäre Speicherung und Verarbeitung von Informationen verantwortlich ist.
- Es unterscheidet sich strukturell und funktional von den verschiedenen Formen des Langzeitgedächtnisses.
- Seine Inhalte bestehen aus Repräsentationen, die durch Interpretationsprozesse im Langzeitgedächtnis aktuell aktiviert wurden.
- Seine Kapazität wird eingeschränkt durch die begrenzte Höhe der neuronalen Aktivierung, die durch das Langzeitgedächtnis ausgebreitet werden kann, und durch die begrenzten

Aufmerksamkeitsressourcen, die verfügbar sind für die Aktivierung und Erhaltung von aufgabenrelevanten Informationen und die Unterdrückung und Entfernung von aufgabenirrelevanten Informationen.

- Sein Kern ist ein zentraler Prozessor, der in die Ausführung einer breiten Palette von exekutiven Funktionen involviert ist.
- Es umfasst zusätzliche Hilfssysteme, die dem zentralen Prozessor untergeordnet sind und zur temporären Speicherung von Informationen dienen. Diese temporären Speicher unterscheiden sich untereinander strukturell und funktional.
- Betrachtet man die theoretischen Eigenschaften, die ihm zugeordnet werden, so spielt es beim Verständnis der intellektuellen Leistungsfähigkeit eine zentrale Rolle.

3 Das Drei-Komponenten-Modell des Arbeitsgedächtnisses von Baddeley und Hitch

Das bekannteste und bislang einflussreichste Modell des Arbeitsgedächtnisses wurde 1974 von Alan Baddeley und Graham Hitch vorgestellt (Anrade, 2001; Logie, 1996). Seine Entwicklung resultierte aus der Notwendigkeit, einen theoretischen Rahmen zu schaffen für die Bewältigung einer breiten Palette kognitiver Aufgaben, bei denen sowohl temporäres Behalten als auch Prozesse der Informationsverarbeitung eine Rolle spielen (Baddeley, 2001). Das Modell verbindet eine einfache Struktur mit sinnstiftenden Erklärungen für viele experimentelle und alltägliche Anforderungen an unmittelbare kognitive Prozesse. Es hilft zudem bei der Erklärung einer Reihe kognitiver Defizite, die aus Gehirnverletzungen oder Gehirnerkrankungen resultieren. Die Popularität des Modells und seine große Erklärungskraft beruhen darauf, dass die Grundlagen der Modellbildung empirische Erkenntnisse waren. Baddeley (1998, S. 167) schreibt hierzu: „While cognitive psychology has always had its share of ambitious modellers, keen to push the latest computer metaphor to its limits, my own approach to memory theory has been resolutely data driven. I see models as resembling maps, in that they represent what we already know in a way that facilitates further investigation and the subsequent development of better models”.

Baddeley (2000, S. 418) beschreibt das Arbeitsgedächtnis folgendermaßen: “The term working memory is used ... to refer to a limited capacity system allowing the temporary storage and manipulation of information necessary for such complex tasks as comprehension,

learning and reasoning". Er ordnet dem Arbeitsgedächtnis folgende Funktionen im Rahmen der kognitiven Informationsverarbeitung zu (Baddeley, 1998):

- Das Zusammenfügen von Informationen aus den einzelnen Sinneskanälen, so dass Wahrnehmungseinheiten entstehen.
- Die zeitweise Speicherung der Sinneswahrnehmungen im Bewusstsein, um sie weiterverarbeiten zu können.
- Das Zusammenfügen der Bestandteile von Episoden, so dass sie später als Einheiten erinnert werden können.
- Das Reflektieren über Informationen, um sie als zweckdienlich oder unzweckdienlich bewerten zu können.
- Die selektive Zuwendung von Aufmerksamkeit, um die aktuell zweckdienlichsten Informationen aus der unmittelbaren Umwelt und dem Langzeitgedächtnis auswählen zu können.
- Die aktive Aufrechterhaltung dieser Informationen bis eine notwendige Handlungsplanung abgeschlossen ist.
- Die Speicherung von gewonnenen Erfahrungen im Langzeitgedächtnis und ihr späterer Abruf.

Baddeley und Hitch bauten auf dem modalen Gedächtnismodell auf, dessen bekannteste Version von Atkinson und Shiffrin aus dem Jahr 1968 stammt (Andrade, 2001; Floyd, 2005; Vallar, 2002) und entwickelten ihr Modell des Arbeitsgedächtnisses, um das Verständnis für die funktionalen Operationen, die im Kurzzeitgedächtnis ausgeführt werden, zu verbessern.

Das modale Modell von Atkinson und Shiffrin differenziert zwischen einem Kurzzeitgedächtnis und einem Langzeitgedächtnis. Diese beiden Systeme sind seriell organisiert: Die temporäre Speicherung von Informationen im Kurzzeitgedächtnis wird als notwendige Voraussetzung für die Speicherung der Informationen im Langzeitgedächtnis angesehen. Das Kurzzeitgedächtnis ist ein einheitliches System mit einer begrenzten Kapazität und spielt eine zentrale Rolle in der kognitiven Aktivität. Es erhält Informationen aus dem sensorischen Register und dem Langzeitgedächtnis und beinhaltet verschiedene Kontrollprozesse. Die Kontrollprozesse, wie Wiederholen, Kodieren und Strategien zum Abruf von Informationen, sind Operationen, die ausgeführt werden, um mit dem Gedächtnis zu arbeiten und es zu kontrollieren. Die Kontrollprozesse können bewusst eingesetzt werden,

und je mehr Kapazität sie in Anspruch nehmen, desto weniger Kapazität verbleibt für das temporäre Speichern von Informationen, und umgekehrt (Logie, 1995; Logie, 1996; Raaijmakers & Shiffrin, 2002; Richardson, 1996a).

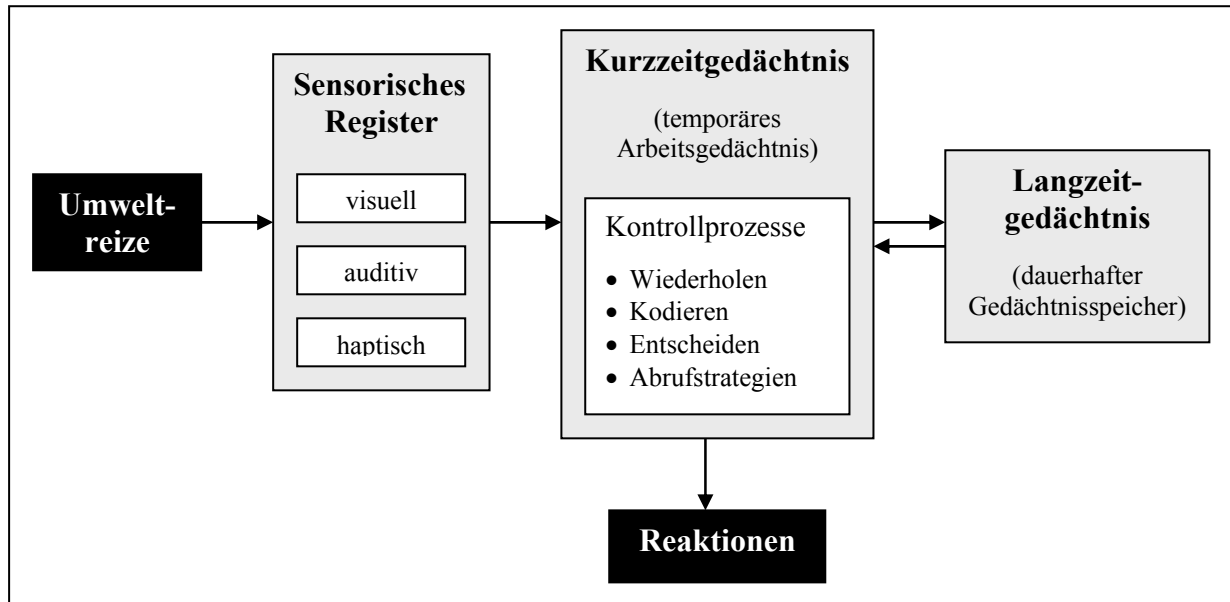


Abb. 3.1: Informationsfluss im modalen Modell des Gedächtnisses von Atkinson und Shiffrin

Die Kernannahme beim Kurzzeitgedächtnis des modalen Modells ist, dass es ein einheitliches System mit begrenzter Kapazität darstellt. Baddeley und Hitch - und andere - konnten jedoch diese Annahme experimentell widerlegen (Andrade, 2001; Baddeley, 1998, Logie, 1995; Logie, 1996; Richardson, 1996a). Sie ermittelten z. B. die maximale Gedächtnisspanne von Versuchspersonen und ließen die Personen anschließend Zahlenfolgen memorieren, die ihre Gedächtnisspanne vollständig ausschöpften. Gemäß dem modalen Modell hätten die Versuchspersonen nun nicht mehr in der Lage sein dürfen, gleichzeitig Aufgaben ohne Probleme zu bewältigen, deren Bearbeitung das Arbeitsgedächtnis in Anspruch nimmt, indem z. B. logisches Denken gefordert wird. Die Ergebnisse zeigten jedoch, dass die Fehlerraten zwischen der Aufgabenbearbeitung ohne gleichzeitige Beanspruchung der Gedächtnisspanne und der Aufgabenbearbeitung mit gleichzeitiger maximaler Beanspruchung der Gedächtnisspanne unverändert blieb. Es verlängerte sich lediglich die Zeit zur Bearbeitung der Aufgaben um etwa 50%. Dieses Ergebnis zeigte deutlich, dass das System des

Arbeitsgedächtnisses, das an der Memorierung von Zahlen beteiligt ist, zwar eine Überlappung mit demjenigen besitzt, das für das logische Denken zuständig ist, dass beide Systeme aber weitgehend unabhängig voneinander arbeiten. Aufgrund dieser und weiterer experimenteller Erkenntnisse ersetzen Baddeley und Hitch das Modell eines einheitlichen Kurzzeitgedächtnisses durch ein strukturiertes Arbeitsgedächtnis, bei dem zwischen Informationsspeicherung und Informationsverarbeitung unterschieden wird und das aus drei Komponenten mit begrenzter Kapazität besteht (Baddeley, 2001).

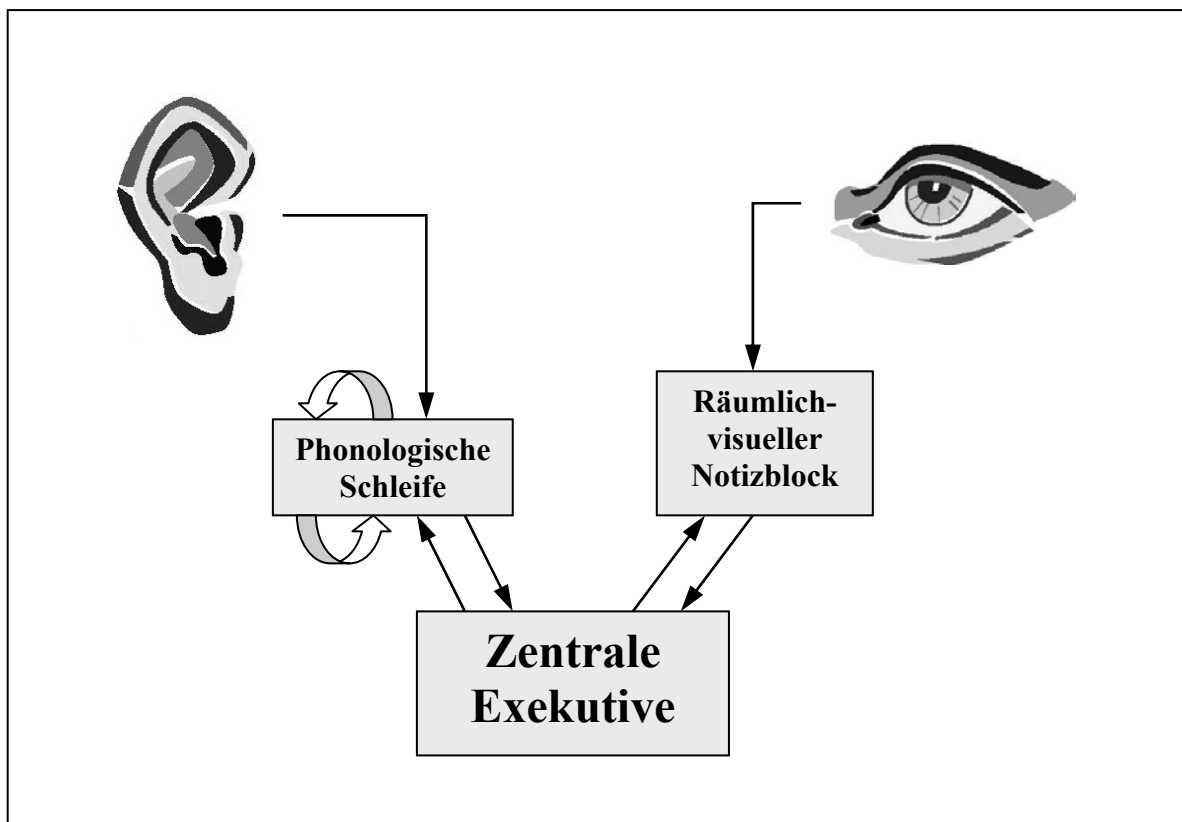


Abb. 3.2: Drei-Komponenten-Modell des Arbeitsgedächtnisses von Baddeley und Hitch

Baddeley (1998, S. 167, 170) beschreibt das Drei-Komponenten-Modell des Arbeitsgedächtnisses wie folgt: „It is suggested that working memory comprises a system for the temporary storage and manipulation of information, forming an important link between perception and controlled action. ... It comprised an attentional control system, which was termed the central executive, aided by two subsidiary slave systems, the phonological loop,

which holds and manipulates acoustic and speech-based information, and the visuo-spatial sketchpad, which performs a similar function for visual information“.

3.1 Die phonologische Schleife

Die phonologische Schleife umfasst einen passiven phonologischen Speicher und einen artikulatorischen Wiederholungsprozess. Verbale Informationen gelangen direkt in den passiven Speicher, in dem sie in einer phonologischen Form für einige Sekunden aufrechterhalten werden. Die gespeicherten Informationen verblassen mit der Zeit und werden auch durch neu eingehendes verbales Material überlagert. Ihr Verlust kann allerdings durch lautlose artikulatorische Wiederholung für prinzipiell unbegrenzte Zeit verhindert werden.

Die Funktion des Speichers wird beeinträchtigt, wenn die verwendeten Sprachmaterialien sich phonologisch ähnlich sind. - Eines der Indizien dafür, dass der Speicher in einem phonologischen Modus arbeitet. - Modellgemäß ist die Existenz der lautlosen artikulatorischen Wiederholung auch der Grund dafür, dass visuell dargebotenes sprachliches Material in einen phonologischen Kode transformiert wird, um es mit Hilfe der phonologischen Schleife temporär zu speichern. Evidenz hierfür ergibt sich aus dem Phänomen, dass der beeinträchtigende Effekt der phonologischen Ähnlichkeit verschwindet, wenn visuell dargebotene Buchstaben oder Wörter nicht in die phonologische Schleife eingespeist werden können. - Dies erreicht man dadurch, indem die Testpersonen während der Memorierungsaufgabe permanent bedeutungslose Wörter laut wiederholen müssen. - Das visuell dargebotene Material wird nun unter Zuhilfenahme eines visuellen oder semantischen Kodes kurzzeitig gespeichert.

Hinweise auf die Kapazitätsbegrenzung der Schleife liefert der so genannte Wortlängeneffekt: Es können mehr kurze Wörter durch den artikulatorischen Wiederholungsprozess aufrechterhalten werden als lange Wörter. Der Grund hierfür liegt in der maximalen Geschwindigkeit, in der Wörter individuell innerhalb von etwa zwei Sekunden lautlos artikuliert werden können (Baddeley, 1998; Baddeley, 2001; Della Sala & Logie, 2002; Logie, 1995; Logie, 1996).

3.2 Der räumlich-visuelle Notizblock

Vom räumlich-visuellen Notizblock wird angenommen, dass er eine der phonologischen Schleife analoge Funktion für visuelles und räumliches Material besitzt (Andrade, 2001; Logie, 1995). Er erlaubt die Repräsentationen von visuell-räumlicher Information für kurze Zeit aufrecht zu erhalten und zu verarbeiten. Er spielt eine wichtige Rolle bei der räumlichen Orientierung und bei der Lösung von visuell-räumlichen Problemen. Vom Notizblock wird angenommen, dass er eine Schnittstelle zwischen visueller und räumlicher Information schafft, mit Zugang durch die Sinne oder durch das Langzeitgedächtnis. In Folge dessen ermöglicht er, dass eine Reihe visueller Informationskanäle mit ähnlichen Informationen motorischer, taktiler oder haptischer Natur zusammengeführt werden können (Baddeley, 2001). Die Art und Weise des Wiederholungsprozesses ist weniger gut untersucht als bei der phonologischen Schleife. Es scheint auch, dass der räumlich-visuelle Notizblock insgesamt weit komplexer ist. Er enthält z. B. visuelle Subsysteme, die zum Aufrechterhalten von Aspekten wie Form und Farbe zuständig sind, und, davon getrennt, räumliche Subsysteme (Baddeley, 1998).

3.3 Die zentrale Exekutive

Die zentrale Exekutive spielt eine entscheidende Rolle im Modell des Arbeitsgedächtnisses: Sie koordiniert und kontrolliert den Einsatz der verschiedenen Hilfssysteme. Sie kann Informationen in die Hilfssysteme einspeisen und aus ihnen abrufen. Sie kann den Kode von Informationen aus einem Hilfssystem in den Kode eines anderen Hilfssystems übersetzen. Sie koordiniert die parallele Ausführung verschiedener Aufgaben. Sie ist bedeutsam für das logische Denken, das Problemlösen und für die Zuteilung von Aufmerksamkeit beim Lernen und beim Erinnern (Della Sala & Logie, 2002; Floyd, 2005; Logie, 1995; Logie, 1996; Logie & Della Sala, 2001).

Baddeley (1998, S. 167) selbst bezeichnet die zentrale Exekutive als „attentional control system“. Er führte mit Kollegen eine ganze Reihe von Untersuchungen durch, um Aufmerksamkeitsprozesse als integrale Bestandteile der Funktionen der zentralen Exekutive zu erforschen. Die Ergebnisse geben klare Hinweise darauf, dass sowohl die Fähigkeit, Aufmerksamkeit zu fokussieren, als auch die Fähigkeit, Aufmerksamkeit zu teilen, wichtige

Funktionen der zentralen Exekutive sind: Die zentrale Exekutive kontrolliert und steuert kognitive Prozesse offensichtlich über die Zuwendung von Aufmerksamkeit.

Den Untersuchungen zur fokussierten Aufmerksamkeit lag die These zugrunde, dass alles, was die Aufmerksamkeitskapazität einschränkt, auch die kognitive Leistungsfähigkeit verschlechtert. Hierzu wurden parallel zur Bearbeitung von Schachaufgaben, von denen angenommen wird, dass sie die zentrale Exekutive stark in Anspruch nehmen, Aufgaben zur Bearbeitung vorgegeben, die zum einen die phonologische Schleife, zum anderen den räumlich-visuellen Notizblock und letztlich die zentrale Exekutive selbst stark beanspruchten. Das Erinnern einer Schachposition und die Auswahl des besten nächsten Zuges wurden gleichermaßen bei Schachexperten und -laien am stärksten durch das gleichzeitige Erzeugen von Zufallszahlen gestört, eine Aufgabe, von der angenommen wird, dass sie die zentrale Exekutive stark beansprucht. Die Untersuchungsergebnisse resümierend schreibt Baddeley (2001, S. 856): „In conclusion, the capacity to focus available attentional capacity is clearly an important feature of the executive“.

Die Funktion der geteilten Aufmerksamkeit wurde mit Hilfe von Alzheimer-Patienten untersucht, die u. a. symptomatisch an Aufmerksamkeitsstörungen leiden. Alzheimer-Patienten und gesunde Kontrollpersonen wurden aufgefordert, Aufgaben zu lösen, die zum einen die phonologische Schleife, zum anderen den räumlich-visuellen Notizblock in Anspruch nahmen. Gegenüber den Leistungen bei der Bearbeitung der Einzelaufgaben, schnitten die Alzheimer-Patienten bei der kombinierten Bearbeitung der Aufgaben deutlich schlechter ab als die gesunden Personen (Baddeley, 2001; Logie, 1996). Dieser Effekt vergrößerte sich mit fortschreitender Erkrankung (Andrade, 2001). Auch hierzu fasst Baddeley (2001, S. 856) zusammen: „These and other studies using both AD* patients and normal participants appear to argue for a separable executive capacity to divide attention“ (*AD – Alzheimer disease).

Baddeley (2001) weist der zentralen Exekutive auch eine Schnittstellenfunktion zwischen den Subsystemen des Arbeitsgedächtnisses und dem Langzeitgedächtnis zu. Z. B. wird die unmittelbare Gedächtnisspanne durch eine semantische Ähnlichkeit der zu memorierenden Wörter, deren Auftretenshäufigkeit oder deren Bedeutungsgehalt positiv beeinflusst; alles Aspekte, die funktional dem Langzeitgedächtnis zugeschrieben werden, nicht aber der phonologischen Schleife, die in einem rein phonologischen Modus arbeitet.

3.4 Weiterentwicklung des Drei-Komponenten-Modells durch Baddeley

Baddeley (2000) erweiterte aufgrund einer Reihe von Phänomenen der kognitiven Informationsverarbeitung, die nicht mit der funktionalen Struktur des Drei-Komponenten-Modells zu vereinbaren waren, das ursprüngliche Modell um das Hilffssystem ‚episodischer Puffer‘. Ergebnisse aus Untersuchungen wiesen auf die Notwendigkeit hin, Informationen aus dem Langzeitgedächtnis und den Arbeitsgedächtnissystemen in einer Weise zu integrieren, so dass sie aktiv aufrecht erhalten und verarbeitet werden können (Baddeley, 2001).

Vom episodischen Puffer wird angenommen, dass er ein kapazitätsbegrenztes Speichersystem darstellt, das in der Lage ist, Informationen zu integrieren, die aus verschiedenen Quellen stammen und in unterschiedlichen Codes repräsentiert sind. Er wird von der zentralen Exekutive kontrolliert, die auf die in ihm enthaltenen Informationen über Bewusstseinsprozesse zugreift, über die Informationen reflektiert und, wo es nötig ist, die Informationen modifiziert und verarbeitet. Die zentrale Exekutive beeinflusst seinen Inhalt, indem sie die Aufmerksamkeit auf bestimmte Informationsquellen richtet, nämlich die Sinne, die anderen Subsysteme des Arbeitsgedächtnisses oder das Langzeitgedächtnis. Der Puffer ist insofern episodisch, als dass er Episoden speichert, wobei weitere Informationen räumlich und zeitlich integriert werden können. Mit seiner Hilfe können also Modelle der Umwelt entworfen werden, deren Weiterverarbeitung z. B. zum Problemlösen oder zur Verhaltensplanung dienen. Die zentrale Exekutive wird in dem erweiterten Modell als reines aufmerksamkeitsbasiertes Kontrollsystem angesehen (Baddeley, 2000; 2001).

3.5 Modifizierung des Drei-Komponenten-Modells durch Logie

Eine für die vorliegende Arbeit bedeutsame Modifizierung des Drei-Komponenten-Modells des Arbeitsgedächtnisses erfolgte durch Robert H. Logie (1995). Sie betrifft vor allem die explizite Einbeziehung des Langzeitgedächtnisses in das Modell.

Logie versuchte folgende Phänomene der menschlichen Informationsverarbeitung in das Drei-Komponenten-Modell zu integrieren: Zum einen gibt es Belege dafür, dass das kurzfristige Behalten von Bewegungssequenzen auf anderen kognitiven Funktionen beruht als das kurzfristige Behalten von Mustern oder Bildern. Zum anderen gibt es klare Belege dafür,

dass externale Reize aus dem sensorischen Register zuerst in das Langzeitgedächtnis gelangen und dort ihre kognitiven Repräsentationen aktivieren. Dem Arbeitsgedächtnis werden dann die aktivierten Repräsentationen verfügbar.

3.5.1 Die Struktur des räumlich-visuellen Notizblocks

Die Kapazität der Memorierung von Bewegungssequenzen wird z. B. mit Hilfe von Block-Tapping-Aufgaben (siehe Schellig, 1997) gemessen. Der Versuchsleiter tippt zufällig auf einem Brett angeordnete Würfel in einer bestimmten Sequenz an. Der Proband muss danach dieselbe Sequenz durch Antippen reproduzieren. Das unmittelbare Erinnern von visuellen Mustern wird häufig mit Hilfe von Schachbrettmustern mit zufällig angeordneten schwarzen und weißen Feldern erfasst, die nach kurzer Vorgabe reproduziert werden sollen (Della Salla & Logie, 2002).

Versuche mit Kindern und Erwachsenen zeigen, dass die Leistungen bei der Bewältigung beider Aufgabentypen über die Zeit unterschiedliche Entwicklungsverläufe haben. Ebenso kann bei unterschiedlichen Formen von Hirnschädigungen jeweils nur eine der beiden Aufgabenbewältigungen beeinträchtigt sein. Auch gibt es experimentelle Belege dafür, dass unterschiedliche kognitive Funktionen dem Behalten von Bewegungssequenzen und dem kurzfristigen Behalten von visuellen Mustern zugrunde liegen. So treten bei Doppelaufgaben folgende Interferenzeffekte auf: Die Block-Tapping-Aufgabe wird z. B. durch das regelmäßige Antippen von vorgegebenen Feldern während der Behaltenszeit beeinträchtigt, nicht jedoch durch das Betrachten von Bildern oder Mustern. Bei den Aufgaben zum Behalten von visuellen Mustern zeigt sich der genau umgekehrte Effekt: Das regelmäßige Antippen von vorgegebenen Feldern während der Behaltenszeit schränkt die Erinnerungsleistung nicht ein; sie wird jedoch durch das Betrachten von Bildern oder Mustern beeinträchtigt (Baddeley, 2001; Della Salla & Logie, 2002).

Um diese und gleichartige Phänomene zu berücksichtigen, teilte Logie (1995) den räumlich-visuellen Notizblock in einen visuellen Speicher (visual cache) und einen visuellen Kontrollprozess (inner scribe) auf.

3.5.2 Die Funktion des Langzeitgedächtnisses bei der Informationsaufnahme

Es existieren klare empirische Belege dafür, dass sensorische Reize ihre kognitiven Repräsentationen im Langzeitgedächtnis automatisch aktivieren. Dem Arbeitsgedächtnis werden diese aktivierten Repräsentationen zur Weiterverarbeitung verfügbar (siehe auch van der Linden, 1998).

Einer der bekanntesten Belege ist der Stroop-Effekt (Stroop, 1935). Werden Personen aufgefordert, die Farbe zu benennen, in denen Wörter geschrieben sind, so verlangsamt sich ihre Reaktionszeit deutlich, wenn die Wörter selbst Bezeichnungen anderer Farben darstellen (z. B. das Wort ‚rot‘ in blauer Farbe geschrieben). Dieser Effekt ist ein Indiz dafür, dass die Wortbedeutung beim Erkennen des Wortes automatisch aktiviert wird. Das bewusste Benennen der Farbe, in der ein Wort geschrieben ist, interferiert dann mit der automatisch aktivierten Wortbedeutung, wenn diese eine andere Farbe darstellt (Anderson, 2001; Bundesen, 2001; Vecera & Luck, 2002; Cohen, Aston-Jones & Gilzenrat, 2004).

Priming-Effekte, d. h. der schnellere Abruf von Inhalten des Langzeitgedächtnisses, wenn diese selbst oder mit ihnen assoziierte kognitive Inhalte zuvor aktiviert worden sind, werden auch dann hervorgerufen, wenn die vorhergehende Aktivierung unbewusst stattgefunden hat. Auch aktivieren visuell dargebotene Wörter automatisch ihre phonologischen Repräsentationen, die dann zum temporären Memorieren in die phonologische Schleife eingespeist werden (Logie, 1996).

Versuche mit Personen mit visuellem Neglect zeigen, dass der Teil eines Bildes, der trotz normaler Sehfähigkeit von ihnen nicht bewusst wahrgenommen werden kann, dennoch einen Einfluss auf ihre Reaktionen ausübt. Della Salla und Logie (2002) beschreiben ein klassisches Experiment, in dem Patienten mit linksseitigem Neglect zwei Bilder mit identischen Häusern gezeigt wurden. Bei einem der beiden Häuser stand die linke Hausseite in Flammen. Wurden die Patienten gefragt, ob es Unterschiede zwischen den beiden Häusern gibt, so verneinten sie dies. Wurden sie allerdings gefragt, in welchem Haus sie lieber wohnen würden, so wählten sie das Haus, das nicht brannte. Obgleich die sensorische Information ‚Feuer‘ scheinbar nicht im Arbeitsgedächtnis verfügbar war, hatte sie dennoch semantisches Wissen über Feuer und seine unangenehmen Auswirkungen aktiviert.

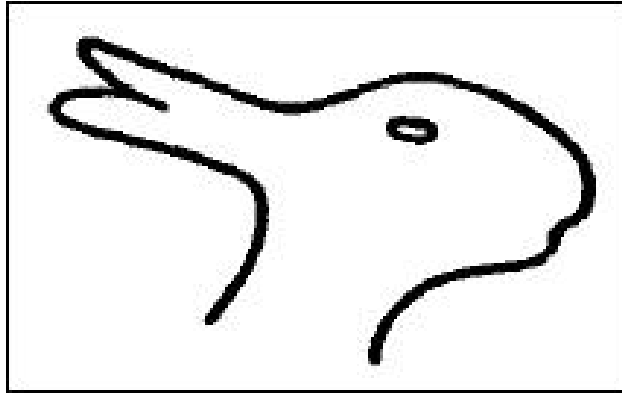


Abb. 3.3: Doppeldeutige Figur aus dem Experiment von Chambers und Reisberg

Chambers und Reisberg (1992) berichten von einem Experiment mit doppeldeutigen Figuren: Die Figur in Abbildung 3.3 kann sowohl als Hase als auch als Ente interpretiert werden. Wenn man sie länger betrachtet springt die Interpretation zwischen Hase und Ente hin und her. In ihrem Experiment zeigten Chambers und Reisberg diese Figur einer großen Anzahl von Versuchspersonen für sehr kurze Zeit. Die Personen wurden anschließend aufgefordert, zu benennen, was die Figur darstellte. Alle benannten sie entweder als Hase oder als Ente. Niemand war jedoch in der Lage, die Figur im Gedächtnis neu zu interpretieren und etwas anderes in ihr zu sehen als das, was die ursprüngliche Interpretation ergeben hatte. Die Neuinterpretation gelang jedoch ohne Probleme, wenn die Versuchspersonen die Figur aus dem Gedächtnis auf Papier zeichneten und ihr gezeichnetes Bild betrachteten. Chambers und Reisberg (1992) berichten weiterhin von Versuchen, die dieses Phänomen der visuellen Vorstellung auch für Prozesse der auditiven Vorstellung belegen.

Logie (1995, S. 3) sieht diese Phänomene als eindeutige Belege dafür, dass die Informationen, die aus den Sinneswahrnehmungen in das Arbeitsgedächtnisses gelangen, dort gespeichert und verarbeitet werden, bereits eine Interpretation im Langzeitgedächtnis durchlaufen haben: „... sensory input accesses long-term memory representations first. These representations are activated, and *ipso facto* the activated representations become available to working memory which holds the information momentarily and implements on-line processing”.

3.5.3 Das modifizierte Modell

Auf Grund der dargestellten Phänomene der kognitiven Informationsverarbeitung, sowohl bezogen auf die Informationsspeicherung im räumlich-visuellen Notizblock als auch bezogen auf die Rolle des Langzeitgedächtnisses bei der Informationsaufnahme, modifizierte Logie (1995, S.127) das Drei-Komponenten-Modell des Arbeitsgedächtnisses von Baddeley und Hitch.

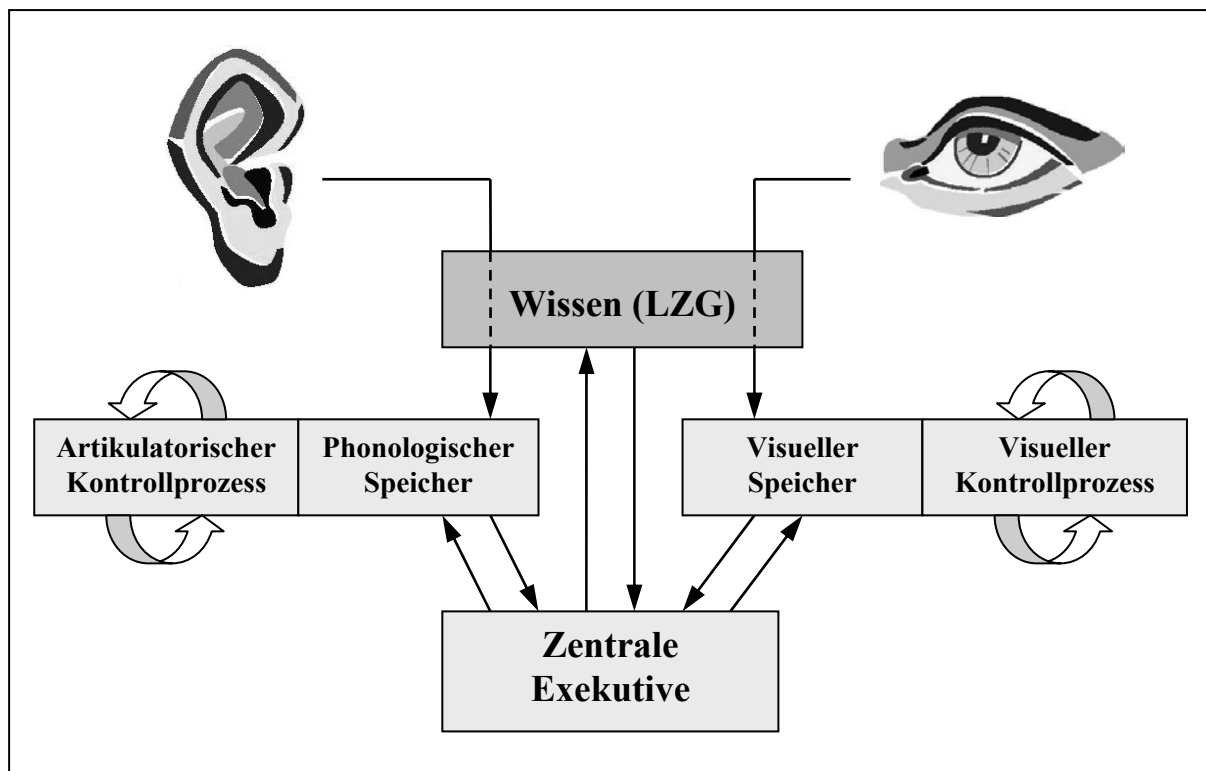


Abb. 3.4: Modifiziertes Drei-Komponenten-Modell des Arbeitsgedächtnisses nach Logie

Das Arbeitsgedächtnis ist modellgemäß ein multifunktionaler Arbeitsbereich des Gehirns, der Informationen, die aus dem sensorischen Register, dem Langzeitgedächtnis und als Resultate der Informationsverarbeitung aus dem Arbeitsgedächtnis selbst stammen, temporär speichert und weiterverarbeitet. Die zentrale Exekutive ist das zentrale Element des Arbeitsgedächtnisses. Sie steuert und kontrolliert die Informationsspeicherung und -verarbeitung über Prozesse der Aufmerksamkeitszuwendung. Die Zuwendung der Aufmerksamkeit auf

einzelne Gedächtnissysteme erfolgt aufgrund der limitierten Kapazität des Arbeitsgedächtnisses, so dass die für die konkrete Anforderung relevantesten Informationen zur Verfügung gestellt und verarbeitet werden können. Die zentrale Exekutive kann die Codes, in denen Informationen repräsentiert werden, in andere Codes transformieren, so dass sie in andere Subsysteme zur Speicherung oder Weiterverarbeitung eingespeist werden können. Die phonologischen und visuell-räumlichen Hilfssysteme dienen der temporären Speicherung von entsprechend kodierten Informationen, die aktuell weiterverarbeitet oder erinnert werden sollen. Sensorische Informationen gelangen aus dem sensorischen Register zuerst in das Langzeitgedächtnis und aktivieren dort ihre kognitiven Repräsentationen. Die aktivierten Repräsentationen können anschließend in das Arbeitsgedächtnis und damit ins Bewusstsein gelangen. Sie sind i. d. R. akustisch enkodiert (Logie, 1996; Zimbardo, 1992).

Das in Abbildung 3.4 dargestellte Modell des Arbeitsgedächtnisses soll als Grundlage für die Identifizierung der kognitiven Prozesse dienen, die bei der Bearbeitung der Tests ablaufen, die zur weiteren Konstruktvalidierung von INKA und ILKO eingesetzt werden.

4 Die zur modellgeleiteten Konstruktvalidierung eingesetzten Tests

Zur modellgeleiteten Konstruktvalidierung des Inventars Komplexer Aufmerksamkeit und des Inventars Lexikalischer Kompetenz wurden Tests ausgewählt, die unterschiedliche Aspekte der intellektuellen Leistungsfähigkeit sprachgebunden und sprachfrei erfassen. Einen besonderen Schwerpunkt bildeten Verfahren, die sprachgebundene Inhalte des Langzeitgedächtnisses überprüfen. Im Einzelnen wurden eingesetzt:

- Figure Reasoning Test - 20 Min. (FRT; Daniels, 1962; Booth & Horn, 2004)
- Verbaler Kurz-Intelligenztest - Form A (VKI; Anger, Mertesdorf, Wegener & Wülfing, 1980)
- Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B; Lehrl, 1995)
- Mannheimer Rechtschreib-Test (MRT; Jäger & Jundt, 1981)
- Intelligenz Struktur Analyse - Beziehungen erschließen (ISA-BE; Fay, Trost & Gittler, 2001)
- Intelligenz-Struktur-Test 70 (IST 70; Amthauer, 1973)

Die Tests wurden zusammen mit ILKO und/oder INKA in folgenden Kombinationen als Testbatterien einer Gesamtstichprobe von 331 deutschen Studenten bzw. Akademikern vorgegeben (siehe Teil 3, Kapitel 6.1):

- Testbatterie 1 (TB 1): ILKO, INKA, FRT
- Testbatterie 2 (TB 2): ILKO, INKA, MWT-B, VKI
- Testbatterie 3 (TB 3): ILKO, INKA, IST-70, MRT
- Testbatterie 4 (TB 4): ILKO, ISA-BE

4.1 Ableitung der für die Bearbeitung der Tests relevanten kognitiven Prozesse

„...information processing is generally defined as the sequence of mental operations and their products involved in performing a cognitive task. ... Thus, an elementary information process can be considered the fundamental mental event in which information is operated on to produce a response. For example, these processes may include the *encoding* of external information into the cognitive system, *comparison* of the new information to information stored in memory, selection of a response (e.g., either ‘same’ or ‘different’), and execution of that response (saying either ‘same’ or ‘different’). ... Models such as the modal model of memory, the working memory model, and the ACT-R theory represent some of the most advanced descriptions of the human information-processing system and its elementary information processes“ (Floyd, 2005, S. 204 f.).

Zur Beschreibung der Prozesse der kognitiven Informationsverarbeitung, die bei der Bearbeitung der ausgewählten Leistungstests ablaufen, soll das von Logie modifizierte Drei-Komponenten-Modell des Arbeitsgedächtnisses dienen (siehe Teil 3, Kapitel 3.5.3). Für jeden der eingesetzten Tests wird zunächst die Aufgabenstellung objektiv dargestellt. Danach erfolgt eine modellgeleitete introspektive Beschreibung der kognitiven Prozesse, die bei der Bearbeitung der jeweiligen Aufgaben stattfinden. Die Ergebnisse werden anschließend in einer Tabelle zusammengefasst. Die Tabelle dient als Grundlage für die Bildung von Hypothesen über mögliche Zusammenhänge zwischen den Tests. Zum Schluss werden die Hypothesen einer empirischen Überprüfung unterzogen.

Den nachfolgenden Beschreibungen der kognitiven Prozesse soll die Feststellung von Logie (1995, S. 131) vorausgeschickt werden: „Subjects, we know, are strategic, and not all subjects will necessarily attempt to perform the same task using the same components of their neuroware“. Dies gilt aus Sicht des Autors insbesondere für komplexe Aufgabenstellungen, wie z. B. beim Figure Reasoning Test.

4.1.1 Inventar Lexikalischer Kompetenz

Die Aufgabenstellung:

Das ILKO ist ein Wortschatztest, bei dem Begriffe durch das Lesen ihrer inhaltlichen Beschreibungen erinnert und anschließend aufgeschrieben werden sollen.

Der kognitive Prozess bei der Aufgabenbearbeitung:

- Lesen der Beschreibung des semantischen Inhalts eines Begriffs
- bei Aktivierung des Begriffs im Langzeitgedächtnis (freies Erinnern): Aufschreiben des Begriffs

4.1.2 Inventar Komplexer Aufmerksamkeit

Die Aufgabenstellung:

Das INKA ist ein Test zur Erfassung der selektiven visuellen Aufmerksamkeit. Bei der Aufgabenbearbeitung müssen zunächst ein bis drei Konsonanten und/oder Konsonantenpärchen mit Hilfe einer Umwandlungstabelle in andere Konsonanten und/oder Konsonantenpärchen transformiert werden. Die transformierten Vorgaben müssen temporär memoriert und anschließend in Buchstabenreihen gesucht werden. Immer wenn eine memorierte Vorgabe in den Buchstabenreihen entdeckt wird, so muss zur Indikation dessen ein bestimmter Zielbuchstabe aufgeschrieben werden.

Der kognitive Prozess bei der Aufgabenbearbeitung:

- Umwandlung von vorgegebenen Konsonanten und/oder Konsonantenpärchen
- temporäre Speicherung der umgewandelten Vorgaben (in der phonologischen Schleife)
- Ausrichten von selektiver visueller Aufmerksamkeit auf die Buchstabenreihe

- kontrollierte Vergleiche der gespeicherten Vorgaben mit den visuellen Stimuli
 - bei Übereinstimmung: Identifizierung und Aufschreiben des Zielbuchstabens

 - Ausrichten von selektiver visueller Aufmerksamkeit auf die Buchstabenreihe
 - ...
- (Fortführung der Prozesse bis zum Erreichen des Endes der Buchstabenreihe)

4.1.3 Figure Reasoning Test

Die Aufgabenstellung:

Der FRT ist ein sprachfreier Test des logischen Denkens. Seine 45 Aufgaben bestehen aus 3x3-Matrizen, die aus einfachen geometrischen Figuren zusammengesetzt sind. In den einzelnen Matrizen verändern sich die Figuren pro Zeile regelmäßig in ein bis drei Parametern (Form, Anzahl, Muster). In der untersten Zeile ist jeweils die letzte Figur ausgespart. Zu jeder Matrize werden zusätzlich sechs Figuren vorgegeben, von denen diejenige ausgewählt werden soll, die die letzte Zeile der 3x3-Matrize logisch richtig, d. h. regelkonform, ergänzt.

Der kognitive Prozess bei der Aufgabenbearbeitung:

- Ausrichten von visueller Aufmerksamkeit auf die Figuren der ersten Zeile
- temporäre Speicherung von (Beschreibungen der) Figuren oder Aspekten der Figuren (in der phonologischen Schleife)
- kontrollierte Herleitung der Veränderungsregel(n) durch multiple Vergleiche
- Ausrichten von visueller Aufmerksamkeit auf die Figuren der zweiten Zeile
- kontrollierte Anwendung der hergeleiteten Veränderungsregel(n), um ihre Gültigkeit zu prüfen
(falls ungültig: Herleitung neuer Regel(n) aus Zeile eins oder zwei bis die Gültigkeitsprüfung mit Hilfe der jeweils anderen Zeile positiv ausfällt)
- Ausrichten von visueller Aufmerksamkeit auf die Figuren der dritten Zeile
- kontrollierte Anwendung der hergeleiteten Veränderungsregel(n) auf die Figuren der dritten Zeile
- Auswahl der regelkonformen Figur aus den sechs Lösungsvorgaben und Markierung

4.1.4 Verbaler Kurz-Intelligenztest

Die Aufgabenstellung:

Der VKI ist ein Wortschatztest, der aus vier gezeichneten Alltagssituationen (z. B. Menschen an einer Rezeption, Straßenszene mit Litfasssäule) und 20 Wörtern (z. B. Administration, Graphik) besteht. Jedes Wort soll derjenigen Zeichnung zugeordnet werden, zu der es semantisch am besten passt.

Der kognitive Prozess bei der Aufgabenbearbeitung:

- temporäre Speicherung eines gelesenen Wortes (in der phonologischen Schleife)
- Ausrichten von selektiver visueller Aufmerksamkeit auf die Zeichnungen
- Suche eines Zeichnungselements, das den gleichen semantischen Inhalt wie das Wort besitzt
- bei Aktivierung des gleichen semantischen Inhalts (Wiedererkennen): Markierung

4.1.5 Mehrfachwahl-Wortschatz-Intelligenztest

Die Aufgabenstellung:

Der MWT-B ist ein Wortschatztest, bei dem jede Aufgabe aus einer Reihe von vier sinnlosen Wörtern und einem umgangs-, bildungs- oder wissenschaftssprachlich bekannten Wort besteht. Das bekannte Wort soll entdeckt und durchgestrichen werden.

Der kognitive Prozess bei der Aufgabenbearbeitung:

- sukzessives Lesen der vorgegebenen Wörter
- bei Aktivierung der kognitiven Repräsentation eines Wortes (Wiedererkennen):
Durchstreichen des Wortes

4.1.6 Mannheimer Rechtschreib-Test

Die Aufgabenstellung:

Der MRT ist ein Rechtschreibtest. Aus jeweils vier vorgegebenen Schreibweisen eines Wortes oder einer kurzen Kombination von Wörtern (z. B. ‚im Dunkel der Nacht‘) soll diejenige Schreibweise ausgewählt werden, die orthografisch korrekt ist.

Der kognitive Prozess bei der Aufgabenbearbeitung:

- sukzessives Lesen der vorgegebenen Worte/Kombinationen von Wörtern
- bei Aktivierung der kognitiven Repräsentation der Orthografie (Wiedererkennen):
Markieren des wiedererkannten Wortes/der wiedererkannten Kombination von Wörtern

4.1.7 Intelligenz-Struktur-Analyse: Beziehungen erschließen

Die Aufgabenstellung:

Der ISA-Untertest ‚Beziehungen erschließen‘ besteht aus sprachlichen Analogieaufgaben: Es werden drei Wörter vorgegeben. Zwischen den ersten beiden besteht eine semantische Relation. Zwischen dem dritten Wort und einem von fünf vorgegebenen Wahlwörtern besteht eine ähnliche Beziehung. Dieses Wahlwort soll gefunden und markiert werden.

Der kognitive Prozess bei der Aufgabenbearbeitung:

- Lesen der ersten beiden Wörter und Erkennen der semantischen Relation
- Lesen des dritten Wortes und sukzessives Lesen der fünf Wahlwörter
- bei Aktivierung einer kongruenten semantischen Relation: Markieren des entsprechenden Wahlwortes

4.1.8 Intelligenz-Struktur-Test 70

Der IST 70 besteht aus insgesamt neun Untertests. Aus Gründen der Übersichtlichkeit wird an dieser Stelle von einer Beschreibung der Aufgabenstellungen und der kognitiven Prozesse bei den Aufgabenbearbeitungen abgesehen. Am Ende des Kapitels werden die signifikanten Zusammenhänge der IST 70-Untertests mit INKA und ILKO inhaltlich interpretiert.

4.2 Zusammenfassung und Vergleich der relevanten kognitiven Prozesse

Aufgrund der vorliegenden Untersuchungsstichproben (Studenten bzw. Absolventen deutscher Hochschulen, deren Muttersprache Deutsch ist) darf davon ausgegangen werden,

dass die zur Bearbeitung der Tests erforderlichen Kulturtechniken (Lesen, Schreiben und elementares Rechnen) von den Testpersonen beherrscht werden.

In Anlehnung an das durch Logie modifizierte Drei-Komponenten-Modell des Arbeitsgedächtnisses (siehe Abbildung 3.4) ergeben sich folgende Gemeinsamkeiten und Unterschiede in der kognitiven Informationsverarbeitung bei der Bearbeitung der Aufgaben der einzelnen Verfahren:

Gemeinsamkeiten der Informationsverarbeitung

- Informationsaufnahme

Alle rezipierten Stimuli sind visueller Natur (Buchstaben, Wörter, Zeichnungen). Sie aktivieren automatisch ihre kognitiven Repräsentationen, die dann dem Arbeitsgedächtnis verfügbar werden.

- Reaktionen

Alle Reaktionen sind kontrolliert und handmotorischer Art: Kreuze bzw. Striche machen, Buchstaben bzw. Wörter schreiben.

Unterschiede der Informationsverarbeitung

- Langzeitgedächtnis

Die Lösung der Aufgaben des ILKO, MWT-B, VKI, MRT und ISA-BE beruht auf der Existenz von spezifischem semantischem Wissen im Langzeitgedächtnis und seiner Aktivierung durch die Inhalte der Aufgaben. Für die Aufgabenlösung beim INKA und FRT spielt die Existenz von spezifischem semantischem Wissen keine Rolle.

- Temporäre Speicherung von Informationen

Bei INKA, FRT, VKI und ISA-BE müssen zur Aufgabenbearbeitung vorgegebene Informationen temporär in den Kurzzeitspeichern aufrechterhalten werden. Bei ILKO, MWT-B und MRT ist dies nicht notwendig.

- Prozess der Informationsverarbeitung

Zur Lösung der Aufgaben des INKA und des FRT müssen Vergleiche zwischen mehreren externalen Stimuli simultan und kontrolliert ausgeführt werden. Die Lösung der Aufgaben des ILKO, MWT-B, VKI, MRT und ISA-BE beruht auf der automatisierten Aktivierung von Inhalten des Langzeitgedächtnisses.

- Aufmerksamkeitsleistung

Beim INKA und FRT ist zur Koordinierung und Ausführung der simultanen Prozesse der kontrollierten Informationsspeicherung und -verarbeitung eine hohe Aufmerksamkeitsleistung notwendig. Die erforderliche Aufmerksamkeitsleistung ist bei der weitgehend automatisierten Informationsverarbeitung bei ILKO, MWT-B, VKI, MRT und ISA-BE allenfalls gering.

Zur besseren Übersicht werden die Unterschiede der kognitiven Informationsverarbeitung in Tabelle 3.1 zusammenfassend dargestellt:

	Aktivierung von spezifischen semantischen LZG-Inhalten	Speicherung von Informationen im KZG	Prozess der Informationsverarbeitung		Aufmerksam- keitsleistung
INKA	nicht notwendig	notwendig	komplex	kontrolliert	hoch
FRT	nicht notwendig	notwendig	komplex	kontrolliert	hoch
VKI	notwendig	notwendig	einfach	automatisiert	gering
ILKO	notwendig	nicht notwendig	einfach	automatisiert	gering
MWT-B	notwendig	nicht notwendig	einfach	automatisiert	gering
MRT	notwendig	nicht notwendig	einfach	automatisiert	gering
ISA-BE	notwendig	notwendig	einfach	automatisiert	gering

Tab. 3.1: Unterschiede der kognitiven Informationsverarbeitung bei der Bearbeitung der Aufgaben der Tests

5 Hypothesen

Zur Bildung der Forschungshypothesen über statistische Zusammenhänge zwischen den Leistungen in den einzelnen Tests dient die Darstellung derjenigen Aspekte der kognitiven Informationsverarbeitung, die bei der Bearbeitung der Testaufgaben Unterschiede aufweisen (Tabelle 3.1).

Bei einigen Tests ist für die Aufgabenbearbeitung eine temporäre Speicherung von Informationen erforderlich (INKA, FRT, VKI, ISA-BE). Die temporären Speicher des Arbeitsgedächtnisses stellen Hilfssysteme dar, die mit eigener Kapazität ausgestattet sind und von den anderen Gedächtniskomponenten weitgehend unabhängig arbeiten (siehe Teil 3, Kapitel 3). Sofern die Kapazität der Kurzzeitspeicher nicht überschritten wird, sollte daher die temporäre Informationsspeicherung modellgemäß keinen systematischen Einfluss auf die Bearbeitungsleistungen ausüben. Wegen der geringen Informationsmengen, die bei der Bearbeitung der vorliegenden Tests temporär gespeichert werden müssen, darf man davon ausgehen, dass die Kapazität der Kurzzeitspeicher nicht überschritten wird. Die temporäre Informationsspeicherung wird daher bei der Hypothesenbildung nicht berücksichtigt.

Die leistungsrelevanten Unterschiede bei der Informationsverarbeitung sind darin begründet, ob zur Lösung der Aufgaben:

- spezifisches semantisches Wissen aktiviert werden muss, und
- die Verarbeitungsprozesse einfacher oder komplexer Natur sind, automatisiert oder kontrolliert ablaufen, und dem entsprechend die erforderliche Aufmerksamkeitsleistung hoch oder niedrig ist.

Hypothesen

Positiv gerichtete signifikante Zusammenhänge werden zum einen zwischen den Testleistungen im INKA und im FRT erwartet, zum anderen zwischen den Testleistungen im ILKO und denen im MWT-B, VKI, MRT und ISA-BE.

Keine signifikanten Zusammenhänge werden zum einen zwischen den Testleistungen im ILKO und denen im INKA und FRT erwartet, zum anderen zwischen den Testleistungen im INKA und denen im MWT-B, VKI und MRT.

Tabelle 3.2 gibt die erwarteten Zusammenhänge zwischen den Leistungen in den einzelnen Tests zusammenfassend wieder.

	ILKO	INKA	VKI
INKA	nicht sig.		
FRT	nicht sig.	positiv sig.	
MWT-B	positiv sig.	nicht sig.	positiv sig.
VKI	positiv sig.	nicht sig.	
MRT	positiv sig.	nicht sig.	
ISA-BE	positiv sig.		

Tab. 3.2: Hypothesen über die Zusammenhänge zwischen den eingesetzten Tests

6 Empirische Überprüfung der Hypothesen

Wie bereits in Teil 3, Kapitel 4 dargestellt, wurden die Tests in folgenden Kombinationen als Testbatterien zur Bearbeitung vorgegeben:

TB 1: ILKO, INKA, FRT

TB 3: ILKO, INKA, IST-70, MRT

TB 2: ILKO, INKA, MWT-B, VKI

TB 4: ILKO, ISA-BE

6.1 Untersuchungsstichproben

Die vier Testbatterien wurden im Zeitraum Oktober 2004 bis April 2006 an einer Gesamtstichprobe von 228 Studenten und 103 Absolventen deutscher Hochschulen, deren Muttersprache Deutsch war, eingesetzt. Die Stichprobe baute sich soziodemografisch wie folgt auf (Tabelle 3.3):

	männlich	weiblich	gesamt
Geschlecht	166 (50,2%)	165 (49,8%)	331
Alter (in Jahren)	AM = 26,3 s = 4,2 Min.: 19 Max.: 47	AM = 24,4 s = 4,1 Min.: 19 Max.: 53	AM = 25,3 s = 4,2 Min.: 19 Max.: 53

Tab. 3.3: Soziodemografische Daten der Gesamtstichprobe zur modellgeleiteten Konstruktvalidierung

Die Datenerhebungen fanden mit den vier Testbatterien mit folgenden Testpersonen an folgenden Orten und in folgenden Zeiträumen statt (Tabelle 3.4):

Tests	n	Studienfachrichtungen	Testorte	Zeiträume
TB 1	99	91 Studenten unterschiedlicher Fachrichtungen, Schwerpunkt Psychologie, und 8 wissenschaftliche Mitarbeiter	eine rheinland-pfälzische und eine sachsen-anhaltinische Universität	Mai 2005 - Dez. 2005
TB 2	80	42 Psychologiestudenten und 38 Studenten der Verwaltungswissenschaften	eine rheinland-pfälzische Universität und eine baden-württembergische Verwaltungsfachhochschule	Nov. 2005 - Apr. 2006
TB 3	73	Fachhochschulabsolventen technischer Studiengänge als Bewerber für den gehobenen öffentlichen Dienst	Berufsfeuerwehr einer hessischen Großstadt	Okt. 2004 - Okt. 2005
TB 4	79	57 Studenten unterschiedlicher Fachrichtungen, 22 Bewerber für einen Aufbaustudiengang	drei hessische Universitäten und eine baden-württembergische Verwaltungsfachhochschule	Jun. 2004 - Jan. 2005

Tab. 3.4: Testbatterien, -personen, -orte und -zeiträume

6.2 Ergebnisse

Vor dem Hintergrund dessen, was Horn und Blankson (2005, S. 61) mit den Worten beschreiben: „It is known that almost any task that can be made up to measure a cognitive ability correlates positively with tests of almost every other cognitive ability“, zeigt die Korrelationsmatrix in Tabelle 3.5 ein bemerkenswert deutliches Resultat: Alle Forschungshypothesen können bestätigt werden.

	ILKO	INKA	VKI
INKA	,0974 (n = 237)		
FRT	,0797 (n = 99)	,4581** (n = 92)	
MWT-B	,4517** (n = 80)	,0241 (n = 72)	,3838** (n = 80)
VKI	,4836** (n = 80)	,1566 (n = 72)	
MRT	,3576** (n = 73)	,1589 (n = 73)	
ISA-BE	,4299** (n = 79)		

Pearsonsche Korrelationen, 1-seitig (*p < ,05; **p < ,01)

Tab. 3.5: Ergebnis der empirischen Überprüfung der Zusammenhangshypothesen

Im Rahmen der Testbatterie 3 wurde der IST-70 gemeinsam mit ILKO und INKA eingesetzt. Tabelle 3.6 gibt die korrelativen Zusammenhänge zwischen den Ergebnissen in den neun Untertests des IST-70 und den Testergebnissen im ILKO und INKA wieder:

	ILKO	INKA		ILKO	INKA
IST-SE	,2805** (n = 73)	,0759 (n = 73)	IST-RA	,1618 (n = 73)	,2368* (n = 73)
IST-WA	,1030 (n = 73)	,0070 (n = 73)	IST-ZR	-,0366 (n = 73)	,0211 (n = 73)
IST-AN	,3672** (n = 73)	,1397 (n = 73)	IST-FA	,0759 (n = 73)	,0231 (n = 73)
IST-GE	,2162* (n = 73)	,0180 (n = 73)	IST-WÜ	-,1132 (n = 73)	-,0895 (n = 73)
IST-ME	,1411 (n = 73)	,1297 (n = 73)			

Pearsonsche Korrelationen, 1-seitig (*p < ,05; **p < ,01)

Tab. 3.6: Korrelative Zusammenhänge zwischen ILKO, INKA und den Untertests des IST-70

Die Interpretation der Ergebnisse soll sich auf eine Bewertung der Plausibilität der signifikant positiven Zusammenhänge beschränken.

ILKO

Signifikant positive Zusammenhänge zeigen sich zwischen dem Testergebnis im ILKO und den Testergebnissen in den verbalen Untertests des IST-70, Analogien, Satzergänzung und Gemeinsamkeiten.

Der IST 70-Untertest *Analogien* (IST-AN) besteht aus sprachlichen Analogieaufgaben: Es werden drei Wörter vorgegeben. Zwischen den ersten beiden besteht eine semantische Beziehung. Zwischen dem dritten Wort und einem von fünf vorgegebenen Wahlwörtern besteht eine ähnliche Relation. Dieses Wahlwort soll gefunden werden.

Beim IST 70-Untertest *Satzergänzung* (IST-SE) sollen Sätze, bei denen ein Wort fehlt, mit Hilfe eines von jeweils fünf vorgegebenen Wahlwörtern richtig vervollständigt werden.

Beim IST 70-Untertest *Gemeinsamkeiten* (IST-GE) sollen aus jeweils sechs vorgegebenen Wörtern die beiden Wörter herausgefunden werden, für die es einen gemeinsamen Oberbegriff gibt.

Die für die Aufgabenbearbeitung der drei IST 70-Untertests maßgebenden kognitiven Prozesse sind automatische Aktivierungen von im Langzeitgedächtnis gespeicherten semantischen Inhalten der dargebotenen Wörter. Die Reaktionen (Antworten) bestimmen sich durch die semantische Nähe zwischen verschiedenen Wörtern.

INKA

Ein signifikant positiver Zusammenhang besteht zwischen dem Testergebnis im ILKO und dem Testergebnis im numerischen Untertest *Rechenaufgaben* des IST-70.

Der Untertest *Rechenaufgaben* des IST 70 (IST-RA) besteht aus Textaufgaben, zu deren Lösung verschiedenen Zwischenberechnungen durchgeführt werden müssen, um aus deren Ergebnissen das Endergebnis berechnen zu können. Bei der Bearbeitung dieser Aufgaben sind verschiedene Prozesse der Informationsverarbeitung und -speicherung kontrolliert auszuführen und zu koordinieren.

Die modellgeleitete Konstruktvalidierung bestätigt, dass ILKO ein Test ist, der die verbale kristalline Intelligenz misst. INKA zeigt sich als ein Verfahren, das, bedingt durch die Struktur seiner Aufgaben, neben selektiven visuellen Aufmerksamkeitsleistungen auch die Basisfunktionen des Arbeitsgedächtnisses erfasst.

Fazit

Mit dem Inventar Komplexer Aufmerksamkeit und dem Inventar Lexikalischer Kompetenz wurden zwei standardisierte psychodiagnostische Verfahren entwickelt, die aus der kognitionspsychologischen Perspektive gesehen die Leistungsfähigkeit zweier unterschiedlicher Komponenten des menschlichen Gedächtnisses, Arbeitsgedächtnis und Langzeitgedächtnis, messen.

Inventar Komplexer Aufmerksamkeit

Die ursprüngliche „klassische“ Validierung des INKA ergab deutliche Belege dafür, dass der Test selektive visuelle Aufmerksamkeitsleistung erfasst. Hierzu zählen vor allem die signifikant positiven Zusammenhänge mit dem Aufmerksamkeits-Belastungs-Test d2 (Test d2; Brickenkamp, 1987) und dem Frankfurter Aufmerksamkeits-Inventar (FAIR; Moosbrugger & Oehlschlägel, 1996). Es zeigten sich aber auch signifikant positive Zusammenhänge mit Maßen der fluiden Intelligenz, wie den Standard Progressive Matrices (SPM; Raven, 1996; Raven, Raven & Court, 1999) oder den Advanced Progressive Matrices (APM; Raven, 1995; Raven, Raven & Court, 1998b). Auch ergab eine Studie von Schmidt-Atzert, Bühner und Enders (2006) zum Messgegenstand der gebräuchlichsten Papier-Bleistift-Aufmerksamkeitstests ein vergleichbares Ergebnis. Schmidt-Atzert et al. konnten Zusammenhänge der Testleistung im INKA sowohl mit Maßen der Aufmerksamkeit als auch mit Maßen der Intelligenz ermitteln. Die konvergente Validität des INKA konnte also bestätigt werden, die diskriminante Validität hingegen nicht.

Eine Aufklärung dieses Phänomens ergab sich durch eine zusätzliche modellgeleitete Konstruktvalidierung. Im Rahmen dieser Untersuchung wurden die kognitiven Prozesse, die bei der Aufgabenbearbeitung ablaufen, beleuchtet. Als Modell der menschlichen Informationsverarbeitung diente hierfür eine von Logie (1995) vorgestellte Variante des Drei-Komponenten-Modells des Arbeitsgedächtnisses von Baddeley und Hitch.

Im Gegensatz zu den gängigen Aufmerksamkeitstests, bei denen einfache Aufgaben möglichst schnell bearbeitet werden sollen, müssen beim INKA komplexe Aufgaben, deren

Schwierigkeit ansteigt, möglichst fehlerfrei bearbeitet werden. Die multiplen Prozesse der Informationsspeicherung und -verarbeitung, die für die Bearbeitung der Aufgaben des INKA erforderlich sind, werden über Aufmerksamkeitsleistungen koordiniert und zeigen große Ähnlichkeit mit Prozessen, die bei der Bearbeitung von Aufgaben zur Messung der fluiden Intelligenz ablaufen. Dem entsprechend zeigten sich keine Zusammenhänge des INKA mit Maßen der kristallinen Intelligenz (ILKO: Inventar Lexikalischer Kompetenz; MWT-B: Mehrfachwahl-Wortschatz-Intelligenztest, Lehrl, 1995; VKI: Verbaler Kurz-Intelligenztest - Form A, Anger, Mertesdorf, Wegener & Wülfing, 1980; MRT: Mannheimer Rechtschreib-Test, Jäger & Jundt, 1981).

Das Inventar Komplexer Aufmerksamkeit ist ein Test der selektiven visuellen Aufmerksamkeit, dessen Bearbeitung, bedingt durch die Struktur seiner Aufgaben, eine hohe Arbeitsgedächtnisleistung erfordert.

Inventar Lexikalischer Kompetenz

Der Wortschatztest ILKO ist ein Maß der verbalen kristallinen Intelligenz. Seine modellgeleitete Konstruktvalidierung, die auch auf der von Logie (1995) vorgestellten Variante des Drei-Komponenten-Modells des Arbeitsgedächtnisses basierte, bestätigte die postulierten kognitiven Prozesse und zeigte signifikant positive Zusammenhänge mit andern Maßen der kristallinen Intelligenz (MWT-B: Mehrfachwahl-Wortschatz-Intelligenztest, Lehrl, 1995; VKI: Verbaler Kurz-Intelligenztest - Form A, Anger, Mertesdorf, Wegener & Wülfing, 1980; MRT: Mannheimer Rechtschreib-Test, Jäger & Jundt, 1981). Keine Zusammenhänge ergaben sich mit einem Maß der fluiden Intelligenz (FRT: Figure Reasoning Test - 20 Min., Daniels, 1962; Booth & Horn, 2004) und mit dem Inventar Komplexer Aufmerksamkeit (INKA; Heyde, 1995).

Der Messgegenstand von Wortschatztests wird oft erweitert um Aspekte, wie „allgemeines Sprachverständnis“ (WST: Wortschatztest, Schmidt & Metzler, 1992), „Fähigkeit zu verbaler Kommunikation“ (WST-ap: Wortschatztest - aktiv und passiv, Ibrahimovic & Bulheller, 2005a) oder „Verständnis der gesprochenen Sprache“ (PPVT: Peabody Picture Vocabulary Test, Bulheller & Häcker, 2003), ohne dass hierfür empirische Belege geliefert werden. Aufgrund der aufgezeigten Vielschichtigkeit von sprachkompetentem Verhalten, musste im

Rahmen der vorliegenden Arbeit von der Überprüfung der durchaus plausiblen Annahme abgesehen werden, dass ILKO auch ein valider Indikator der Sprachkompetenz sein könnte.

Literatur

- Alfonso, V. C., Flanagan, D. P. & Radwan, S. (2005). The Impact of the Cattell-Horn-Carroll Theory on Test Development and Interpretation of Cognitive and Academic Abilities. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary Intellectual Assessment, Second Edition: Theories, Tests, and Issues*. (S. 185-202). New York: Guilford.
- Amelang, M. & Bartussek, D. (1985). *Differentielle Psychologie und Persönlichkeitsforschung*. 2., erw. Aufl. Stuttgart: Kohlhammer.
- Amthauer, R. (1973). *Intelligenz-Struktur-Test 70 (IST 70)*. 4. Aufl. Göttingen: Hogrefe.
- Anderson, J. R. (2001). *Kognitive Psychologie*. 3. Aufl. Heidelberg: Spektrum, Akademischer Verlag.
- Andrade, J. (2001). An introduction to working memory. In J. Andrade (Hrsg.), *Working memory in perspective*. (S. 3-30). Hove, England: Psychology Press.
- Anger, H., Mertesdorf, F., Wegener, R. & Wülfing, G., (1971). *Wort-Bild-Test (WBT 10+)*. Weinheim: Beltz.
- Anger, H., Mertesdorf, F., Wegener, R. & Wülfing, G. (1980). *Verbaler Kurz-Intelligenztest (VKI)*. Weinheim: Beltz.
- Atkinson, L. (2001). Intellectual functioning, Assessment of. In N. J. Smelser & P. B. Baltes (Hrsg.), *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 11, S. 7600-7605). Amsterdam: Elsevier.
- Baddeley, A. D. (1998). Working memory. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie*, 321, 167-173.
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423.
- Baddeley, A. D. (2001). Is working memory still working? *American Psychologist*, 56, 851-864.
- Bartram, D. & Lindley, A. (1994). *Psychological Testing: The BPS 'Level A' Open Learning Programme*. Leicester: The British Psychological Society, BPS Books.
- Booth, J. & Horn, R. (2004). *Figure Reasoning Test (FRT) - Manual für FRT und FRT-J*. Frankfurt am Main: Swets Test Services.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. 4., vollst. überarb. Aufl. Berlin: Springer.
- Bourdon, B. (1955). *Bourdon-Test*. Göttingen: Hogrefe.

- Brickenkamp, R. (Hrsg.) (1975). Handbuch psychologischer und pädagogischer Tests. Göttingen: Hogrefe.
- Brickenkamp, R. (1987). Test d2. Aufmerksamkeits-Belastungs-Test. 6. Aufl. Göttingen: Hogrefe.
- Brickenkamp, R. (1991a). Fehlinterpretationen von Testleistungen? Anmerkungen zum Beitrag „Konzentrationsleistung ohne Konzentration“. *Diagnostica*, 37, 52-57.
- Brickenkamp, R. (1991b). Die Überraschung blieb aus. *Report Psychologie*, 11, 46-47.
- Brickenkamp, R. (1993). Zur Lösung der Problematik von Gesamtttestwerten in Konzentrationstests. *Report Psychologie*, 3, 24-26.
- Brickenkamp, R. (2002). Test d2: Aufmerksamkeits-Belastungs-Test (9. überarb. u. neu normierte Aufl.). Göttingen: Hogrefe.
- Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press.
- BROCKHAUS (2003). Der Brockhaus in einem Band. 10., neu bearb. Aufl. Leipzig: Brockhaus.
- Brosius, F. (1998). SPSS 8 Professionelle Statistik unter Windows. Bonn: International Thomson Publishing.
- Bulheller, S. & Häcker, H. O. (2002). Rechtschreibtest - Neue Rechtschreibregelung. 2., korr. Aufl. Frankfurt am Main, Swets Test Services.
- Bulheller, S. & Häcker, H. O. (2003). Peabody Picture Vocabulary Test (PPVT) - Deutschsprachige Fassung des PPVT-III für Jugendliche und Erwachsene von L. M. Dunn & L. M. Dunn. Frankfurt am Main, Swets Test Services.
- Bundesen, C. (2001). Attention: Models. In N. J. Smelser & P. B. Baltes (Hrsg.), *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 2, S. 878-884). Amsterdam: Elsevier.
- Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence*, 24, 25-52.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues*. 2. Aufl. (S. 69-76). New York: Guilford.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Chambers, D. & Reisberg, D. (1992). What an image depicts depends on what an image means. *Cognitive Psychology*, 24, 145-174.

- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustic Society of America*, 25, 975-979.
- Cohen, J. D., Aston-Jones, G. & Gilzenrat, M. S. (2004). A Systems-Level Perspective on Attention and Cognitive Control. In M. I. Posner (Hrsg.), *Cognitive Neuroscience of Attention*. (S. 71-90). New York: Guilford.
- COLLINS (1991). Deutsch-Englisch, Englisch-Deutsch. 2. Aufl., Stuttgart: Klett.
- Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- Colom, R. & Andrés-Pueyo, A. (2000). The study of human intelligence: A review at the turn of the millennium. *Psychology in Spain*, 4, 167-182.
- Coren, S. (2002). Left-Handedness. In V. S. Ramachandran (Hrsg.). *Encyclopedia of the Human Brain*. (Bd. 2, S. 685-694). Amsterdam: Academic Press.
- Coseriu, E. (1988). Sprachkompetenz: Grundzüge der Theorie des Sprechens. Tübingen: UTB.
- Daniel, M. H. (1997). Intelligence testing: Status and trends. *American Psychologist*, 52, 1038-1045.
- Daniels, J.C. (1962). Figure Reasoning Test (FRT). Stuttgart: Testzentrale.
- Della Salla, S. & Logie, R. H. (2002). Working Memory. In V. S. Ramachandran (Hrsg.), *Encyclopedia of the Human Brain*. (Bd. 4, S. 819-830). Amsterdam: Academic Press.
- Deutsch, J. A. & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70, 80-90.
- Dimitz, E. & Formann, A.K. (1979). Programm zur Schätzung der Item- und Personenparameter im dichotomen logistischen Modell von Rasch sowie Modellkontrollen und ein interaktives System zur Eingabe und Änderung der Programmparameter, Steuerbefehle und Daten. Institutsarbeit, 115. Wien: Institut für Höhere Studien.
- DUDEN (1989). Der Duden in 10 Bänden. Band 7: Etymologie - Herkunftswörterbuch der deutschen Sprache. 2., völlig neu bearb. u. erw. Aufl. Mannheim: Dudenverlag.
- DUDEN (1997). Der Duden in 10 Bänden. Band 8: Die sinn- und sachverwandten Wörter. 2. Aufl. Mannheim: Dudenverlag.
- DUDEN (2003a). Deutsches Universalwörterbuch. 5., überarb. Aufl. Mannheim: Dudenverlag.
- DUDEN (2003b). Der Duden in 10 Bänden. Band 5: Das große Fremdwörterbuch. 3., überarb. Aufl. Mannheim: Dudenverlag.

- Dunn, L. M. & Dunn L. M. (1997). Peabody Picture Vocabulary Test-Third Edition (PPVT-III). Circle Pines, MN: American Guidance Service.
- Engle, R. W. (1996). Working memory and retrieval: An inhibition-resource approach. In J. T. E. Richardson, R. W. Engle, L. Hasher, R.H. Logie, E. R. Stoltzfus & R. T. Zacks, *Working memory and human cognition*. (S. 89-119). New York: Oxford University Press.
- Fay, E. & Stumpf, H. (1992). Leistungsdaten. In Jäger, R. S. & Petermann, F. (Hrsg.). *Psychologische Diagnostik*. 2., veränd. Aufl. (S. 380-396). Weinheim: Psychologie Verlags Union.
- Fay, E., Trost, G. & Gittler, G. (2001). Intelligenz-Struktur-Analyse (ISA). 2., erw. Aufl. Frankfurt am Main: Swets Test Services.
- Fischer, G. (1974). Einführung in die Theorie psychologischer Tests. Bern: Huber.
- Fisseni, H.-J. (1990). Lehrbuch der psychologischen Diagnostik. Göttingen: Hogrefe.
- Floyd, R. G. (2005). Information-processing approaches to interpretation of contemporary intellectual assessment instruments. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues*. 2. Aufl. (S. 203-233). New York: Guilford.
- Geschwind, D. H. & Crabtree, E. (2003). Handedness and cerebral laterality. In M. J. Aminoff & R. Daroff (Hrsg.). *Encyclopedia of the Neurological Sciences*. (Bd. 2, S. 504-507). Amsterdam: Academic Press.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues*. 2. Aufl. (S. 136-181). New York: Guilford.
- Hahn, U. & Heit, E. (2004) Semantic similarity, Cognitive psychology of. In N. J. Smelser & P. B. Baltes (Hrsg.), *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 20, S. 13878-13881). Amsterdam: Elsevier.
- van der Heijden, A. H. C. (2001). Attention and action. In N. J. Smelser & P. B. Baltes (Hrsg.), *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 2, S. 868-871). Amsterdam: Elsevier.
- Heyde, G. (1995). Inventar Komplexer Aufmerksamkeit (INKA). Frankfurt am Main: Swets Test Services.
- Heyde, G. (2000). Inventar Komplexer Aufmerksamkeit (INKA). 2., überarb. u. erw. Aufl. Frankfurt am Main: Swets Test Services.
- Horn, W. (1962). Leistungsprüfsystem (LPS). Göttingen: Hogrefe.

- Horn, J. L. & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues*. 2. Aufl. (S. 41-68). New York: Guilford.
- Horn, J. L. & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107-129.
- Hunt, E. (2000). Let's hear it for crystallized intelligence. *Learning and Individual Differences*, 12, 123-129.
- Ibrahimovic, N. & Bulheller, S. (2005a). Wortschatztest - aktiv und passiv. Deutschsprachige Version des Mill-Hill Vocabulary Tests. Frankfurt am Main: Harcourt Test Services.
- Ibrahimovic, N. & Bulheller, S. (2005b). Mathematiktest. Grundkenntnisse für Lehre und Beruf. Frankfurt am Main: Harcourt Test Services.
- Jäger, R. S. & Jundt, E. (1981). Mannheimer Rechtschreib-Test (MRT). 2., erw. Aufl. Göttingen: Hogrefe.
- Jäger, A. O., Süß, H. M. & Beauducel, A. (1997). Berliner Intelligenzstruktur-Test: BIS-Test Form 4. Göttingen: Hogrefe.
- Johnson, A. & Proctor, G. (2004). Attention: Theory and practice. Thousand Oaks, Ca: Sage Publications.
- Kahneman, D. (1973). Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall.
- Kamphaus, R. W., Winsor, A. P., Rowe, E. W. & Kim, S. (2005). A history of intelligence test interpretation. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues*. 2. Aufl. (S. 23-38). New York: Guilford.
- Kaufman, A. S. & Kaufman, N. L. (1990). Kaufman Brief Intelligence Test (K-BIT). Circle Pines, MN: American Guidance Service.
- Kaufman, A. S. & Kaufman, N. L. (1993). Kaufman Adolescent and Adult Intelligence Test (KAIT). Circle Pines, MN: American Guidance Service.
- Krauth, J. (1995). Testkonstruktion und Testtheorie. Weinheim: Beltz, Psychologie Verlags Union.
- Kubinger, K. D. (1988). Aktueller Stand und kritische Würdigung der probabilistischen Testtheorie. In Kubinger, K.D. (Hrsg.). *Moderne Testtheorie - Ein Abriss samt neuesten Beiträgen*. (S. 19-83). Weinheim und München: Psychologie Verlags Union.
- Kubinger, K. D. (1995). Einführung in die psychologische Diagnostik. Weinheim: Beltz, Psychologie Verlags Union.

- Kubinger, K. D. (2003a). Gütekriterien. In Kubinger, K. D. & Jäger, R. S. (Hrsg.). *Schlüsselbegriffe der Psychologischen Diagnostik*. (S. 195-204). Weinheim: Beltz, Psychologie Verlags Union.
- Kubinger, K. D. (2003b). Testtheorie, Probabilistische. In Kubinger, K. D. & Jäger, R. S. (Hrsg.). *Schlüsselbegriffe der Psychologischen Diagnostik*. (S. 415-423). Weinheim: Beltz, Psychologie Verlags Union.
- Kyllonen, P. C. (1996). Is working memory capacity Spearman's *g*? In I. Dennis & P. Tapsfield (Hrsg.). *Human abilities: Their nature and measurement*. (S. 49-76). Mahwah, NJ: Erlbaum.
- Lehrl, S. (1995). Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B). 3. überarb. Aufl. Balingen: Perimed-Spitta.
- Lehrl, S. (1999). Mehrfachwahl-Wortschatz-Intelligenztest MWT-B. (4. überarb. Aufl.). Balingen: Spitta Verlag.
- Lehrl, S., Merz, J., Erzigkeit, H. & Galster, V. (1974). Der MWT-A - ein wiederholbarer Intelligenz-Kurztest, der weitestgehend unabhängig von seelisch-geistigen Störungen ist. *Nervenarzt*, 45, S. 364-369.
- Li, S.-C. & Schmiedek, F. (2001). Intelligence: Central conceptions and psychometric models. In N. J. Smelser & P. B. Baltes (Hrsg.), *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 11, S. 7635-7641). Amsterdam: Elsevier.
- Lienert, G. A. & Ratz U. (1994). Testaufbau und Testanalyse. 5., völlig Neubearb. und erw. Aufl. Weinheim: Beltz, Psychologie Verlags Union.
- van der Linden, M. (1998). The relationship between working memory and long-term memory. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie*, 321, 175-177.
- Lockhart, R. S. (2004). Memory retrieval. In N. J. Smelser & P. B. Baltes (Hrsg.), *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 14, S. 9613-9618). Amsterdam: Elsevier.
- Logie, R. H. (1995). Visuo-spatial working memory. Hove, England: Erlbaum.
- Logie, R. H. (1996). The seven ages of working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R.H. Logie, E. R. Stoltzfus & R. T. Zacks, *Working memory and human cognition*. (S. 31-65). New York: Oxford University Press.
- Logie, R. H. & Della Sala, S. (2001). Working memory, Psychology of. In N. J. Smelser & P. B. Baltes (Hrsg.), *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 24, S. 16587-16593). Amsterdam: Elsevier.

- Lucas, J. A. (2003). Memory, overview. In V. S. Ramachandran (Hrsg.), *Encyclopedia of the Human Brain*. (Bd. 2, S. 817-833). Amsterdam: Academic Press.
- Lukesch, H. & Mayrhofer, S. (2001). Konzentrations-Leistungs-Test: Revidierte Fassung. Göttingen: Hogrefe.
- Marschner, G. (1972). Revisions-Test (Rev.T.) nach Dr. Berthold Stender: Ein allgemeiner Leistungstest zur Untersuchung anhaltender Konzentration bei geistiger Tempoarbeit. Göttingen: Hogrefe.
- Moosbrugger, H. (1994). Testtheorie: Klassische Ansätze. In Jäger, R. S. & Petermann, F. (Hrsg.). *Psychologische Diagnostik*. 2., veränd. Aufl. (S. 310-322). Weinheim: Beltz, Psychologie Verlags Union.
- Moosbrugger, H. & Hartig, J. (2003). Testtheorie, Klassische. In Kubinger, K. D. & Jäger, R. S. (Hrsg.). *Schlüsselbegriffe der Psychologischen Diagnostik*. (S. 408-415). Weinheim: Beltz, Psychologie Verlags Union.
- Moosbrugger, H. & Oehlschlägel, J. (1996). Frankfurter Aufmerksamkeits Inventar (FAIR). Bern: Huber.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 56-60.
- Neisser, U., Boodoo, G., Bouchard, Jr., Th. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77-101.
- Nodari, C. (2002). Was heißt eigentlich Sprachkompetenz? In: Barriere Sprachkompetenz. Dokumentation zur Impulstagung vom 2. Nov. 01 im Volkshaus Zürich. *SIBP Schriftenreihe* 18, 9-14.
- Oehlschlägel, J. & Moosbrugger, H. (1991a). Konzentrationsleistung ohne Konzentration? Zur Schätzung wahrer Leistungswerte im Aufmerksamkeits-Belastungs-Test d2. *Diagnostica*, 37, 42-51.
- Oehlschlägel, J. & Moosbrugger, H. (1991b). Überraschende Validitätsprobleme im Aufmerksamkeits-Belastungs-Test d2. *Report Psychologie*, 09, 16-25.
- Oehlschlägel, J. & Moosbrugger, H. (1993). Bagatellisieren oder Abhilfe schaffen? *Report Psychologie*, 3, 20-23.
- Phillips, L. H. & Hamilton, C. (2001). The working memory model in adult aging research. In J. Andrade (Hrsg.), *Working memory in perspective*. (S. 101-125). Hove, England: Psychology Press.
- PONS (1985). Deutsch-Französisch = Teil 2. 2. Aufl. Stuttgart: Klett.

- Raaijmakers, J. G. W. & Shiffrin, R. M. (2002). Models of memory. In H. Pashler & D. Medin (Hrsg.). *Stevens' Handbook of Experimental Psychology*. Band 2, 3. Aufl. (S. 43-76). New York: Wiley.
- Raven, J. C. (1995). *Advanced Progressive Matrices - APM*. Set I & Set II. Oxford: Oxford Psychologists Press.
- Raven, J. C. (1996). *Standard Progressive Matrices - SPM*. Sets A-E. Oxford: Oxford Psychologists Press.
- Raven, J. C., Court, J. H. & Raven, J. (1994). *Mill Hill Vocabulary Scale*. 1994 Edition. Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 5A. Oxford: Oxford Psychologists Press.
- Raven, J. C., Raven, J. & Court, J. H. (1998a). *Raven's Progressive Matrices and Vocabulary Scales*. Grundlagen. Frankfurt am Main: Swets Test Services.
- Raven, J. C., Raven, J. & Court, J. H. (1998b). *Raven's Progressive Matrices and Vocabulary Scales*. APM Manual. Frankfurt am Main: Swets Test Services.
- Raven, J. C., Raven, J. & Court, J. H. (1999). *Raven's Progressive Matrices and Vocabulary Scales*. SPM Manual. Frankfurt am Main: Swets Test Services.
- Raven, J. C., Styles, I. & Raven, M. (1998). *Standard Progressive Matrices Plus - SPM Plus*. Sets A-E. Oxford: Oxford Psychologists Press.
- Raz, A. (2004). Attention. In Ch. D. Spielberger (Hrsg.), *Encyclopedia of Applied Psychology*. (Bd. 1, S. 203-208). Amsterdam: Academic Press.
- Resing, W. C. M. (2005). Intelligence testing. In K. Kempf-Leonard (Hrsg.), *Encyclopedia of Social Measurement*. (Bd. 2, S. 307-315). Amsterdam: Elsevier.
- Richardson, J. T. E., Engle, R. W., Hasher, L., Logie, R. H., Stoltzfus, E. R. & Zacks, R. T. (1996). *Working memory and human cognition*. New York: Oxford University Press.
- Richardson, J. T. E. (1996a). Evolving concepts of working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R.H. Logie, E. R. Stoltzfus & R. T. Zacks, *Working memory and human cognition*. (S. 3-30). New York: Oxford University Press.
- Richardson, J. T. E. (1996b). Evolving issues in working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R.H. Logie, E. R. Stoltzfus & R. T. Zacks, *Working memory and human cognition*. (S. 120-154). New York: Oxford University Press.
- Roediger III, H. L. & Karpicke, J. D. (2005). Learning and memory. In K. Kempf-Leonard (Hrsg.), *Encyclopedia of Social Measurement*. (Bd. 2, S. 479-486). Amsterdam: Elsevier.
- Rost, J. (1996). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Huber.

- Samstag, K., Sander, A. & Schmidt, R. (1971). DRE 3: Diagnostischer Rechentest für 3. Klassen. Weinheim: Beltz.
- Sanders, A. F. (2004). Dual Task Performance. *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 6, S. 3888-3892). Amsterdam: Elsevier.
- Sattler, J. B. (2000). Links und Rechts in der Wahrnehmung des Menschen. Zur Geschichte der Linkshändigkeit. Donauwörth: Auer.
- Schäfer, N. & Moosbrugger, H. (1993). Einige Validitätsaspekte des Frankfurter Aufmerksamkeits-Inventar (FAIR). *Arbeiten aus dem Institut für Psychologie der J.-W.-Goethe-Universität, Heft 10*, Frankfurt am Main.
- Schellig, D. (1997). Block-Tapping-Test. Frankfurt am Main: Swets Test Services.
- Schmidt, K.-H. & Metzler, P. (1992). Wortschatztest (WST). Weinheim: Beltz.
- Schmidt-Atzert, L., Bühner, M. & Enders, P. (2006). Messen Konzentrationstests Konzentration? - Eine Analyse der Komponenten von Konzentrationsleistungen. *Diagnostica*, 52, 33-44.
- Schmidt-Atzert, L., Büttner G. & Bühner M. (2004). Theoretische Aspekte von Aufmerksamkeits-/Konzentrationsdiagnostik. In Büttner G. & Schmidt-Atzert, L. (Hrsg.). Diagnostik von Konzentration und Aufmerksamkeit. (Jahrbuch der pädagogisch-psychologischen Diagnostik : Tests und Trends : Neue Folge Band 3). (S. 3-22). Göttingen: Hogrefe.
- Schonemann, P. H. (2005). Psychometrics of intelligence. In K. Kempf-Leonard (Hrsg.), *Encyclopedia of Social Measurement*. (Bd. 3, S. 193-201). Amsterdam: Elsevier.
- Siéroff, E. (2001). Attention: Multiple resources. In N. J. Smelser & P. B. Baltes (Hrsg.), *International Encyclopedia of the Social & Behavioral Sciences*. (Bd. 2, S. 884-888). Amsterdam: Elsevier.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Stankov, L. (2000). The theory of fluid and crystallized intelligence: New findings and recent developments. *Learning and Individual Differences*, 12, 1-3.
- Sternberg, R. J. (2004). Intelligence in humans. In Ch. D. Spielberger (Hrsg.), *Encyclopedia of Applied Psychology*. (Bd. 2, S. 321-328). Amsterdam: Academic Press.
- Sternberg, R. J. & Kaufman, J. C. (2002). Intelligence. In V. S. Ramachandran (Hrsg.), *Encyclopedia of the Human Brain*. (Bd. 2, S. 587-597). Amsterdam: Academic Press.

- Stoltzfus, E. R., Hasher, L. & Zacks, R. T. (1996). Working memory and aging: Current status of the inhibitory view. In J. T. E. Richardson, R. W. Engle, L. Hasher, R.H. Logie, E. R. Stoltzfus & R. T. Zacks, *Working memory and human cognition*. (S. 66-88). New York: Oxford University Press.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Süß, H.-M. (2003). Intelligenztheorien. In Kubinger, K. D. & Jäger, R. S. (Hrsg.). *Schlüsselbegriffe der Psychologischen Diagnostik*. (S. 217-224). Weinheim: Beltz, Psychologie Verlags Union.
- Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O. & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, 30, 261-288.
- Testkuratorium (der Föderation deutscher Psychologenverbände) (1986). Mitteilung. *Diagnostica*, 32, 358-360.
- Thurstone, L. L. (1938). Primary mental abilities. Chicago: University of Chicago Press.
- Towse, J. N. & Houston-Price, C. M. T. (2001). Reflections on the concept of the central executive. In J. Andrade (Hrsg.), *Working memory in perspective*. (S. 240-260). Hove, England: Psychology Press.
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12, 242-248.
- Trim, J., North, B. & Coste, D. (2001). Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin: Langenscheidt.
- Vallar, G. (2002). Short-term memory. In V. S. Ramachandran (Hrsg.), *Encyclopedia of the Human Brain*. (Bd. 4, S. 367-381). Amsterdam: Academic Press.
- Vecera, S. P. & Luck, S. J. (2002). Attention. In V. S. Ramachandran (Hrsg.), *Encyclopedia of the Human Brain*. (Bd. 1, S. 269-284). Amsterdam: Academic Press.
- Waltz, J. A. (2003). Memory, Working. In M. J. Aminoff & R. Daroff (Hrsg.), *Encyclopedia of the Neurological Sciences*. (Bd. 3, S. 90-95). Amsterdam: Academic Press.
- Wasserman, J. D. & Tulskey, D. S. (2005). A history of intelligence assessment. In D. P. Flanagan & P. L. Harrison (Hrsg.), *Contemporary intellectual assessment: Theories, tests, and issues*. 2. Aufl. (S. 3-22). New York: Guilford.
- Wechsler, D. (1991). Wechsler Intelligence Scale for Children-Third Edition (WISC-III). San Antonio, TX: The Psychological Cooperation.

- Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Hrsg.). *Varieties of attention*. (S. 63-102). New York: Academic Press.
- Yerkes, R. M. & Dodson, J. D. (1908). The relation of strengths of stimulus to rapidity of habit-formation. *Journal of Comparative and Neurological Psychology*, 18, 459-482.
- Zimbardo, P. G. (1992). *Psychologie*. 5., neu übers. u. bearb. Aufl. Berlin: Springer.

Anhang

A1 ILKO-Testbogen

Datum: _____

Name: _____ Alter (in Jahren): _____ Geschlecht: männlich weiblich

Schulbildung: Hauptschulabschluss Realschulabschluss Fachhochschulreife Abitur

Muttersprache: Deutsch andere: _____

WORTSCHATZTEST

Dieser Test dient dazu, Ihren Wortschatz zu überprüfen. Seine Aufgaben sind wie folgt aufgebaut:

Sie erhalten die Beschreibung eines Wortes, das gesucht wird. Falls Ihnen das Wort einfällt, schreiben Sie es bitte als Antwort auf. (**WICHTIG**: Es wird immer nur **ein** Wort gesucht!)

Beispiele: a) Geduldsspiel, bei dem viele kleine Einzelteile zu einem Bild, einer Figur zusammengesetzt werden müssen.

Antwort: Puzzle

b) Tochter des Bruders oder der Schwester.

Antwort: Nichte

Arbeiten Sie bitte konzentriert und zügig. Falls Ihnen ein gesuchtes Wort „auf der Zunge liegt“, es Ihnen aber nicht gleich einfällt, halten Sie sich bitte nicht zu lange bei der Aufgabe auf. Gehen Sie lieber zur nächsten, denn am Schluss zählt die Anzahl der Wörter, die Sie gefunden haben.

Der Test ist so aufgebaut, dass es kaum möglich ist, zu jeder Beschreibung sofort das passende Wort zu finden, auch wenn man über einen guten Wortschatz verfügt. Lassen Sie sich also nicht davon irritieren, wenn Ihnen nicht alle Wörter einfallen. Wichtig ist, so viele Wörter wie möglich zu finden.

Bitte schreiben Sie die Lösungswörter gut lesbar auf!

Der Test besteht aus insgesamt 18 Aufgaben. Für ihre Bearbeitung stehen Ihnen 9 Minuten zur Verfügung.

Falls Sie Fragen haben, wenden Sie sich bitte jetzt an den Testleiter.

STOPP! Beginnen Sie mit der Bearbeitung der Aufgaben erst dann, wenn der Testleiter Sie dazu auffordert!

1. Wort, das mit einem anderen oder mehreren Wörtern derselben Sprache bedeutungsgleich ist.

Antwort: _____

2. Zerstörungsarbeit von Wasser, Eis und Wind an der Erdoberfläche.

Antwort: _____

3. Vorrede, Vorwort, Einleitung eines literarischen Werkes.

Antwort: _____

4. Besonders durch Hypnose erreichter, dem Schlaf ähnlicher Dämmerzustand.

Antwort: _____

5. Zeitraum von 10 Tagen, Wochen, Monaten oder Jahren.

Antwort: _____

6. Deckname, angenommener Name, auch Künstlername.

Antwort: _____

7. Am Vormittag stattfindende künstlerische Veranstaltung.

Antwort: _____

8. Aufbau von Kraftwagen oberhalb des Fahrgestells.

Antwort: _____

9. Zusammenschluss von Unternehmen, die rechtlich und wirtschaftlich weitgehend selbstständig bleiben, aber durch Preisabsprachen o. Ä. den Wettbewerb ausschalten.

Antwort: _____

10. Vorwiegend im Handel übliche Form der Vergütung, die meist in Prozenten vom Umsatz berechnet wird.

Antwort: _____

11. Differenzbetrag, der sich nach Aufrechnung der Soll- und Habenseite des Kontos ergibt.

Antwort: _____

12. Bedeutsamkeit, Wichtigkeit in einem bestimmten Zusammenhang, Belang.

Antwort: _____

13. Vorzeichen, Vorbedeutung.

Antwort: _____

14. Beziehung, Zusammenhang, Verhältnis zwischen zwei Begriffen (Subjekt-Objekt), Dingen (Ursache-Wirkung), Größen u. a.

Antwort: _____

15. Vom eigenen Staatsgebiet eingeschlossener Teil eines fremden Staatsgebiets.

Antwort: _____

16. Allzu bereitwillige Anpassung an die jeweilige Lage aus Nützlichkeitsabwägungen.

Antwort: _____

17. Schwere Last, die als Fracht von geringem Wert zum Gewichtsausgleich mitgeführt wird.

Antwort: _____

18. Meist schmales, langes (Rand-)Beet mit Zierpflanzen, besonders als Begrenzung von Wegen oder Rasenflächen.

Antwort: _____

A2 Abstract

The present dissertation describes the development of two psychometric ability tests, Inventar Komplexer Aufmerksamkeit (INKA) and Inventar Lexikalischer Kompetenz (ILKO), the theoretical embedding of their measurement contents, and their validation. First, a classical approach is used to examine the validity of both instruments by applying additional tests either stating explicitly to measure the same or a different construct. The second validation is based on a cognitive model of human information processing, an expanded version of the working memory model introduced by Alan D. Baddeley and Graham J. Hitch in 1974. From the view of cognitive psychology, INKA appears as a measure of selective visual attention. Because of the specific content of its tasks, INKA additionally indicates the functioning of working memory. ILKO measures the retrieval of contents from semantic long term memory. From the view of psychometric theories of intelligence, INKA shows strong relations to tests of fluid intelligence. ILKO appears as a test of crystallized intelligence.

A3 Akademischer Lebenslauf des Verfassers

- 1985-92 Studium der Psychologie, Schwerpunkt Arbeits-, Betriebs- und Organisationspsychologie, Nebenfach Arbeitsrecht, an der Technischen Hochschule Darmstadt
- 1992 Erwerb des Diploms in Psychologie (Gesamtnote: sehr gut)

A4 Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit – abgesehen von den in ihr ausdrücklich genannten Hilfen – selbständig verfasst habe.

Gerd Heyde Groß-Gerau, 16. August 2006