



DIPLOMARBEIT

Titel der Diplomarbeit

AID 2 als Gruppentestung? Eine Machbarkeitsstudie

Verfasserin

Johanna Böck

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, März 2010

Studienkennzahl lt.
Studienblatt:

A 298

Studienrichtung lt.
Studienblatt:

Psychologie

Betreuerin / Betreuer:

Univ.-Prof. Dr. Mag. Klaus D. Kubinger

Danksagung

An erster Stelle möchte ich meinen Eltern und meiner Großmutter einen besonders großen Dank aussprechen, die mich während meiner gesamten Studienzeit nicht nur finanziell sondern auch emotional unterstützt haben.

Danken möchte ich auch Univ. Prof. Dr. Kubinger, der den Anstoß zu dieser Diplomarbeit gegeben hatte. Bei Unklarheiten und Problemen jeglicher Art stand er mit seinem Expertenwissen stets zur Verfügung. Mag. Poinstingl danke ich für die Unterstützung bei der Auswertung mit dem Programm *R*.

Des Weiteren bedanke ich mich bei den DirektorInnen, LehrerInnen und SchülerInnen der Volks- und Hauptschule sowie dem Gymnasium in Oberösterreich, die mir die Durchführung der Testung ermöglichten.

Großer Dank gebührt auch meinem Lebenspartner Andreas Endl für seine moralische Unterstützung und wissenschaftliche Kompetenz.

Sarah Böck und Sandra Weiss, die als Korrekturleserinnen fungierten und mich auf etwaige Unklarheiten und Fehler in meiner Arbeit hingewiesen hatten, sei an dieser Stelle ebenfalls gedankt.

Literaturverzeichnis

I. Einleitung	1
1. Zielsetzung	2
II. Theorie	4
2. AID 2	4
3. Testentwicklung	5
3.1. Theoretische Fundierung	5
3.2. Definition	5
3.3. Der Geltungsbereich	6
3.4. Test- und Antwortformat	6
3.5. Entwicklung der Items	8
3.6. Testanweisung	9
3.7. Wahl der Analysestichprobe	9
3.8. Testvorgabe	9
3.9. Aufgabenbewertung	9
3.10. Aufgabenanalyse	10
3.11. Weitere Vorgehensweise	10
III. Empirischer Teil	11
4. Soziale und Sachliche Folgerichtigkeit	11
4.1. Originalitems des AID 2	11
4.2. Theoretischer Hintergrund	11
4.3. Definition der zu messenden Fähigkeit	13
4.4. Geltungsbereich	13
4.5. Testformat	13
4.6. Antwortformat	14
4.7. Entwicklung der Items	14
5. Synonyme finden alias Wortschatztest	20
5.1. Originalitems des AID 2	20
5.2. Theoretischer Hintergrund	20
5.3. Definition der zu messenden Fähigkeit	22
5.4. Geltungsbereich	22
5.5. Testformat	22
5.6. Antwortformat	22
5.7. Entwicklung der Items	23
5.8. Testinstruktion	26
5.9. Wahl der Analysestichprobe	27
5.10. Testvorgabe	27
5.11. Aufgabenbewertung	28
5.12. Aufgabenanalyse	28
6. Funktionen abstrahieren	30
6.1. Originalitems des AID 2	30
6.2. Theoretischer Hintergrund	30
6.3. Definition der zu messenden Fähigkeiten	31
6.4. Geltungsbereich	32
6.5. Testformat	32
6.6. Antwortformat	32
6.7. Itementwicklung	33
6.8. Testinstruktion	36
6.9. Wahl der Analysestichprobe	36

6.10.	<i>Testvorgabe</i>	36
6.11.	<i>Aufgabenbewertung</i>	37
6.12.	<i>Aufgabenanalyse</i>	37
7.	Beobachtungen zur Testung	38
8.	Deskriptive Statistik	38
8.1.	<i>Geschlecht</i>	38
8.2.	<i>Region</i>	39
8.3.	<i>Schultyp</i>	40
8.4.	<i>Schulstufe</i>	41
9.	Statistische Berechnungen	43
9.1.	<i>Überprüfung der Daten auf Rasch-Modell-Konformität</i>	43
9.2.	<i>Likelihood-Quotienten-Test der ersten Schätzung</i>	44
9.3.	<i>Parameterschätzungen und Modelltests nach Itemselektion</i>	53
9.4.	<i>Mittelwertvergleiche</i>	83
10.	Diskussion und Ausblick	86
11.	Zusammenfassung	89
12.	Literatur	i
13.	Anhang	v
13.1.	<i>Wortschatz: Instruktion & Items</i>	v
13.2.	<i>Funktionen Abstrahieren: Instruktion & Items</i>	xxv
13.3.	<i>Itemleichtigkeitsparameter</i>	xxxvii
13.4.	<i>Personenparamter</i>	xlvi
13.5.	<i>Ansuchen an Landesschulrat</i>	liii
13.6.	<i>Beschreibung des Forschungsprojekts</i>	liv
13.7.	<i>Elternbrief</i>	lv
13.8.	<i>Abstract - Deutsch</i>	lvi
13.9.	<i>Abstract – English</i>	lvii
13.10.	<i>Lebenslauf</i>	lviii

Abbildungsverzeichnis

Abb.1: Grafischer Modelltest „Wortschatz“ Testform B, Teilungskriterium „Höhe des Rohwerts“	45
Abb.2: Grafischer Modelltest „Wortschatz“ Testform B, Teilungskriterium „Geschlecht“	46
Abb.3: Grafischer Modelltest „Wortschatz“ Testform B, Teilungskriterium „Region“ ..	46
Abb.4: Grafischer Modelltest „Wortschatz“ Testform D, Teilungskriterium „Höhe des Rohwerts“	48
Abb.5: Grafischer Modelltest „Wortschatz“ Testform D, Teilungskriterium „Geschlecht“	48
Abb.6: Grafischer Modelltest „Wortschatz“ Testform D, Teilungskriterium „Region“ ..	49
Abb.7: Grafischer Modelltest „Funktionen Abstrahieren“ Testform C, Teilungskriterium „Höhe des Rohwerts“	51
Abb.8: Grafischer Modelltest „Funktionen Abstrahieren“ Testform C, Teilungskriterium „Geschlecht“	51
Abb.9: Grafischer Modelltest „Funktionen Abstrahieren“ Testform C, Teilungskriterium „Region“	52
Abb.10: Grafischer Modelltest „Wortschatz“ Testform A, TK „Höhe des Rohwerts“ ..	56
Abb.11: Grafischer Modelltest „Wortschatz“ Testform A, TK „Geschlecht“	57
Abb.12: Grafischer Modelltest „Wortschatz“ Testform A, TK „Region“	57
Abb.13: Grafischer Modelltest „Wortschatz“ Testform C, TK „Höhe des Rohwerts“ ..	62
Abb.14: Grafischer Modelltest „Wortschatz“ Testform C, TK „Geschlecht“	63
Abb.15: Grafischer Modelltest „Wortschatz“ Testform C, TK „Region“	63
Abb.16: Grafischer Modelltest „Funktionen Abstrahieren“ Testform A, TK „Höhe des Rohwerts“	68
Abb.17: Grafischer Modelltest „Funktionen Abstrahieren“ Testform A, Teilungskriterium „Geschlecht“	69
Abb.18: Grafischer Modelltest „Funktionen Abstrahieren“ Testform A, Teilungskriterium „Region“	69
Abb.19: Grafischer Modelltest „Funktionen Abstrahieren“ Testform B, Teilungskriterium „Höhe des Rohwerts“	73
Abb. 20: grafischer Modelltest „Funktionen Abstrahieren“ Testform B, Teilungskriterium „Geschlecht“	74
Abb.21: Grafischer Modelltest „Funktionen Abstrahieren“ Testform B, Teilungskriterium „Region“	74
Abb.22: Grafischer Modelltest „Funktionen Abstrahieren“ Testform D, Teilungskriterium „Höhe des Rohwerts“	79
Abb.23: Grafischer Modelltest „Funktionen Abstrahieren“ Testform D, Teilungskriterium „Geschlecht“	79
Abb.24: Grafischer Modelltest „Funktionen Abstrahieren“ Testform D, Teilungskriterium „Region“	80

Tabellenverzeichnis

Tab. 1: Verteilung Geschlecht pro Untertest.....	39
Tab. 2: Verteilung Geschlecht pro Testform.....	39
Tab. 3: Verteilung Region pro Untertest.....	39
Tab. 4: Verteilung Region pro Testform.....	40
Tab. 5: Verteilung Schultyp pro Untertest.....	40
Tab. 6: Verteilung Schultyp pro Testform.....	41
Tab. 7: Verteilung Schulstufe pro Untertest.....	41
Tab. 8: Verteilung Schulstufe pro Testform.....	42
Tab. 9: Erste Parameterschätzung, Test Wortschatz A, nicht Rasch-Modell-konform ..	44
Tab. 10: Erste Parameterschätzung, Test Wortschatz B, Rasch-Modell-konform.....	45
Tab. 11: Erste Parameterschätzung, Test Wortschatz C, nicht Rasch-Modell-konform.	47
Tab. 12: Erste Parameterschätzung, Test Wortschatz D, Rasch-Modell-konform	47
Tab. 13: Erste Parameterschätzung, Test Funktionen Abstrahieren Testform A, nicht Rasch-Modell-konform	49
Tab. 14: Erste Parameterschätzung, Test Funktionen Abstrahieren Testform B, nicht Rasch-Modell-konform	50
Tab. 15: Erste Parameterschätzung, Test Funktionen Abstrahieren Testform C, Rasch-Modell-konform	50
Tab. 16: Erste Parameterschätzung, Test Funktionen Abstrahieren Testform D, nicht Rasch-Modell-konform	52
Tab. 17: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Wortschatz Testform A	54
Tab. 18: Itemfitstatistik Wortschatz Testform A	55
Tab. 19: Vierte Parameterschätzung, Test Wortschatz A, Rasch-Modell-konform.....	59
Tab. 20: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Wortschatz Testform C.....	60
Tab. 21: Itemfitstatistik Wortschatz Testform C.....	61
Tab. 22: Zweite Parameterschätzung, Test Wortschatz C, Rasch-Modell-konform.....	65
Tab. 23: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Funktionen Abstrahieren Testform A.....	66
Tab. 24: Itemfitstatistik Funktionen Abstrahieren Testform A.....	67
Tab. 25: Zweite Parameterschätzung, Funktionen Abstrahieren Testform A, Rasch-Modell-konform	71
Tab. 26: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Funktionen Abstrahieren Testform B.....	71
Tab. 27: Itemfitstatistik Funktionen Abstrahieren Testform B.....	72
Tab. 28: Vierte Parameterschätzung, Test Funktionen abstrahieren Testform B, Rasch-Modell-konform	76
Tab. 29: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Funktionen Abstrahieren Testform D.....	77
Tab. 30: Itemfitstatistik Funktionen Abstrahieren Testform D.....	78
Tab. 31: Vierte Parameterschätzung, Test Funktionen abstrahieren Testform D, Rasch-Modell-konform	82
Tab. 32: Dreifache Varianzanalyse, Wortschatz und Funktionen Abstrahieren.....	85

I. Einleitung

Die Anzahl der Anwendungsbereiche der Psychologischen Diagnostik sowie die Anforderungen, die an sie gestellt werden, liegen in der heutigen Zeit sehr hoch und werden in Zukunft wahrscheinlich noch wachsen. So ist sie aus derart verschiedenen Bereichen, wie dem Kranken- und Gesundheitswesen, Berufs- und Bildungsberatung, Personalauswahl und –selektion, Unternehmensberatung und Justizbereich nicht mehr wegzudenken.

Psychologische Tests sind dabei für das Sammeln diagnostisch relevanter und im Sinne der Fragestellung interpretierbarer Informationen essentiell und spielen aus diesem Grund neben dem Anamnesegespräch eine zentrale Rolle in der diagnostischen Praxis. Ziel ihrer Anwendung ist es, unter standardisierten Bedingungen Informationen über eine Person zu erheben und diese mit einer Referenzgruppe vergleichen zu können (Kubinger, 2009b). Diese Kriterien sowie die wissenschaftliche Fundiertheit unterscheiden psychologische Tests von anderen Untersuchungsverfahren (Lienert & Raatz, 1994).

Die Relevanz und Notwendigkeit psychologischer Tests zeigt sich dabei nicht nur in der theoretischen Argumentation sondern auch in der Praxis. Eine Metaanalyse von Schmidt und Hunter (1998) über Erfahrungen in der Personalselektion zeigte die Überlegenheit allgemeiner Fähigkeitstests hinsichtlich ihrer prädiktiven Validität für Leistung im Beruf oder Training im Vergleich zu anderen Verfahren.

Wie schon im ersten Satz erwähnt, wachsen mit den Anwendungsfeldern auch die Anforderungen an die Psychologische Diagnostik. Neben zentralen Gütekriterien wie Validität, Reliabilität und Objektivität steht auch immer mehr Ökonomie und Zumutbarkeit eines Verfahrens im Vordergrund. So soll die Diagnostik eine faire, valide und gute Entscheidungsfindung vorbereiten und unterstützen, sie soll aber auch zeitsparend und ökonomisch sein.

Eine naheliegende Lösung ist die Gruppentestung am Computer. Hier werden gleich mehrere Personen zur selben Zeit getestet, zudem erspart sich der Testleiter Zeit und Mühe bei der Vorgabe und Auswertung. An dieser Stelle sei aber erwähnt, dass eine Einzeltestung schon alleine auf Grund der möglichen Verhaltensbeobachtung und der genaueren Beschäftigung mit einer Person in der Regel mehr Informationen liefert und

für jede Fragestellung individuell die Kosten und Nutzen der verschiedenen Vorgabemodi abgewogen werden müssen.

Ein relativ moderner jedoch bereits bewährter und guter Ansatz, eine ökonomische Vorgehensweise ohne Messgenauigkeitseinbußen zu ermöglichen, ist das adaptive Testen. Der Name „adaptiv“ rührt daher, dass für die Testperson nach jeder bearbeiteten Aufgabe ihre Fähigkeit in der zu messenden Dimension geschätzt wird und eine passende – adaptierte – nächste Aufgabe vorgegeben wird (z.B. Kubinger, 2003).

Die Testperson erhält somit nur Items, die ihrem geschätzten Fähigkeitsniveau entsprechen. Dieses Vorgehen optimiert die Itemanzahl und –güte bezüglich einer Testperson und wirkt sich auch positiv auf deren Motivation aus, da keine Reihe von Items vorgegeben wird, die sie über- oder unterfordert (Hornke, 2001; Kubinger, 2003). Schließlich besteht ein weiterer Vorteil des adaptiven Testens darin, dass Items nicht so einfach weitergegeben werden können und auch das Abschreiben während der Testvorgabe erschwert bzw. verhindert wird, da individuell zusammengestellte Batterien vorliegen (vgl. Kubinger, 2009b).

1. Zielsetzung

Ziel dieser Diplomarbeit ist die Entwicklung einer Gruppentestversion der Untertests „Soziale und sachliche Folgerichtigkeit“, „Synonyme finden“ und „Funktionen abstrahieren“ des Adaptiven Intelligenz Diagnostikum 2 Version 2.2 (Kubinger, 2009a). Wie der Name schon vermuten lässt, erfolgt die Itemvorgabe im AID 2 adaptiv und ist daher schon per se als ökonomisch anzusehen. Acht von elf Untertests messen aufgrund des adaptiven Testens trotz geringer Testlänge definitionsgemäß genau, zudem liegen für fünf Untertests Kurzformen vor. Je nach gewünschter Messgenauigkeit variiert so die Dauer einer Testung zwischen 30 und 75 Minuten (Kubinger, 2009a, S.3).

Im Sinne einer Einzeltestung und Individualberatung repräsentiert der AID 2 somit ein höchst ökonomisches und zumutbares Verfahren. Werden jedoch ganze Schulklassen oder andere größere Gruppen getestet, würde eine Gruppentestversion des AID 2 zumindest für den Testleiter eine große Zeitersparnis bedeuten. Eine Testung in dieser Form ließe sich auch einfacher in den Schulalltag einbetten und würde dessen Ablauf weniger durcheinander bringen. SchulleiterInnen und LehrerInnen wäre es möglich, in kurzer Zeit Informationen über den Stand ganzer Schulklassen in verschiedenen Fähigkeitsbereichen einzuholen und gegebenenfalls Interventionen zu setzen. Auch für Eltern und Kinder würden derartige Testungen einen Informationsgewinn darstellen und

könnten besonders bei Entscheidungen wie „Hauptschule oder AHS?“ sehr hilfreich sein.

Eine Gruppentestversion des AID 2 könnte auch im Sinne eines Screeningsverfahren eingesetzt werden, um beispielsweise potenziell hochbegabte SchülerInnen in einer Klasse festzustellen oder auch Kinder mit einer eventuell vorliegenden Teilleistungsstörung. Im Zuge einer Einzeltestung könnten diese Diagnosen schließlich genauer überprüft werden und gegebenenfalls passende Interventionen vorgeschlagen werden.

Zu guter letzt wäre eine Gruppentestversion für wissenschaftliche Untersuchungen und Fragestellungen sehr hilfreich. So könnten in derselben Zeit rund zwanzigmal so viele Kinder getestet werden.

Aufgrund all dieser angeführten Vorteile stellt eine Gruppentestversion des AID 2 eine Erleichterung und Bereicherung für die psychodiagnostische Praxis dar. Diese Diplomarbeit ist im Sinne einer Machbarkeitsstudie zu verstehen, welche abklären soll, ob eine Vorgabe dreier Untertests des AID 2 im Gruppentestmodi möglich wäre, ohne an wichtigen Gütekriterien wie der Rasch-Modell-Konformität einbüßen zu müssen.

II. Theorie

2. AID 2

Eine genaue Vorstellung des AID 2 wird an dieser Stelle nicht geboten, der interessierte Leser, der sich nicht auf die hier stark gekürzte Repräsentation des Verfahrens beschränken will, sei auf Kubinger (2009a) verwiesen.

Der AID 2 ist eine Testbatterie für Kinder und Jugendliche, welche aus elf Untertests und drei Zusatztests besteht und der Erfassung komplexer sowie basaler Kognitionen dient. Der Großteil der Untertests lässt sich dabei deutlich einem der folgenden vier Faktoren zuordnen: „Informationsverarbeitung in der gesellschaftlichen Umwelt“, „Informationsverarbeitung neuer Inhalte“, „Auffassungskapazität“ oder „(Re-)Produktionsfähigkeit durch Strukturierung“. Zudem lassen sich die Untertests in förderungsabhängige und förderungsunabhängige unterscheiden (Kubinger, 2009a).

Unter Intelligenz wird sensu eines pragmatischen Ansatzes ein Bündel kognitiver Voraussetzungen verstanden, welche für den Erwerb von Wissen und der Entwicklung von Handlungskompetenzen notwendig sind. Das Testkonzept orientiert sich dabei an der Theorie von Wechsler (1944) und vage an Cattell (1963) und seiner Investmenttheorie, welche Wissen als investierte Intelligenz versteht. Nicht konform geht der AID 2 hingegen mit Spearman's Generalfaktortheorie (1927) und lehnt so die Berechnung eines Intelligenzquotienten, als Durchschnittswert aller geprüften Fähigkeiten, ab. Vielmehr verfolgt der AID 2 einen förderungsorientierten Ansatz und ermöglicht mittels einer Profilinterpretation das Erkennen von Stärken und Schwächen einer Testperson. Zur globaleren Beurteilung der Intelligenz kann die „Intelligenzquantität“, welche Aufschluss über die kognitive Mindestfähigkeit gibt, sowie der „Range der Intelligenz“ als Grad der Differenziertheit aller erfassten Fähigkeiten, zu Rate gezogen werden (Kubinger, 2009a).

Methodisch gesehen erlaubt der AID 2 unter Verwendung des *branched testing* eine adaptive Vorgehensweise ohne dabei auf einen Computer angewiesen zu sein. Die Testperson erhält so nur Aufgaben, die auf ihren individuellen Leistungsgrad abgestimmt sind. Der AID 2 ermöglicht somit zum einen eine extrem testökonomische Vorgehensweise, zum anderen wird die Leistungsmotivation, die ohnehin durch das ansprechende und vielseitige Testmaterial als gegeben angenommen werden kann, durch diese Vorgehensweise zusätzlich erhöht (Kubinger, 2009a).

Im Anschluss werden die Schritte einer Testentwicklung zunächst allgemeingültig und dann für jeden der drei Untertests im speziellen dargestellt. Dabei wird auch kurz auf die Originalversion des jeweiligen Untertests eingegangen.

3. Testentwicklung

3.1. Theoretische Fundierung

Je nachdem ob ein Test theoriegeleitet entwickelt wird oder nicht, spielt eine theoretische Fundierung eine mehr oder weniger wichtige Rolle. In jedem Falle ist jedoch eine Durchsicht der (falls) vorhandenen Literatur, die von dem zu messenden Merkmal handelt, hilfreich und wichtig. Denn wie Lewin (1963, S.205) schon anmerkte: „Es ist nichts praktischer als eine gute Theorie.“

So kann sie dazu beitragen, das zu messende Merkmal definitorisch einzugrenzen und greifbarer zu machen, was nun genau mit einem Test erfasst werden soll. Durch eine genaue und exakte Definition können auch Doppeldeutigkeiten eines Begriffs ausgeschaltet werden. Außerdem kann die Literatur helfen, ein Fähigkeitskonstrukt besser zu verstehen und in all seinen Facetten zu erfassen. Fragen wie: „Von welchen Komponenten hängt die Fähigkeit ab? Welche grundlegenden Kognitionen sind daran beteiligt? Wie entwickelt sich diese Fähigkeit und unter welchen Bedingungen?“ können wahrscheinlich anhand der Literatur zumindest ansatzweise geklärt werden.

Zu guter letzt kann sich der Leser Anregungen aus der Literatur zur Testentwicklung per se holen. Was ist der aktuelle Status quo in der Testentwicklung? Welche Faktoren sind zu beachten? Welche Tests existieren bereits, die dasselbe Fähigkeitskonstrukt messen? Welche Stärken und Schwächen haben diese?

3.2. Definition

Wie im Unterpunkt Theoretische Fundierung bereits angeschnitten, sind exakte, verständliche und genaue Definitionen für die erfolgreiche Entwicklung eines Tests entscheidend.

So stellt sich die Brauchbarkeit eines Tests in Frage, wenn nicht klar ist, was mit diesem überhaupt gemessen werden soll bzw. über welches Merkmal er eine Aussage trifft.

Besonders bei Begriffen aus der Psychologie, welche im Alltag eine oft weitere oder gänzlich andere Bedeutung tragen, ist eine definitorische Eingrenzung wichtig. Diese hilft auch, bei der Generierung von Items stets im Auge zu behalten, welches

Persönlichkeitskonstrukt gemessen werden soll und was eben genau nicht gemessen werden soll.

In diesem Zusammenhang sind auch die Konzepte *affordances* und *constraints* nach Greeno, Moore und Smith (1993) zu nennen. Unter *affordances* sind jene Merkmale zu verstehen, die das zu messende Konstrukt umreißen und als Regeln und generative Elemente bei der Itementwicklung mit einfließen sollen. *Constraints* hingegen sind kognitive Operationen welche im Konstrukt nicht vorgesehen sind und bei der Itemgenerierung somit ausgeschlossen werden sollten.

3.3. Der Geltungsbereich

Neben dem Zweck, den ein Test erfüllen soll, stellt sich auch die Frage, für welchen Personenkreis er dienen soll. Zu berücksichtigen ist dabei, dass sich die Entscheidung bezüglich des Geltungsbereichs sowohl bei der Auswahl einer geeigneten, repräsentativen Analytestichprobe als auch in einer späteren Normierung oder Eichung niederschlägt.

Je breiter der Geltungsbereich eines Tests, desto mehr müssen die Aufgabenschwierigkeiten streuen, desto vielseitiger und heterogener sind in der Regel die Aufgabeninhalte und desto einfacher sind die Anweisungen und Durchführungen zu halten (Lienert & Raatz, 1994).

3.4. Test- und Antwortformat

Individual – versus Gruppentest

Die Entscheidung für eine der beiden Möglichkeiten hängt stark mit dem Verwendungszweck und der Intention eines Tests ab. Eine Individualtestung bietet den Vorteil – der in der Praxis und je nach Verfahren mehr oder weniger genutzt wird – dass nur eine Testperson im Zentrum der Aufmerksamkeit steht. Ihr Verhalten, ihr verbaler und nonverbaler Ausdruck, ihre Interaktionen und Reaktionen können direkt beobachtet werden. Zum Teil sind Rückschlüsse auf ihre Anstrengungsbereitschaft, Motivation und ihr Verständnis für einzelne Aufgaben möglich. Ein weiterer Vorteil der Individualtestung besteht darin, dass interaktivere Aufgabenformate möglich sind.

Einen Nachteil hingegen bildet der Aufwand dieser Vorgehensweise – so kann in einer Gruppentestung in derselben Zeit das Vielfache an Personen getestet werden und dies in der Regel unter höherer Objektivität (vgl. Kubinger, 2009b). So finden weniger

Interaktionen zwischen Testleiter und Testperson statt, welche die Testleiterunabhängigkeit gefährden. Zudem können dank der neuen Technik auch in Gruppentestungen sehr vielseitige und zumindest virtuell interaktive Antwortformate vorgegeben werden. Es ist jedoch zu beachten, dass Gruppentests aufgrund der Administrierbarkeit zeitlich beschränkt vorgegeben werden müssen und es fraglich ist, ob schneller erbrachte Leistungen Ausdruck höherer Fähigkeiten sind (Kubinger, 2009a).

All diese Faktoren sollten bei der Entscheidung für oder gegen einen Individual- bzw. Gruppentest bedacht werden. Bei der Entwicklung eines dieser Verfahren sollten die jeweiligen Vorteile bewusst genutzt werden und nicht, wie es in der Praxis üblich ist, ein und dasselbe Verfahren als Einzel- und Gruppentest vorgelegt werden.

Antwortformat

Im Wesentlichen wird zwischen sechs Antwortformaten unterschieden (vgl. Lienert & Raatz, 1994):

- Richtig-Falsch Aufgabe
- Mehrfach-Wahl Aufgabe (= Multiple-Choice Aufgabe)
- Zuordnungs-Aufgabe
- Umordnungs-Aufgabe
- Ergänzungsaufgabe
- Kurzaufsatzaufgabe

Die ersten vier genannten Möglichkeiten zählen zu den geschlossenen Antwortformaten, die letzten beiden zu den offenen. Um sich bei der Testentwicklung für den passenden Antworttypus zu entscheiden, muss man sich mit seinen Vor- und Nachteilen sowie Eigenarten auseinandersetzen.

So haben Ergänzungsaufgaben den großen Vorteil, quasi frei von Zufallseinflüssen zu sein. Die Antworten können hier von der Testperson selbst formuliert werden, was als Vorteil aber auch als Nachteil gesehen werden kann. Positiv zu sehen ist, dass die inhaltliche Auswertung sehr bereichernd sein kann, die Testperson zu gründlichem Nachdenken angehalten wird und nicht bloß die richtige Antwort wiedererkennen oder gar nur erraten muss. Nachteilig ist jedoch, dass Bearbeitung sowie Auswertung mehr Zeit in Anspruch nehmen und nicht immer ganz eindeutig ist, ob eine Antwort als richtig oder falsch zu werten ist. Zudem werden Testpersonen mit Schwierigkeiten im

Ausdruck und in der Formulierung benachteiligt (Kubinger 2009b; Lienert & Raatz, 1994).

Diese Nachteile werden im geschlossenen Antwortformat verhindert. Die Bearbeitung und Auswertung ist hier äußerst ökonomisch, jedoch treten andere Probleme auf. Zum einen gilt es, die Ratewahrscheinlichkeit möglichst gering zu halten, was bei Richtig-Falsch Aufgaben nicht der Fall ist. Zum anderen sind gute Distraktoren zu entwickeln, das bedeutet, dass sie möglichst plausibel erscheinen und untereinander einigermaßen gleichwertig sein müssen (Lienert & Raatz, 1994).

3.5. Entwicklung der Items

Nachdem all diese grundlegenden Entscheidungen getroffen wurden, heißt es, sich Gedanken über die Entwicklung der einzelnen Items zu machen. Besonders beim adaptiven Testen ist ein relativ großer Itempool von Nöten und mit diesem wohl überlegte, präzise Regeln für die Itementwicklung sowie ausreichende Kreativität seitens des/der Testentwicklers/in.

So sollen die Items einerseits das mittlerweile gut definierte Persönlichkeitsmerkmal messen - und zwar nur dieses und kein anderes - andererseits unabhängig voneinander und möglichst vielseitig und unterschiedlich sein.

Dabei ist zu differenzieren, ob sich die Items zwar oberflächlich unterscheiden, aber dieselben psychometrischen Eigenschaften aufweisen (z.B. Schwierigkeit eines Items), oder ob sie sich auch in diesen unterscheiden. Im ersteren Fall treten Elemente des Items in Kraft – die sogenannten *Incidentals* - die dieses nur oberflächlich verändern ohne Einfluss auf dessen Schwierigkeit zu haben, im letzteren Fall hingegen sind es Elemente – die *Radicals* – welche auch die psychometrischen Eigenschaften verändern (Arendasy, Sommer, Gittler & Hergovich, 2006).

Zusammenfassend ist zu sagen, dass im Zuge der Itementwicklung einerseits berücksichtigt werden muss, welche Regeln bei der Itemgenerierung und damit einhergehend welche kognitiven Prozesse bei der Lösung der Items miteinbezogen werden oder nicht (*Affordances* vs. *Constraints*) und andererseits wie sich die Items rein inhaltlich bzw. in ihrer Schwierigkeit variieren lassen (*Incidentals* vs. *Radicals*).

3.6. Testanweisung

Noch vor der endgültigen Formulierung der einzelnen Items ist eine geeignete Testanweisung zu entwickeln. Diese kann generell für einen gesamten Test mit homogenen Items, speziell für einzelne Untertests oder bei sehr heterogenem Material separat für jede Aufgabe formuliert werden. Die Vorgabe von Beispielen spielt dabei eine wichtige Rolle, um das Verständnis der Instruktion, welche möglichst einfach und eindeutig formuliert sein sollte, zu überprüfen (Lienert & Raatz, 1994).

3.7. Wahl der Analytestichprobe

Im Idealfall sollte die Analytestichprobe für den Geltungsbereich im höchstmöglichen Maße repräsentativ sein. In der Realität ist dies nicht immer möglich, zumeist werden Stichproben herangezogen die eben gerade verfügbar sind und somit keine Zufallsauswahl realisiert. Die Analytestichprobe darf sich aber in keinem Falle von dem späteren Testkollektiv offensichtlich unterscheiden (Kubinger, 2009b; Lienert & Raatz, 1994).

3.8. Testvorgabe

Bei der Testvorgabe muss darauf geachtet werden, dass Verfälschungsversuche jeglicher Art (z.B. durch Verwendung unerlaubter Hilfsmittel oder Abschreiben vom Nachbarn) unterbunden werden. Zudem ist festzulegen, was genau die Aufgaben des Testleiters sind um möglichst hohe Objektivität zu gewährleisten. Auch die Durchführungsbedingungen sind zu berücksichtigen und so zu gestalten, dass eine störungsfreie Testung möglich ist (Lienert & Raatz, 1994).

3.9. Aufgabenbewertung

Nach der Testvorgabe erfolgt naturgemäß die Auswertung. Schon während der Testkonstruktion soll entschieden werden, wie die Bewertung erfolgt. Welche Antworten als richtig oder falsch gewertet werden, ob es eine Gewichtung oder eine Vergabe von „Teilpunkten“ gibt. Pauschal ist zu sagen, dass die Aufgabenbewertung so einfach wie nur irgend möglich zu gestalten ist (Lienert & Raatz, 1994).

So können komplexe oder nicht ganz eindeutige Auswertungsmodi die Verrechnungssicherheit gefährden. Aber auch bei eindeutigen und einfachen Regeln sind Auswertungsfehler keine Seltenheit (vgl. Kubinger, 2009b, S. 43). Deshalb sollte zumindest stichprobenartig die Auswertung auf Fehler überprüft werden.

3.10. Aufgabenanalyse

Nach der Vorgabe der Items und der Auswertung dieser ist eine Aufgabenanalyse notwendig, um ungeeignete Items auszuschalten bzw. verbesserungswürdige zu revidieren. Außerdem ist von Interesse, ob die vorher festgelegten „Radicals“ tatsächlich einen Einfluss auf die Schwierigkeit des Items haben und ob es noch andere nicht beachtete Regeln gibt.

Die Analyse kann dabei am Testrohwert oder an einem Außenkriterium oder auch an beiden vorgenommen werden. Bei der Entscheidung für die eine oder andere Analyseverfahren sind mehrere Faktoren wie das Ausmaß der Homogenität der Items, die Absicht eine spätere Validierung etc. zu berücksichtigen (Lienert & Raatz, 1994).

3.11. Weitere Vorgehensweise

Die nächsten Schritte in der Testentwicklung werden hier nur kurz skizziert - der interessierte Leser sei auf Lienert & Raatz (1994) verwiesen.

Nachdem die Testvorform ausreichend genau analysiert und revidiert wurde, wird die endgültige Aufgabenzusammenstellung, deren Anordnung und Darstellung festgelegt. Diese Testendform ist zumindest auf ihre Validität, Reliabilität und Objektivität zu überprüfen, aber auch eine kritische Begutachtung bezüglich Nützlichkeit, Zumutbarkeit, Ökonomie, Verfälschbarkeit und Fairness ist wünschenswert (Lienert & Raatz, 1994).

Zudem ist das Gütekriterium der Skalierung zu überprüfen, welches erfüllt ist wenn „die laut Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen adäquat abbilden“ (Kubinger, 2009 b, S. 82).

Schließlich ist eine Eichung des Tests vorzunehmen, wobei darauf geachtet werden muss, dass diese repräsentativ für die vorher definierte Stichprobe ist und gemäß der DIN 33430 die Aktualität der Eich Tabellen zumindest alle acht Jahre überprüft werden muss.

III. Empirischer Teil

4. Soziale und Sachliche Folgerichtigkeit

Allem vorweg ist für diesen Untertest zu sagen, dass eine Vorgabe der entwickelten Items im Zuge dieser Diplomarbeit nicht erfolgte. Der Grund hierfür war, dass eine Programmierung am Computer die zeitlichen und personalen Ressourcen sowie den Umfang und Aufwand einer Diplomarbeit überstiegen hätte.

Eine Vorgabe im Paper-Pencil Format wäre im Zuge einer Schulstunde nicht realisierbar gewesen und hätte zudem einen hohen Materialverbrauch bedeutet, da jeder Testperson sämtliche Kärtchen der Bildergeschichten vorgelegt werden hätten müssen. Außerdem hätte ein derartiger Vorgabemodus viele Störvariablen mit sich gebracht (richtiges Aufschreiben der gewählten Bilderreihenfolge auf den Testbögen, Verlust von Bilderkärtchen, entstehendes Durcheinander und Chaos durch hohe Anzahl von Bildergeschichten und Kärtchen etc.).

Im Folgenden wird daher nur auf den theoretischen Hintergrund und die Itementwicklung eingegangen. Die darauf folgenden Schritte können erst präzisiert und endgültig festgelegt bzw. durchgeführt werden, wenn eine Computerversion dieses Untertests vorliegt.

4.1. Originalitems des AID 2

Mit dem Untertest 4 (Soziale und Sachliche Folgerichtigkeit) soll das Verständnis für die Abfolge sozialen Geschehens oder alltäglicher Sachgegebenheiten erfasst werden. Dabei werden der Testperson Bilder einer Bildergeschichte vorgegeben, die sie in die richtige Reihenfolge bringen soll (Kubinger, 2009a).

4.2. Theoretischer Hintergrund

Wie bereits an früherer Stelle erwähnt, lassen sich nahezu alle Untertests des AID 2 einem von vier Faktoren („Informationsverarbeitung in der gesellschaftlichen Umwelt“, „Informationsverarbeitung neuer Inhalte“, „Auffassungskapazität“ oder „(Re-)Produktionsfähigkeit durch Strukturierung“) klar zuordnen. Eine Ausnahme bildet neben dem Untertest 2 „Realitätssicherheit“ der Untertest 4 „Soziale und Sachliche Folgerichtigkeit“ (Kubinger, 2009a). Die Fähigkeit, welche in diesem Test gemessen wird, scheint sich aus einem komplexen, vielseitigen und verstrickten

Kognitionskonstrukt zu ergeben und ist dadurch auch definitorisch nicht so einfach einzugrenzen.

Ganz allgemein gesehen kann jedoch davon ausgegangen werden, dass die Testperson bei der Bearbeitung dieses Untertests pro Item vor einem Problem steht. Es gibt einen Anfangszustand – die ungeordneten Bilder – und einen definierten Zielzustand – die Bilder so zu ordnen, dass sie eine in sich schlüssige Geschichte ergeben. Um vom Anfangszustand zum Ziel zu gelangen, muss die Testperson die richtigen Entscheidungen treffen und geeignete Operatoren wählen. So ist ein möglicher Lösungsweg, Ursache – Wirkungszusammenhänge bzw. in sich logische konsistente Abfolgen zu erkennen und damit die abgebildeten Ereignisse zu verstehen.

Kotovsky und Simon (1990) stellten fest, dass je mehr Dimensionen und Merkmale bei der Lösung eines Problems zu beachten sind, desto schwieriger gestaltet sich seine Lösung. So besteht ein schwieriges Problem darin, dass eine so lange Folge von Operationen zu planen und durchzuführen ist, dass der Weg vom Anfangs- zum Zielzustand kaum mehr erkennbar ist.

Auch ein großer Zeitabstand erschwert das Erkennen von Ursache-Wirkungs-Zusammenhängen (vgl. Goswami, 2001).

Ein weiterer Operator, um den Zielzustand zu erreichen, ist, sich in der Vergangenheit bewährte Lösungen oder Regeln ins Gedächtnis zu rufen. Beim Ordnen von Bildern können so eigene Erfahrungen, Erlebnisse, Alltagsroutinen oder auch sogenannte *Skripts*¹ als Vorlage dienen.

Diese Vorgehensweise wird analoges Problemlösen genannt, da Analogien zwischen der aktuellen und einer vergangen Situation hergestellt werden (Chen & Honomichl, 2004). Das analoge Problemlösen kann in vielen Fällen hilfreich sein, nämlich dann, wenn die Erfahrungen und Anwendungen in relativ gleicher Form erfordert werden.

Wird jedoch ein wesentlicher Unterschied zwischen aktueller und vergangener Situation übersehen oder ist bei der Lösung eines Problems neuartiges Denken und Kreativität gefragt, behindert das gedankliche Festhalten an frühere Situationen den Lösungsprozess (Mayer, 1979; Wessels, 1994).

¹„Eine Eintragung (Speicherung) im Langzeitgedächtnis, die die reguläre Ergebnisabfolge in bestimmten (spezifischen) Situationen, Kontexten (z.B. Restaurant, Arzt, Bahnfahrt) beschreibt.“ (Häcker & Stapf, 2009, S. 891)

4.3. Definition der zu messenden Fähigkeit

Affordances

Die Reihenfolge der Bilder soll durch logisches Schlussfolgern, Beachten und Kombinieren der verschiedenen Merkmale und Inhalte der Bilder sowie Erkennen von Ursache–Wirkungszusammenhängen erschließbar sein. Das Wissen über den Ablauf sozialer Interaktionen und die Verfügbarkeit von *Skripts* bieten eine Stütze beim Finden der richtigen Bilderfolge, sofern die Bildergeschichte eine reguläre und übliche Ergebnisabfolge darstellt. Weicht sie hingegen von dieser ab, muss die Testperson in der Lage sein, sich von ihren Alltagserfahrungen und *Skripts* zu lösen und durch schlussfolgerndes Denken und kritischer Überprüfung ihrer Annahmen auf den richtigen Weg finden.

Constraints

Beim Lösen der Items soll die Verfügbarkeit von Fachwissen keine Rolle spielen. So dürfen z.B. keine chemischen Verbindungsprozesse in einer Bildergeschichte abgebildet werden, zu deren Lösung die Testperson über die Reaktionseigenschaften verschiedener Moleküle Bescheid wissen müsste etc.

Auch die Seh- und Beobachtungsgenauigkeit darf keinen Einfluss auf die Lösungswahrscheinlichkeit des Items haben. Alle relevanten Merkmale der Bildergeschichte müssen gut sichtbar und erkennbar sein.

4.4. Geltungsbereich

Da die Instruktion schriftlich erfolgt und die Testperson in der Lage sein soll, diese selbstständig zu lesen und zu verstehen, wurde der ursprüngliche Altersbereich des AID 2 (6;0 bis 15;11 Jahre) auf 8;0 bis 15;11 Jahre eingeschränkt. Zudem muss die Testperson die Muttersprache „Deutsch“ haben oder diese Sprache zumindest fließend in Wort und Schrift beherrschen, damit eine Vorgabe und Interpretation der Tests sinnvoll ist.

4.5. Testformat

Wie aus dem Titel dieser Arbeit ersichtlich und durch bereits zu Beginn angeführten Punkten begründet, fiel die Entscheidung auf eine Gruppentestversion.

4.6. Antwortformat

Das ursprüngliche Antwortformat des AID 2, welches einer Umordnungsaufgabe entspricht, wurde auch für die Gruppentestversion beibehalten.

4.7. Entwicklung der Items

Radicals

Auf Basis der in 4.2. angeführten Literaturrecherche und der Analyse der Originalitems in Bezug auf ihre Unterschiede in den verschiedenen Schwierigkeitsgraden, stellte sich heraus, dass vor allem fünf Eigenschaften der Bildergeschichten ausschlaggebend für deren Schwierigkeitsgrad sind.

Anzahl der Bilder

Die Anzahl der Bilder nimmt mit aufsteigendem Schwierigkeitsgrad zu. Je länger eine Geschichte ist, umso schwieriger ist es scheinbar für die Testperson diese in die richtige Reihenfolge zu bringen und ihren Ablauf zu erkennen. Anders formuliert: Da es mehr Bilder gibt ist der Weg zwischen Anfangszustand und Zielzustand länger, es sind mehr Schritte zu setzen und Entscheidungen zu fällen, wodurch die Lösung wesentlich schwerer fällt.

Anzahl der Akteure

Mit Anzahl der Akteure sind die in der Bildergeschichte aktiv handelnden oder beteiligten Personen bzw. Lebewesen gemeint. Diese Komponente spielt aber nur bei der Unterscheidung der niedrigsten mit sämtlichen anderen Schwierigkeitsstufen eine Rolle. So darf bei den einfachsten Items nur ein Akteur vorkommen, während ansonsten bis zu 4 Akteure möglich sind. Abfolgen sind demnach für Testpersonen leichter zu erfassen und nachzuvollziehen, wenn nur ein Subjekt daran beteiligt ist.

Zeitlicher Abstand

Damit ist die Zeit gemeint, die in etwa zwischen zwei Bildern einer Geschichte vergangen ist. Ein kleiner Zeitabstand macht die Abfolge der Bilder nachvollziehbarer, da sie direkt und unmittelbar zusammenhängen. Umso größer der zeitliche Abstand, desto schwieriger ist das Item.

Bei der Entwicklung der Bildergeschichten wurde zwischen 3 Zeitabständen unterschieden:

- kleiner Zeitabstand: nicht länger als 5 Minuten
- mittlerer Zeitabstand: max.1 Stunde
- langer Zeitabstand: über 1 Stunde (können auch Monate, und sofern noch nachvollziehbar sogar Jahre sein)

Hilfsmittel

Bei Hilfsmittel handelt es sich um Gegenstände oder Prozesse, die nicht unmittelbar zum Hauptgeschehen zählen, jedoch helfen, die Bildergeschichte in die richtige Reihenfolge zu bringen. Dies können eher offensichtliche wie z.B. eine Uhr im Hintergrund, einer untergehende Sonne, oder auch etwas weniger offensichtliche sein (z.B. leer werdende Teller beim Essen). Die Testperson muss sich so im Grunde nur auf eine Dimension des Problems beziehen, um es lösen zu können.

Wichtig ist jedoch, dass die Bildergeschichten auch ohne diese Hilfsmittel noch immer eindeutig gelöst werden könnten.

Abweichen von Skripts und achten auf Details

Die Komponente bezieht sich darauf, dass bei Abweichungen von regulären sachlichen und sozialen Abfolgen Details oft eine entscheidende Rolle spielen und auf diese geachtet werden muss, um den Sinn und Ablauf der Geschichte zu erfassen.

Ist diese Komponente ausgeprägt, gestaltet sich das Lösen einer Bildergeschichte schwieriger, da das Achten auf Details notwendig ist um die Bilder in die richtige Reihenfolge zu bringen, wären diese Details nicht abgebildet, würde die Reihenfolge der Geschichte nicht eindeutig bestimmbar sein. Zudem muss sich die Testperson von ihren Alltagserfahrungen lösen und mögliche Abfolgen der Bilder kritisch überprüfen.

Beispiel: Eine Morgenzeremonie wird dargestellt

- Bild 1: Ein Mann wacht auf, er liegt noch im Bett, neben ihm läutet der Wecker
- Bild 2: Er zieht sich sein Hemd an
- Bild 3: Er frühstückt (hat sein Hemd, aber immer noch Pyjamahose an)
- Bild 4: Er putzt sich die Zähne (mit Hemd und Pyjamahose)
- Bild 5: Er befindet sich im Stiegenhaus (mit Hemd und Pyjamahose) und schlägt sich mit flacher Hand auf die Stirn

- Bild 6: Er ist wieder in seinem Zimmer und zieht sich seine schwarze Anzugshose an
- Bild 7: Er geht die Stiegen hinab (mit Hemd und Anzugshose)

Das Skript der normalen Morgenzeremonie wird dadurch gebrochen, dass der Mann vergisst, sich seine Hose umzuziehen und erst im Stiegenhaus seinen Fehler bemerkt. Die Testperson muss, um die Geschichte zu verstehen und die Bilder in die richtige Reihenfolge zu bringen, auf das Detail „Hose“ achten.

Ortswechsel

Ein Ortswechsel in der Geschichte gestaltet deren Lösung schwieriger. So treten in den einfachen Bildergeschichten keine Ortswechsel auf, d.h. die Geschichte spielt immer am selben Ort und vor demselben Hintergrund (z.B. in einem bestimmten Raum, im Garten, ...). In den schwierigeren Items können Ortswechsel auftreten, dabei wird zwischen Ortswechsel unterschieden, die in der Bilderfolge beobachtbar sind (z.B. Person geht bei der Tür raus) und jenen, die es nicht sind: hier befindet sich die Person auf einmal in einer anderen Umgebung ohne dass in einem Bild auf einen Wechsel der Umgebung hingedeutet wurde. Diese Art von Ortswechsel tritt erst ab der vierten Schwierigkeitsstufe auf. Zudem wird zwischen „kleinen Ortswechsel“, welche sich dadurch kennzeichnen, dass die Orte sehr nahe beieinander liegen (z.B. von einem Raum in den Nebenraum, vom Haus in den Garten etc.) und „großen Ortswechsel“ unterschieden, bei denen eine längere Strecke zwischen den beiden Orten liegt (z.B. von zu Hause zu einem Einkaufshaus, von zu Hause zum Flughafen etc.).

Incidentals

Während die oben angeführten Komponenten die Schwierigkeit eines Items beeinflussen, hat das Thema der Bildergeschichten voraussichtlich keinen derartigen Effekt. Zu beachten ist aber, dass als Thema nur soziale und sachliche Abfolgen in Frage kommen, die der Testperson kein Fachwissen z.B. in Physik oder Chemie abverlangen. Mögliche Themenbereiche sind:

- Helfen
- Reisen und Straßenverkehr
- Spielen
- etwas konstruieren (z.B. malen, bauen, erforschen, Gartenarbeit, ...), darunter fällt auch etwas präsentieren, vorstellen
- etwas zerstören
- kaufen, etwas erwerben
- Sport
- Vorgänge in Natur, darunter fallen auch Tiere
- Streit, schimpfen
- Feste, Rituale
- Darstellen von Berufen (v.a. jene, die nicht unmittelbar mit Konstruieren zusammenhängen)
- Erfüllung von Grundbedürfnissen, Alltagsroutinen (essen, schlafen gehen, aufstehen)
- etwas erlernen, erkennen von Zusammenhängen

Sofern keine verbreiteten Rollenklischees damit gebrochen werden, sind auch Geschlecht, Alter sowie Aussehen der Akteure in der Bildergeschichte variierbar, ohne die Schwierigkeit eines Items zu beeinflussen.

Richtlinien zum Erstellen von Items

Für das Erstellen einer Bildergeschichte ist zunächst eine Idee über deren Inhalt von Nöten.

Obig angeführte Themenbereiche dienen dabei als Denkanstoß.

Nach der ersten grundlegenden Idee über den Inhalt der Bildergeschichte, muss der Testentwickler festlegen, welchen Schwierigkeitsgrad diese Geschichte erreichen soll und dementsprechend den im Anschluss angeführten Regeln befolgend den Inhalt und die Abfolge der Bilder ausbauen und präzisieren.

Bildergeschichten des ersten Schwierigkeitsgrades:

Die Anzahl der Bilder ist drei. Der zeitliche Abstand zwischen den Bildern muss klein sein. In den Bildergeschichten darf kein Ortswechsel stattfinden und es darf nur eine Person (oder Lebewesen: z.B. Hund, Katze) vorkommen. Es muss möglich sein, die Bildergeschichte in die richtige Reihenfolge zu bringen, ohne auf Details achten zu müssen (= keine Abweichung von *Skripts*). Hilfsmittel sind, da die Bildergeschichten ohnehin sehr einfach gestaltet sind, nicht von Nöten.

Bildergeschichten des zweiten Schwierigkeitsgrades:

Die Anzahl der Bilder beträgt zwischen 3 und 6. Der zeitliche Abstand zwischen den Bildern soll klein, maximal mittelmäßig groß sein. In den Bildergeschichten dürfen höchstens kleine Ortswechsel stattfinden (z.B. vom Haus in den Garten), wobei im vorangegangenen Bild ersichtlich sein muss, dass der Ort gewechselt wird (z.B. Person geht zur Türe). Es dürfen bis zu 4 Personen vorkommen, welche direkt am Geschehen beteiligt sind; „Statisten“ sind zusätzlich erlaubt. Es muss möglich sein die Bildergeschichte in die richtige Reihenfolge zu bringen ohne auf Details achten zu müssen (= keine Abweichung von *Skripts*). Hilfsmittel sollten bei 5 oder 6 Bildern oder wenn ein mittelmäßig großer Zeitraum vorliegt vorhanden sein. Bei 3-4 Bildern sollten wenn überhaupt nur subtile Hilfsmittel verwendet werden.

Bildergeschichten des dritten Schwierigkeitsgrades:

Die Anzahl der Bilder beträgt zwischen 5 und 6. Es sind auch längere zeitliche Abstände möglich, jedoch muss ein/e Gegenstand/Person vorhanden sein, der sich in dieser Zeit entwickelt (z.B. Blume), diese/r Gegenstand/Person ist Hauptthema der Geschichte; ansonsten sollten kurze bis mittlere Zeitabstände gewählt werden. In den Bildergeschichten dürfen maximal kleine Ortswechsel stattfinden (z.B. vom Haus in den Garten), wobei im vorangegangenen Bild ersichtlich sein muss, dass der Ort gewechselt wird (z.B. Person geht zur Türe). Es dürfen bis zu 4 Personen vorkommen, welche direkt am Geschehen beteiligt sind; „Statisten“ sind zusätzlich erlaubt. Es muss möglich sein die Bildergeschichte in die richtige Reihenfolge zu bringen, ohne auf Details achten zu müssen (= keine Abweichung von *Skripts*). Hilfsmittel sollen nicht vorhanden sein.

Bildergeschichten des vierten Schwierigkeitsgrades:

Die Anzahl der Bilder beträgt zwischen 4 und 6. Die Zeitabstände sollen kurz bis mittellang sein, bei 4 Bildern sind auch längere zeitliche Abstände möglich. In den Bildergeschichten dürfen auch größere Ortswechsel stattfinden (z.B. von Haus in Geschäft). Es dürfen bis zu 4 Personen vorkommen, welche direkt am Geschehen beteiligt sind, „Statisten“ sind zusätzlich erlaubt. Um die Bilder in die richtige Reihenfolge zu bringen, soll bei der Bestimmung der Reihenfolge von 2 (max. 3) Bildern auch auf Details geachtet werden müssen (= kleine Abweichung von *Skripts*), die übrigen Bilder sollen ohne das Beachten von Details zeitlich richtig eingeordnet werden können. Hilfsmittel sollten nicht vorhanden sein.

Bildergeschichten des fünften Schwierigkeitsgrades:

Die Anzahl der Bilder beträgt zwischen 6 und 8. Die Zeitabstände sollen lang sein. In den Bildergeschichten dürfen auch größere Ortswechsel stattfinden (z.B. vom Haus in ein Geschäft). Es dürfen bis zu 4 Personen vorkommen, welche direkt am Geschehen beteiligt sind; „Statisten“ sind zusätzlich erlaubt. Um die Bilder in die richtige Reihenfolge zu bringen, muss auch auf Details geachtet werden (= Abweichung von *Skripts*). Hilfsmittel sollten nicht vorhanden sein.

5. Synonyme finden alias Wortschatztest

5.1. Originalitems des AID 2

Der Untertest „Synonyme finden“ dient der Messung des elementaren Sprachverständnisses, indem überprüft wird, ob die Testperson die Bedeutung sprachgebundener Begriffe erfasst und in der Lage ist, diese Begriffe alternativ auszudrücken. Dabei werden der Testperson Wörter vorgelesen, zu denen sie Synonyme nennen soll (Kubinger, 2009a).

5.2. Theoretischer Hintergrund

Der Erwerb der Sprache, welche den Menschen von den anderen Lebewesen unterscheidet, stellt eine der wichtigsten Entwicklungsaufgaben in der (frühen) Kindheit dar und ist weit komplexer als sie auf den ersten Blick scheinen mag. So sind gleich sechs zum Teil eigenständige Wissenssysteme aufzubauen, die dem Kind ermöglichen, die prosodischen, phonologischen, morphologischen, syntaktischen und lexikalisch-semantischen Regularitäten sowie den kontextangemessenen handlungsorientierten Gebrauch seiner Muttersprache zu erwerben. Für den Erwerb des Wortschatzes und seiner Bedeutungsstruktur ist die Komponente Lexikon verantwortlich (Oerter & Montada, 2008).

Ein Kind lernt ab seinem ersten bis zu seinem sechzehnten Lebensjahr im Durchschnitt 60.000 Wörter, wobei die Varianz beträchtlich ist. Kinder, die mit 24 Monaten die 50-Wort Grenze noch nicht erreicht haben (sog. Late Talker), sind gefährdet, eine bleibende Störung der Sprachentwicklung und daraus folgende Defizite für die kognitive und psychosoziale Entwicklung auszubilden (Oerter & Montada, 2008, S. 510).

Ein besonderer Anstieg des Wortschatzes tritt beim Eintritt in die Grundschule auf (Neuhaus, 1962).

Interessant ist, dass am Aufbau der Sprache die verschiedenen Wortarten unterschiedlich stark beteiligt sind. Ab dem 18. Lebensmonat ist ein schneller Erwerb von Objektbezeichnungen zu erkennen, 12 Monate später nimmt der Anteil an Verben stark zu (Oerter & Montada, 2008).

Im Laufe der Schulzeit zeigt sich eine Strukturveränderung: Während der Anteil der Substantiva (ca. 24 %) am Gesamt der verwendeten Wörter zwischen dem 8. und 14. Lebensjahr und ebenso jener der Verben (ca. 21 %) gleichbleibt, nimmt der Anteil der

Adjektiva zu und jener der Konjunktionen und Bindewörter (vor allem „und“) ab (Neuhaus, 1962, S. 134 f.).

Der prozentuelle Anteil der Adjektiva bleibt dennoch verhältnismäßig klein, so zeigte eine Untersuchung von Beckmann (1928, zitiert nach Busemann, 1965), dass dieser im ersten Schuljahr (1,1%) und im 4. Schuljahr (3,6%) beträgt.

Neben der Breite des Wortschatzes ist auch entscheidend, ob ein Kind einem Wort dieselbe Bedeutung zuschreibt wie ein Erwachsener, so kommt es häufig zu Phänomenen der Übergeneralisierung oder Überdiskriminierung, was bedeutet, dass Kinder den Geltungsbereich eines Wortes sehr viel breiter oder enger als Erwachsene setzen (vgl. Oerter & Montada, 2008). Es ist daher nicht nur von Interesse, ob ein Kind ein Wort kennt, sondern auch, was es darunter versteht.

Während das Kleinkind den Dingen gegenüber noch eine ausschließlich konkrete Haltung zeigt, sind für die entwickelte situationsfreie Sprache Abstrakta, ohne welche Verständigungen über unanschauliche Sachverhalte nicht möglich wären, die eigentlichen Bedeutungsträger. Anhand von Erfahrungen, durch die Fähigkeit zu abstrahieren und den Erwerb von Erkenntnissen lernt das Kind im Laufe seiner Entwicklung seinen Wortschatz über konkrete Anschauungen hinaus auszubilden (Neuhaus, 1962).

Bei der Bedeutungsbestimmung einzelner Wörter geht das Kind wahrscheinlich wie ein Wissenschaftler vor. Es entwickelt Hypothesen über die Bedeutung jedes neu gelernten Wortes und widerlegt oder bestätigt diese anhand seiner Erfahrungen (vgl. Gerrig & Zimbardo, 2008).

Die Ausbildung des Wortschatzes und eines richtigen Wortverständnisses ist demnach stark von der Umwelt, in der eine Person aufwächst, abhängig. So ordnet auch Cattell (1963) die Fähigkeitskomponente, welche im Wortschatz gefragt ist, in seinem Intelligenzkonstrukt der kristallinen Intelligenz zu. Die kristalline Intelligenz ist im Gegensatz zur fluiden Intelligenz stark kultur-, erfahrungs- und bildungsabhängig und kann im Laufe des Lebens durch die Nutzung sämtlicher Lernmöglichkeiten gebildet und weiterentwickelt werden. Mit dieser Anschauung konform gehend erwies sich der Untertest 6 „Synonyme finden“ in Untersuchungen zum AID 2 als förderungsabhängig (Angerer, 1991; Mayer-Frühwirth, 1988).

Auch andere Studien zeigten die Abhängigkeit der verbalen Intelligenz im Allgemeinen und dem Wortschatz im Speziellen von Umweltvariablen wie Schultyp, familiären

Prozess- und Strukturmerkmalen und sozioökonomischen Hintergrund (McElvany, Becker & Lüdtke, 2009; Schneider & Stefanek, 2004).

5.3. Definition der zu messenden Fähigkeit

Affordances

Im Zuge dieses Untertests soll zum einem der Wortschatz der Testperson erfasst werden, zum anderen, ob sie den gegebenen Wörtern die richtige Bedeutung zuweist.

Erläuterung zu „Wortschatz“

Unter Wortschatz versteht man „das gesamte einem Mensch zur rezeptiven und produktiven Sprachbenützung verfügbare Repertoire an Wörtern“ (Häcker & Stapf, 2009, S. 1100).

Constraints

Keinen Einfluss beim Lösen dieser Items darf hingegen Fachwissen – in welchem Bereich auch immer – spielen. So sind Fachtermini (es sei denn, sie werden im Alltag in ihrer richtigen Bedeutung häufig verwendet) strengstens zu vermeiden, wie auch veraltete, nicht mehr verwendete Begriffe. Schließlich ist darauf zu achten, die Satzstruktur der Items möglichst einfach zu halten. Mangelndes Verständnis aufgrund zu komplexer Sätze sollte nicht der Grund für falsche Antworten sein.

5.4. Geltungsbereich

Siehe 4.4.

5.5. Testformat

Siehe 4.5.

5.6. Antwortformat

Das ursprüngliche freie Antwortformat des Untertests „Synonyme finden“ wurde für die Gruppentestversion durch ein Multiple-Choice-Format, welches jenem im Lexikon-Wissen-Test von Wagner-Menghin (2004) nachempfunden ist, ersetzt. Dadurch wurde eine objektivere und ökonomischere Auswertung ermöglicht und zum anderen das Kind

nicht gezwungen, Wörter aufzuschreiben, wo Schwierigkeiten bei der Rechtschreibung als Hemmschuh fungieren könnten.

Auch die Herangehensweise, den Wortschatz durch die Nennung von Synonymen zu messen, wurde aufgegeben. (Aus diesem Grund wird die Gruppentestversion des Untertests 6 nicht „Synonyme finden“ sondern „Wortschatz“ genannt). So sind sich SprachwissenschaftlerInnen uneinig, ob es wirkliche Synonyme überhaupt gibt und falls ja, ab welcher Stufe zwei Wörter als Synonyme gelten (Lyons, 1968).

Stattdessen soll die Testperson die vorgegebene Definition eines Wortes durch das Wählen der richtigen Antworten vervollständigen.

Beispiel:

Symbiose

Symbiose bezeichnet das *a) Kennenlernen* unterschiedlicher Arten, das für
b) sich Bekriegen
c) Zusammenleben
d) Auslöschen

alle Beteiligten *e) bedenklich* ist.
f) vorteilhaft
g) negativ
h) ungewohnt

Durch diese Vorgehensweise soll überprüft werden, ob die Testperson das Wort kennt und ihm die richtige Bedeutung zuschreibt. Die Ratewahrscheinlichkeit ist aufgrund der hohen Anzahl an Distraktoren – die Testperson muss zweimal hintereinander aus vier möglichen Antworten die einzig richtige wählen – gering.

5.7. Entwicklung der Items

Radicals

Aus den Untersuchungen, welche im Unterpunkt „Theoretischer Hintergrund“ vorgestellt wurden, sowie aus der Analyse der Originalitems des AID 2 ließen sich drei Merkmale, die aller Voraussicht nach einen Einfluss auf die Schwierigkeit eines Items haben, ableiten.

Wortart

Adjektiva sind im Vergleich zu Nomen und Verben weniger stark im Wortschatz vertreten. Die Annahme, dass eine Testperson ein Adjektiv mit höherer Wahrscheinlichkeit nicht kennt oder falsch definiert, liegt daher nahe.

Konkret versus Abstrakt

Wie schon weiter oben ausführlich erläutert, sind Kinder in ihren Anschauungen und Beschreibungen noch sehr konkret und bedienen sich abstrakter Begriffe nur selten.

Dahingehend dürfte Kindern das Erkennen der richtigen Bedeutung von Wörtern, die sich auf anschauliche bzw. greifbare Dinge, Objekte, Lebewesen, Verhaltensweisen etc. beziehen, einfacher fallen.

Bei dieser Komponente wurde zwischen zwei Ausprägungen unterschieden:

- Konkrete Begriffe: darunter fallen Objekte, Lebewesen etc., die mit den Händen angefasst werden können oder Verhaltensweisen bzw. Abläufe, die direkt (d.h. mit bloßem Auge; es sind keine Schlussfolgerungen oder Interpretationen notwendig, um das Verhalten als solches zu erkennen) beobachtet werden können und sich nicht (nur) im Inneren einer Person abspielen (z.B. Kleidung, Keller, spielen, ...).
- Abstrakte Begriffe: beziehen sich auf Sachverhalte,² die nicht angefasst werden können oder direkt beobachtbar sind (Liebe, denken, radikal, ...).

Relevanz im Alltag

Eine wichtige Komponente scheint schließlich die Häufigkeit der Verwendung eines Wortes im Alltag zu sein. Wird ein Wort sehr häufig verwendet oder spielt der bezeichnete Sachverhalt im Alltag des Kindes eine große Rolle, wird das Kind mit hoher Wahrscheinlichkeit ein Verständnis dafür entwickelt haben, was mit dem Wort ausgedrückt werden soll und was damit gemeint ist.

Bei dieser Komponente gibt es drei Abstufungen:

- Hohe Relevanz im Alltag: Das Kind verwendet das Wort (mit hoher Wahrscheinlichkeit) häufig im Zuge von alltäglichen Gesprächen, oder der durch das Wort ausgedrückte Sachverhalt an sich spielt eine zentrale Rolle im Alltag des Kindes (es hat z.B. den Gegenstand mit sehr hoher

² Sachverhalt steht hier stellvertretend für Objekte, Personen, Dinge, Tätigkeiten, Eigenschaften

Wahrscheinlichkeit schon verwendet, oder die Tätigkeit schon mehrmals ausgeführt etc.)

- Relevanz im Alltag: Das Kind hat das Wort (höchstwahrscheinlich) schon selbst in Alltagsgesprächen verwendet und es des Öfteren von wichtigen Bezugspersonen (Eltern, Lehrer, ...) gehört, oder der durch das Wort ausgedrückte Sachverhalt an sich spielt eine zentrale Rolle im Leben der Eltern, in der Schule oder in den Medien
- Geringe Relevanz im Alltag: Das Kind hat das Wort wahrscheinlich schon von wichtigen Bezugspersonen gehört, selbst verwendet es das Wort jedoch kaum oder nie; der durch das Wort ausgedrückte Sachverhalt an sich spielt eine Rolle – jedoch keine zentrale - im Leben der Eltern, in der Schule oder in den Medien

Incidentals

Eine inhaltliche Variation der Items ohne ihre Schwierigkeit zu beeinflussen ist möglich, indem Wörter welche in ein- und dieselbe Gruppe fallen, gewählt werden (z.B. abstrakte Nomen welche im Alltag selten verwendet werden).

Richtlinien zum Erstellen von Items

Zur Erstellung eines Items ist zunächst ein „Ausgangswort“ zu überlegen. Folgende Themenkreise können dabei als Denkanstoß dienen:

- Natur (z.B. Tiere, Witterung,...)
- Haushalt, Hausarbeit
- Verkehr
- Studium- und Schulbereich
- Berufe
- Gefühle
- Unterhaltung/Freizeit/Sportarten

Um die Schwierigkeit der Items zu variieren sind verschiedene Ausprägungen obig genannter *Radicals* zu realisieren. Die Ausprägungen der Komponente „Wortart“ sind objektiv bestimmbar, jene der Komponenten „Konkret vs. Abstrakt“ und insbesondere „Relevanz im Alltag“ sind zum Teil stark vom subjektiven Ermessen des/der Testentwicklers/in abhängig. Daher sollte für diese beiden Komponenten die Kodierung

von zumindest zwei TestentwicklerInnen unabhängig voneinander durchgeführt und etwaige Differenzen in der Einschätzung besprochen werden.

Der nächste Schritt besteht darin, einen Satz zu erstellen, der das „Ausgangswort“ definiert, dabei soll die Gültigkeit und Richtigkeit der Definition durch das Verwenden offizieller und aktueller Lexika gesichert werden. Bei der Definition ist zu beachten, dass diese nicht mehr als 20 Wörter enthält, da ansonsten die Länge des Satzes ein Hindernis beim Verstehen und Lösen des Items darstellen könnte. Es soll schließlich nicht das Satzverständnis des Kindes, sondern sein Wortschatz gemessen werden. Aus dem selbigen Grund soll die Definition nicht mehr als zwei Satzteile enthalten. Schließlich soll die Definition insgesamt so einfach wie möglich gehalten werden, da die Kenntnis des definierten Wortes geprüft werden soll und nicht die Kenntnis von Wörtern, die im Satz verwendet werden.

Nach der Entwicklung einer geeigneten Definition sind zwei sinntragende Wörter zu wählen. Für jedes dieser beider Wörter sind drei Alternativen zu überlegen, die an ihrer Stelle in den Satz eingefügt werden können. Dabei ist darauf zu achten, dass der Satz bei jeglicher Kombination der Alternativen grammatikalisch richtig bleibt. Falls notwendig sind dazu mehrere Artikel oder Verbindungen hintereinander anzuschreiben (z.B. der/die/das). Zudem müssen Alternativen in verschiedenen Kombinationen inhaltlich sinnvoll sein, jedoch darf nur eine Kombination -und zwar die ursprüngliche Definition- das Wort richtig definieren.

Insgesamt wurden 78 Items für diesen Untertest entwickelt, die im Anhang angeführt sind.

5.8. Testinstruktion

Da im Zuge der Gruppentestung mehrere Untertests vorgegeben wurden, die sich aus jeweils unter sich homogenen Items zusammensetzten, erfolgte pro Untertest eine sogenannte „spezielle Testanweisung“, wobei die Anweisungen schriftlich gegeben und durch ein nachfolgendes Übungsbeispiel erläutert wurden. Die Auflösung des Übungsbeispiels wurde jedoch mündlich durchgenommen, um zu überprüfen, ob jede Testperson die Instruktion richtig verstanden hatte.

Die genaue Instruktion zum Untertest 6 (ebenso zum Untertest 9) findet sich im Anhang.

5.9. Wahl der Analysestichprobe

Um den Geltungsbereich abzudecken, wurden SchülerInnen aus verschiedenen Schultypen im Alter von 8;00 bis 15;11 Jahren getestet. Die Stichprobe setzte sich dabei aus zwei Hauptschulen und der dritten und vierten Klasse einer Volksschule aus dem Mühlviertel sowie einem Privatgymnasium aus der Stadt Linz zusammen, wobei hier von der 5. bis 9. Schulstufe getestet wurde.

Die Auswahl der Schulen erfolgte dabei nicht zufällig, sondern es wurde in jenen Schulen getestet, in denen die Direktoren und der oberösterreichische Landesschulrat ihr Einverständnis gaben. Die gewählte Analysestichprobe ist aus diesem Grund und da nur Schulen aus Oberösterreich getestet wurden, für den vorher festgelegten Geltungsbereich (8;00-15;11 jährige, deutschsprachige Personen) nicht im höchsten Maße repräsentativ, unterscheidet sich aber auch nicht offensichtlich davon.

Das Schreiben an den Landesschulrat, sowie der Elternbrief befinden sich im Anhang.

5.10. Testvorgabe

Für die Testvorgabe wurde das gesamte Testmaterial in vier Testformen aufgeteilt, wobei versucht wurde, diese Testformen gleich schwierig zu gestalten. Diese Aufteilung hatte einerseits den Zweck, die Versuchung und Möglichkeit des Abschreibens zu eliminieren oder zumindest zu reduzieren. Andererseits zeigte sich in Voruntersuchungen mit einem 8-, einem 10- und einem 13-jährigen Kind, dass die Bearbeitung der Untertests, nachdem sie in vier Testformen aufgeteilt worden waren, in der Regel in 50 min (das entspricht einer Schulstunde) gut zu bewältigen war. Dabei erhielt jede Testperson neben den Untertests „Wortschatz“ und „Funktionen abstrahieren“ die Untertests „Alltagswissen“ und „Angewandtes Rechnen“, welche im Zuge einer anderen Diplomarbeit entwickelt wurden (Eiter, in Arbeit).

Die Gruppentestungen erfolgten in den Klassenzimmern und in Anwesenheit von einer oder zwei Testleiterinnen. Nach einer kurzen allgemeinen Instruktion, in der die SchülerInnen informiert wurden, dass ihnen nun eine Reihe von Aufgaben vorgelegt wird, anhand derer gemessen werden soll, was sie schon alles können bzw. wissen und die jede/r für sich bearbeiten soll, sowie dass Blicke auf die Bögen der Nachbarn aufgrund der verschiedenen Gruppen ohnehin sinnlos wären, wurde der erste Untertest („Funktionen abstrahieren“) ausgeteilt. Die SchülerInnen lasen sich die Instruktion leise durch, im Anschluss fand die Auflösung des Übungsbeispiels mündlich statt. Nachdem

alle SchülerInnen, zumindest dem Ermessen der Testleiterin nach, die Aufgabenstellung verstanden hatten, wurden die SchülerInnen gebeten, nun die eigentlichen Aufgaben zu beantworten und wenn sie mit allen Aufgaben fertig seien, ein kurzes Handzeichen zu geben, sich leise ihre Antworten noch einmal durchzulesen und zu überdenken. Spätestens nach acht Minuten oder schon vorher, falls ein Großteil der Klasse bereits fertig war, wurden die SchülerInnen gebeten, den Testbogen beiseite zu legen und den nächsten Testbogen („Angewandtes Rechnen“), welcher Ihnen nun ausgeteilt wurde, zu bearbeiten. Die SchülerInnen wurden darauf hingewiesen, dass sie die nicht fertiggestellten Aufgabenbögen später, falls ihnen noch Zeit bliebe, weiterbearbeiten dürften. Die Vorgabe und Durchführung des Untertests „Angewandtes Rechnen“ erfolgte analog zu „Funktionen abstrahieren“. Nach spätestens zehn Minuten wurde der Untertest „Wortschatz“ ausgeteilt. Im Anschluss daran –nach spätestens 15 Minuten – erhielten die SchülerInnen den letzten Untertest („Alltagswissen“). Hatten die SchülerInnen alle Untertests fertig beantwortet bzw. nahte das Ende der Unterrichtseinheit, wurden alle Testbögen eingesammelt.

Als Dankeschön erhielten die Kinder eine kleine Süßigkeit, von der sie jedoch im Vorhinein noch nichts wussten.

5.11. Aufgabenbewertung

Die Aufgabenbewertung wurde so einfach wie möglich gestaltet. So wurde nur zwischen richtiger und falscher Antwort unterschieden, eine Vergabe von Teilpunkten fand nicht statt. Eine Antwort galt nur dann als richtig, wenn sich die Testperson beide Male für die richtige Option entschieden hatte, d.h. den Satz richtig vervollständigt hatte. Alle anderen Antwortkombinationen wurden als falsch gewertet.

5.12. Aufgabenanalyse

Für die Aufgabenanalyse wurde ein Ansatz aus der probabilistischen Testtheorie verwendet: Die Überprüfung der Testformen auf ihre Konformität mit dem dichotom logistischen Modell nach Rasch.

Dies dient der Abklärung, ob die vorgegebenen Testformen eindimensional messen und die Anzahl der gelösten Aufgaben eine „erschöpfende Statistik“ für die Ausprägung der zu messenden latenten Eigenschaft einer Person darstellt. Eine genauere Beschreibung des Rasch-Modells, die, um verständlich zu sein, einigermaßen ausführlich sein müsste, ist an dieser Stelle weder sinnvoll noch notwendig, da sie den Umfang der Diplomarbeit

sprengen würde und ohnehin gute Werke, die diese Modell ausführlich und nachvollziehbar vorstellen, vorliegen (der interessierte Leser kann sich bei Fischer, 1974 oder Kubinger, 1989 vertieft damit auseinandersetzen).

An dieser Stelle sei nur erwähnt, dass die Konformität der Testformen mit dem Rasch-Modell mittels Likelihood-Quotienten-Test nach Andersen (1973) überprüft wurde. Dieser Modelltest bedient sich der Eigenheit der „Stichprobenunabhängigkeit“. So kann das Rasch-Modell angenommen werden, wenn die Parameterschätzungen in verschiedenen Teilstichproben statistisch gleich sind (Kubinger, 1989).

Musste für eine Testform das Rasch-Modell verworfen werden, fand eine itemweise Überprüfung mittels Z-Test nach Fischer und Scheiblechner, Grafischem Modelltest und Itemfitstatistik nach Hatzinger und Mair sowie eine qualitative Analyse der Items statt. Mit dieser Vorgehensweise wurde entschieden, welches Item bzw. welche Items aus dem Pool zu entfernen sind, damit das Rasch-Modell zumindest a posteriori angenommen werden konnte.

Die genaue Vorgehensweise pro Testform sowie die Ergebnisse finden sich im Unterpunkt „Überprüfung der Daten auf Rasch-Modell-Konformität“.

6. Funktionen abstrahieren

6.1. Originalitems des AID 2

Ziel des Untertests „Funktionen abstrahieren“ ist es, die Fähigkeit, durch Abstraktion zu einer Begriffsbildung zu gelangen, zu erfassen. Dazu werden der Testperson zwei Begriffe vorgelesen und sie muss eine gemeinsame zentrale Funktion oder funktionale Eigenschaft der beiden nennen (Kubinger, 2009a).

Erläuterung zu „Abstraktion“

Abstraktion ist „der gedankliche Prozess, der Teilinhalte vom Ganzen des Bewusstseinsinhaltes ablöst, heraushebt oder zurückdrängt.“ (Häcker & Stapf, 2009, S. 5)

6.2. Theoretischer Hintergrund

Um uns in unserer Umwelt aus unendlich vielen unterschiedlichen Stimuli ³ zurechtzufinden, geben wir den Gegenständen, Personen, Ereignissen etc. um uns Namen und verbinden sie mit bestimmten Eigenschaften, über die sie verfügen. Zudem behandeln wir nicht jeden Stimulus als Einzelfall, sondern ordnen ihn einer bekannten Kategorie zu. Beim Bilden dieser Kategorien hat der Grad der äußeren Ähnlichkeit zwischen den verschiedenen Stimuli einen entscheidenden Einfluss, aber auch andere Eigenschaften wie zum Beispiel die Funktion können der Ausgangspunkt von Kategorienbildungen sein (Goswami, 2001; Neuhaus, 1962).

Von welchen Faktoren hängt es nun ab, ob aufgrund der äußeren Ähnlichkeit, der Funktion, beider zusammen oder anderer Merkmale kategorisiert wird? Und ist beim Sehen ein Stimulus oder beim Lesen seines Namens das Ausgangsmerkmal der Kategorisierung präziser und zentraler als andere Eigenschaften und daher einfacher zu abstrahieren?

Nach Busemann (1965) neigen Kinder dazu, Dinge ganzheitlich zu sehen und sie nicht zu zerpfücken. Sie sehen oft nur eine zentrale Eigenschaft eines Gegenstands, welche häufig seine Funktion ist.

Erst im Jugendalter gelingt es zunehmend kognitive Aktivitäten zu koordinieren und kontrollieren sowie irrelevante Informationen effizienter auszublenden (Case, 1992).

³ Im Kapitel 7 ist „Stimulus“ ganz allgemein zu verstehen und steht stellvertretend für Gegenstände, Lebewesen, Einrichtungen, Ereignisse, Verhaltensweisen, Gefühle etc.

Nach Busemann (1965) ist die Gegenstandswelt für das Kind noch eine Welt von brauchbaren, unnützen oder gar schädlichen Sachverhalten. Oft sind es Erlebnisse der Handhabung der Gegenstände, die entscheiden, ob zwei Dinge als zusammengehörig oder nicht empfunden werden. Seine Theorie bestätigte sich auch in dem von Busemann (1965) durchgeführten „*Kleingerümpel*-Versuch“. Dabei sollten die Kinder etwa 20 kleine Gegenstände wie Nägel, Schrauben, Bleistifte, Farbstifte, Kreide usw. ordnen. Während das Kleinkind nach Größe, Materialgleichheit, Farbähnlichkeit und Formähnlichkeit sortierte, war für das intellektuell normale Kind von 6 Jahren der Verwendungszweck entscheidend.

Die Funktion scheint demnach zumindest für (kleine) Gegenstände eine zentrale Eigenschaft und entscheidend für Kategorienbildungen zu sein. Doch gilt dies auch für andere Stimuli?

Auch nach Keil (1994) steht bei der Kategorisierung von Geräten und Apparaten die Funktion im Vordergrund, bei Lebewesen ist jedoch die innere und äußere Struktur entscheidend. Dies legt die Vermutung nahe, dass die Abstraktion von Funktionen bei Gegenständen einfacher als bei Lebewesen fällt.

Ganz allgemein ist zu sagen, dass das Denken in der Kindheit noch im engen Zusammenhang mit der Erfahrung verläuft. Je augenscheinlicher das Vergleichsmerkmal ist, umso leichter fällt es dem Kind, Verknüpfung zu finden (Oerter & Montada, 2008).

Dahingehend müsste das Erkennen gemeinsamer Funktionen bei Stimuli, welche das Kind selbst in Anspruch nimmt oder zumindest damit vertraut ist, leichter fallen.

6.3. Definition der zu messenden Fähigkeiten

Affordances

Es soll die Fähigkeiten erfasst werden, gemeinsame zentrale Funktionen von Gegenständen, Objekten, Personen, Einrichtungen etc. zu erkennen und gleichzeitig andere Gemeinsamkeiten, die sich nicht auf die Funktion beziehen, auszublenden. Der Testperson muss es gelingen, die gemeinsame Funktion vom Gesamthalt zu abstrahieren.

Constraints

Keinen Einfluss auf die Lösungswahrscheinlichkeit sollen Fachwissen in einem bestimmten Bereich und der Wortschatz einer Testperson haben. Die letztgenannte Störvariable ist nicht vollständig auszuschalten, da zum Lösen der Aufgaben ein Grundwortstock vorausgesetzt wird, es sollten aber bei der Erstellung der Items weder selten verwendete Wörter noch Fachjargon verwendet werden.

6.4. Geltungsbereich

Siehe 5.4.

6.5. Testformat

Siehe 5.5.

6.6. Antwortformat

Das ursprünglich freie Antwortformat des Untertests „Funktionen abstrahieren“ wurde in der Gruppentestversion durch ein Multiple-Choice-Format ersetzt, da es zum einen eine objektivere und ökonomischere Auswertung ermöglicht, zum anderen das Kind nicht gezwungen ist, seine Antworten selbst zu formulieren und so Schwierigkeiten bei der Rechtschreibung, im Ausdruck etc. keinen Einfluss auf das Testergebnis haben.

Beispiel:

Ampel

- | | |
|------------------|--------------------|
| a) Lampe | d) Bodenmarkierung |
| b) Auto | e) Verkehr |
| c) Straßenschild | |

Aufgabe des Kindes ist es pro Item zu entscheiden, welche zwei der fünf Begriffe dieselbe zentrale Funktion wie der Ausgangsbegriff (hier „Ampel“) erfüllen.

Durch das Antwortformat zwei aus fünf wird die Ratewahrscheinlichkeit relativ gering gehalten.

6.7. Itementwicklung

Radicals

Im Zuge der in 6.2. präsentierten Literaturrecherche und der Analyse der Originalitems in Bezug auf ihre Unterschiede in den verschiedenen Schwierigkeitsstufen stellte sich heraus, dass der Schwierigkeitsgrad der Items vor allem durch folgende vier Eigenschaften beeinflusst wird.

Nähe der Stimuli

Diese Komponente bezieht sich darauf, ob sich die pro Item genannten Stimuli abgesehen von der gemeinsamen Funktion auf Grundlage anderer Dimensionen nahe stehen, wie zum Beispiel, dass sie über eine ähnliche innere oder äußere Struktur verfügen, einen gemeinsamen Oberbegriff haben bzw. in dieselbe Kategorie fallen, im Alltag im selben Kontext auftreten oder eine räumliche Nähe aufweisen.

Dabei werden drei Ausprägungen dieser Komponente unterschieden, wobei Items mit der ersten Ausprägung die am einfachsten zu lösenden sein müssten und jene mit der dritten am schwersten.

1. Die drei Stimuli eines Items, welche dieselbe zentrale Funktion erfüllen, sind sich auch untereinander auf anderen Dimensionen ähnlicher als mit jenen drei Stimuli, welche nicht dieselbe zentrale Funktion erfüllen; sie gehören z.B. derselben Klasse von Objekten an, sind demselben Oberbegriff unterzuordnen, sind sich optisch ähnlich, treten im selben Kontext auf etc.
2. Zwischen den sechs genannten Stimuli pro Item gibt es auf anderen Dimensionen als jenen der Funktionalität keine nennenswerten Unterschiede bezüglich ihrer Ähnlichkeit.
3. Der als Ausgangsbegriff genannte Gegenstand weist mit zwei oder drei Gegenständen, welche nicht dieselbe zentrale Funktion erfüllen wie er, mehr Ähnlichkeiten in Bezug auf andere Dimensionen auf als mit den zwei Gegenständen, welche dieselbe Funktionalität wie er haben.

Erfahrung aus dem Alltag

Da, wie in der Literatur bereits erläutert, das Denken vor allem bei Kindern erfahrungsgeleitet ist, müssten gemeinsame Funktionen von Stimuli, welche im Alltag

des Kindes eine Rolle spielen, eher erkannt werden. Es wird zwischen drei Ausprägungen unterschieden:

1. Die Testperson hat mit hoher Wahrscheinlichkeit den Stimulus schon selbst benutzt; es ist sehr unwahrscheinlich, dass ihn die Testperson nicht kennt und ihm seine Funktion nicht bewusst ist.
2. Der Stimulus befindet sich mit hoher Wahrscheinlichkeit im unmittelbaren Lebensumfeld der Testperson und wird von deren Bezugspersonen (Eltern, Freunden, Lehrern,...) verwendet oder in Medien wie Fernsehen, Zeitung etc. sehr häufig präsentiert, diskutiert oder auch benutzt. Es ist gut möglich, dass das Kind den Stimulus selbst noch nicht verwendet hat, es hat aber mit hoher Wahrscheinlichkeit schon von ihm und seiner Funktion gehört.
3. Es ist möglich, dass der Gegenstand weder im Alltags- noch im Medienleben eine große Rolle spielt und seine Funktion für die Testperson daher nicht präsent ist.

Art der Funktion

Diese Komponente bezieht sich darauf, dass Kinder mit manchen Funktionen eher vertraut sind als mit anderen und, wie auch in der vorigen Komponente erwähnt, erhöht Vertrautheit aller Voraussicht nach die Lösungswahrscheinlichkeit.

Dabei werden Funktionen einer der folgenden Kategorien zugeordnet:

1. Erfüllung von Grundbedürfnissen (Nahrung, Wohnung, Bekleidung)
2. Physikalische Funktionen (Wärme, Licht, Energie etc.)
3. Transport, Verkehr
4. Information, Kommunikation
5. Soziale Funktion (z.B. Unterstützung von Randgruppen)
6. Unterhaltung
7. Ästhetik, Komfort, Kunst
8. Handel, Tausch, Kauf
9. Ordnung, Organisation, Schutz
10. Aufbewahrung, Behälter, Speicher, Stauraum, konservieren, verschließen/aufschließen,
11. Hygiene

Art des Stimulus

Während bei kleinen Gegenständen, Geräten und Apparaten die Funktion im Vordergrund steht, ist bei Lebewesen die äußere und innere Struktur zentral (Busemann, 1965; Keil, 1994). Dahingehend müsste die gemeinsame Funktion bei ersteren besser zu erkennen sein. Es wird unterschieden zwischen:

1. Stimuli, die ich aktiv verwende, in der Hand halten kann, gehören in der Regel einer Person bzw. Familie. (Bsp.: Werkzeug, Kleidung etc.)
2. Person/Lebewesen
3. Einrichtung, Institution, Allgemeingut
4. Verhaltensweisen
5. Abstrakte Begriffe

Incidentals

Eine inhaltliche Variation der Items ohne Einfluss auf ihr Schwierigkeitsniveau ist möglich, indem Stimuli, welche in ein- und dieselbe Gruppe fallen, gewählt werden (z.B. Einrichtungen, welche der Unterhaltung dienen und im Alltag häufig verwendet werden).

Richtlinien zum Erstellen von Items

Zur Erstellung eines Items ist zunächst ein „Ausgangsstimulus“ zu überlegen. Folgende Themenkreise können dabei als Denkanstoß dienen:

- Lebensmittel
- Hygieneartikel
- Elektrische Geräte
- Öffentliche oder soziale Einrichtungen
- Kommunikationsmittel
- Natur (z.B. Tiere, Pflanzen...)
- Haushaltsgeräte
- Verkehr (Verkehrsmittel, Hilfsmittel zur Regelung des Verkehrs,...)
- Studium- und Schulbereich
- Berufe
- Gefühle
- Unterhaltung, Kommunikationstechniken

- Freizeit, Sportarten
- Physik

Im Anschluss sind zwei Stimuli zu überlegen, welche dieselbe grundlegende Funktion wie der Ausgangsstimulus erfüllen. Je nachdem, wie schwierig das Item gestaltet werden soll, können sich diese drei Stimuli in anderen Dimensionen ebenfalls nahe stehen oder aber sehr unterschiedlich voneinander sein.

Schließlich sind drei Stimuli zu wählen, welche bezüglich ihrer Funktionalität keine Gemeinsamkeit mit dem Ausgangsstimulus aufweisen. Je schwieriger das Item sein soll, umso ähnlicher sollen diese drei Stimuli aber dem Ausgangsstimulus auf anderen Dimensionen z.B. der äußeren Erscheinung sein.

Auch durch Verwendung der verschiedenen Stufen der restlichen *Radicals* kann die Schwierigkeit der Items variiert werden.

Die Ausprägungen der vier *Radicals* sind mehr oder weniger stark vom subjektiven Ermessen des/der Testentwicklers/in abhängig. Bei jeder könnte es aber zu Streitfällen kommen.

Daher sollte auch bei diesem Untertest die Kodierung von zumindest zwei TestentwicklerInnen unabhängig voneinander durchgeführt werden und etwaige Differenzen in der Einschätzung besprochen werden.

Insgesamt wurden 60 Items entwickelt, die im Anhang angeführt sind.

6.8. Testinstruktion

Siehe 5.8.

6.9. Wahl der Analytestichprobe

Siehe 5.9.

6.10. Testvorgabe

Siehe 5.10.

6.11. Aufgabenbewertung

Die Aufgabenbewertung wurde so einfach wie möglich gestaltet. So wird nur zwischen richtiger und falscher Antwort unterschieden, eine Vergabe von Teilpunkten fand nicht statt. Eine Antwort galt nur dann als richtig, wenn die Testperson beide Stimuli, welche dieselbe zentrale Funktion wie der Ausgangsstimulus erfüllen, als solche erkannt hatte. Alle anderen Antwortkombinationen sind als falsch zu werten.

6.12. Aufgabenanalyse

Siehe 5.12.

7. Beobachtungen zur Testung

Die Testung verlief in allen Klassen einwandfrei und ohne nennenswerte Vorfälle. Selbst in den unteren Schulstufen wurde die Instruktion verstanden und die Übungsbeispiele gelöst. Das bemessene Zeitintervall von 50 min. war für nahezu alle Kinder ausreichend, um jeden Untertest vollständig zu bearbeiten. Testpersonen, die wesentlich früher als ihre KollegInnen fertig geworden sind, wurde, um sie zu beschäftigen, eine andere Testform des Untertests „Alltagswissen“ vorgelegt.

Die im Anschluss der Testung ausgeteilte Süßigkeit als kleines Dankeschön wurde von den Kindern sehr positiv wahrgenommen.

8. Deskriptive Statistik

Die Daten wurden hinsichtlich ihrer Verteilung in den Variablen „Geschlecht“, „Region“, „Schultyp“ und „Schulstufe“ überprüft, wobei dies zum einen für die gesamten Untertests, zum anderen für die einzelnen Testformen erfolgte.

Die Verteilungen „Funktionen Abstrahieren“ und „Wortschatz“ in den Untertests waren naturgemäß deckungsgleich, da jede Testperson beide Untertests erhalten hatte. Ein kleiner Unterschied in der Verteilung resultierte jedoch daraus, dass beim Untertest „Funktionen Abstrahieren“ drei Testbögen nicht abgegeben wurden (ob von dem/der SchülerIn beabsichtigt sei dahingestellt); diese wurden als fehlende Werte kodiert.

Bezüglich der Testformen wurde eine möglichst gleichmäßige Vorgabe angestrebt. In der Regel erhielt eine Testperson in den verschiedenen Untertests dieselbe Testform; selten kam es aus verschiedenen Gründen (z.B. um die Häufigkeit einer bisher seltener vorgegeben Testform zu erhöhen) zu einem Wechsel, keinesfalls aber besaßen Sitznachbarn dieselbe Testform.

8.1. Geschlecht

In den Tabellen 1 und 2 ist ersichtlich, dass die Variable Geschlecht sowohl in den beiden Untertests als auch in den verschiedenen Testformen annähernd gleich verteilt ist, wobei insgesamt etwas mehr Jungen (52,1 %) als Mädchen (47,9%) getestet wurden.

Tab. 1: Verteilung Geschlecht pro Untertest

Geschlecht	Häufigkeit		Prozent	
	Funktionen abstrahieren	Wortschatz	Funktionen Abstrahieren	Wortschatz
Männlich	354	354	52,1%	52,1%
Weiblich	322	325	47,4%	47,9%
Fehlend	3	0	0,4%	0,0%
Gesamt	679	679	100	100
Gesamtgültig	676	679	99,6	100

Tab. 2: Verteilung Geschlecht pro Testform

Geschlecht	Funktionen Abstrahieren				Wortschatz			
	A	B	C	D	A	B	C	D
Männlich	83	90	98	83	88	90	92	84
Weiblich	89	81	70	82	86	79	76	84
Fehlend	0	0	0	3	0	0	0	0
Gesamt	172	171	168	168	174	169	168	168
Gesamtgültig	172	171	168	165	174	169	168	168

8.2. Region

Auch im Bezug auf die Variable Region kann von einer Gleichverteilung der Daten gesprochen werden, wobei geringfügig mehr Kinder aus der Stadt getestet wurden (Tab. 3). Die größte Differenz zwischen Stadt und Land lag mit 10 SchülerInnen beim Untertest „Wortschatz“ Testform C vor (Tab. 4).

Tab. 3: Verteilung Region pro Untertest

Region	Häufigkeit		Prozent	
	Funktionen abstrahieren	Wortschatz	Funktionen Abstrahieren	Wortschatz
Land	323	323	47,6%	47,6%
Stadt	353	356	52%	52,4%
Fehlend	3	0	0,4%	0,0%
Gesamt	679	679	100	100
Gesamtgültig	676	679	99,6	100

Tab. 4: Verteilung Region pro Testform

Region	Funktionen Abstrahieren				Wortschatz			
	A	B	C	D	A	B	C	D
Land	82	82	80	79	83	81	79	80
Stadt	90	89	88	86	91	88	89	88
Fehlend	0	0	0	3	0	0	0	0
Gesamt	172	171	168	168	174	169	168	168
Gesamtgültig	172	171	168	165	174	169	168	168

8.3. Schultyp

Die Verteilung der Variable „Schultyp“ ähnelte jener der Variable „Region“. Grund dafür ist, dass in ländlichen Regionen ausschließlich Volks- und Hauptschulen getestet wurden, während in der Stadt die Testung in einem Gymnasium erfolgte.

Anhand der Tabellen 5 und 6 ist ersichtlich, dass etwas mehr GymnasiastInnen als HauptschülerInnen getestet wurden. Wesentlich geringer ist hingegen der Anteil der VolksschülerInnen mit insgesamt 35 Personen.

Tab. 5: Verteilung Schultyp pro Untertest

Schultyp	Häufigkeit		Prozent	
	Funktionen abstrahieren	Wortschatz	Funktionen Abstrahieren	Wortschatz
Volksschule	35	35	5,2	5,2
Hauptschule	288	288	42,4	42,4
Gymnasium	353	356	52,0	52,4
Fehlend	3	0	0,4	0
Gesamt	679	679	100	100
Gesamtgültig	676	679	99,6	100

Tab. 6: Verteilung Schultyp pro Testform

Schultyp	Funktionen Abstrahieren				Wortschatz			
	A	B	C	D	A	B	C	D
Volksschule	9	8	9	9	10	8	8	9
Hauptschule	73	74	71	70	73	73	71	71
Gymnasium	90	89	88	86	91	88	89	88
Fehlend	0	0	0	3	0	0	0	0
Gesamt	172	171	168	168	174	169	168	168
Gesamtgültig	172	171	168	165	174	169	168	168

8.4. Schulstufe

Aus den folgenden Tabellen (Tab. 7 & 8) ist abzulesen, dass in der 5., 6., 7. und 8. Schulstufe annähernd gleich viele Personen getestet wurden. In der 3. und 4. Schulstufe – den Volksschulklassen - sowie in der 9. Schulstufe war die Anzahl an Testpersonen wesentlich geringer. Die Volksschule und 9. Schulstufe wurden aufgrund der zu kleinen Stichprobe in die Mittelwertvergleiche nicht miteinbezogen. Die Anzahl der Testpersonen reicht aber aus um festzustellen, dass die Vorgabe der Untertests „Wortschatz“ und „Funktionen Abstrahieren“ auch in diesem Altersbereich möglich und zumutbar ist.

Tab. 7: Verteilung Schulstufe pro Untertest

Schulstufe	Häufigkeit		Prozent	
	Funktionen abstrahieren	Wortschatz	Funktionen Abstrahieren	Wortschatz
3.	13	13	1,9	1,9
4.	22	22	3,2	3,2
5.	140	140	20,6	20,6
6.	160	162	23,6	23,9
7.	150	150	22,1	22,1
8.	139	140	20,5	20,6
9.	52	52	7,7	7,7
Fehlend	3	0	0,4	0
Gesamt	679	679	100	100
Gesamtgültig	676	679	99,6	100

Tab. 8: Verteilung Schulstufe pro Testform

Schulstufe	Funktionen Abstrahieren				Wortschatz			
	A	B	C	D	A	B	C	D
3.	3	3	3	4	4	3	3	3
4.	6	5	6	5	6	5	5	6
5.	34	36	35	35	35	35	35	35
6.	40	40	41	39	41	39	41	41
7.	40	39	34	37	39	38	35	38
8.	36	35	36	32	36	36	36	32
9.	13	13	13	13	13	13	13	13
Fehlend	0	0	0	3	0		0	0
Gesamt	172	171	168	168	174	169	168	168
Gesamtgültig	172	171	168	165	174	169	168	168

9. Statistische Berechnungen

Die Konformität der acht vorliegenden Itempools mit dem dichotom logistischen Modell von Rasch wurde mittels des bedingten Likelihood-Quotienten-Tests (LQT) nach Andersen (1973) überprüft und anhand der grafischen Modelltests veranschaulicht. Eine detaillierte Auflistung der Item- und Personenparameter findet sich im Anhang der Diplomarbeit.

Die Parameterschätzungen und Modelltests erfolgten dabei mit Hilfe des Programms R 2.10.0. Für die im Anschluss daran durchgeführten Mittelwertvergleiche wurde das Programmpaket PASW Statistics 17 verwendet.

9.1. Überprüfung der Daten auf Rasch-Modell-Konformität

Die forschungsleitende Hypothese ist in diesem Fall die H_0 :

„Die Daten verhalten sich konform dem dichotomen logistischen Modell von Rasch.“

Um nicht zu konservativ zu testen wurde die Wahrscheinlichkeit einen Fehler 1. Art zu begehen mit $\alpha = 0,05$ festgesetzt. Da jedoch die Überprüfung der H_0 an einem Datensatz durch mehrere voneinander nicht unabhängigen Tests (LQTs) erfolgte, musste, damit der Fehler 1. Art nicht zu wahrscheinlich wurde, eine α Adjustierung stattfinden.

Teilungskriterien

Die Modellgültigkeit wurde für jeden der acht vorgegebenen Tests anhand folgender drei Teilungskriterien überprüft:

Intern:

- Höhe des Rohwertes (Personen mit einem Rohwert niedriger oder gleich dem Median vs. Personen mit einem Rohwert höher dem Median).

Extern:

- Geschlecht (männlich vs. Weiblich)
- Region (Stadt vs. Land)

Es wurden somit drei abhängige Modelltests pro Testform durchgeführt, dies ergab nach Durchführung der Adjustierung ein $\alpha = 0,013$.

9.2. Likelihood-Quotienten-Test der ersten Schätzung

Wortschatz Testform A

Für den Itempool der Testform A musste die H0 „Das Rasch-Modell gilt“ verworfen werden, da im Teilungskriterium Geschlecht auf dem 1% Niveau ein signifikantes Ergebnis auftrat. Der Itempool umfasste beim Likelihood-Quotienten-Test mit dem Teilungskriterium Geschlecht bzw. Region 20 Items, beim internen Teilungskriterium nur 19 Items, da das Item A19 wegen ungünstigen Antwortmustern in den Subgruppen von den Berechnungen ausgeschlossen wurde. Eine Übersicht sowie Veranschaulichung der Resultate zeigt die folgende Tabelle (Tab. 9):

Tab. 9: Erste Parameterschätzung, Test Wortschatz A, nicht Rasch-Modell-konform

Andersen LQT	Intern	Kritischer Wert df=18	Geschlecht	Region	Kritischer Wert df=19
LQ-Wert	32,815	33,89	43,505	30,912	35,26
Chi-square df	18	18	19	19	19
p-Wert	0,018	0,013	0,001	0,041	0,013

Wortschatz Testform B

Von der ursprünglichen Testform B wurde a priori Item B09 (s. Anhang) ausgeschlossen, da sich im Laufe der Testung herausgestellt hatte, dass dieses nicht eindeutig zu lösen ist. So löst eine Massage nicht nur Verspannungen sondern kann im weitem Sinne auch Schadstoffe im Körper lösen. Zudem ist die Formulierung „mechanische Beeinflussung“ unglücklich gewählt.

Für den Itempool der Testform B, welcher 18 Items umfasste, konnte die H0 „Das Rasch-Modell gilt“ beibehalten werden, da die Andersen Likelihood-Quotienten-Tests auf dem 1% Niveau nicht signifikant wurden.

Eine Übersicht sowie Veranschaulichung der Resultate zeigen die folgende Tabelle (Tab. 10) und Grafiken (Abb.1 – Abb. 3):

Tab. 10: Erste Parameterschätzung, Test Wortschatz B, Rasch-Modell-konform

Andersen LQT	Intern	Geschlecht	Region	Kritischer Wert df=17
LQ-Wert	24,235	29,699	26,312	32,51
Chi-square df	17	17	17	17
p-Wert	0,113	0,029	0,069	0,013

Abb.1: Grafischer Modelltest „Wortschatz“ Testform B, Teilungskriterium „Höhe des Rohwerts“

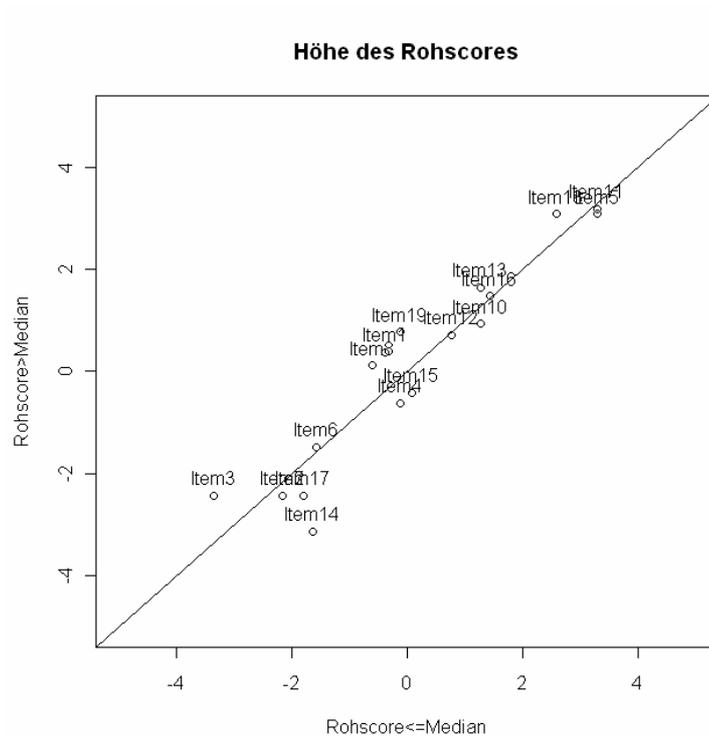


Abb.2: Grafischer Modelltest „Wortschatz“ Testform B, Teilungskriterium „Geschlecht“

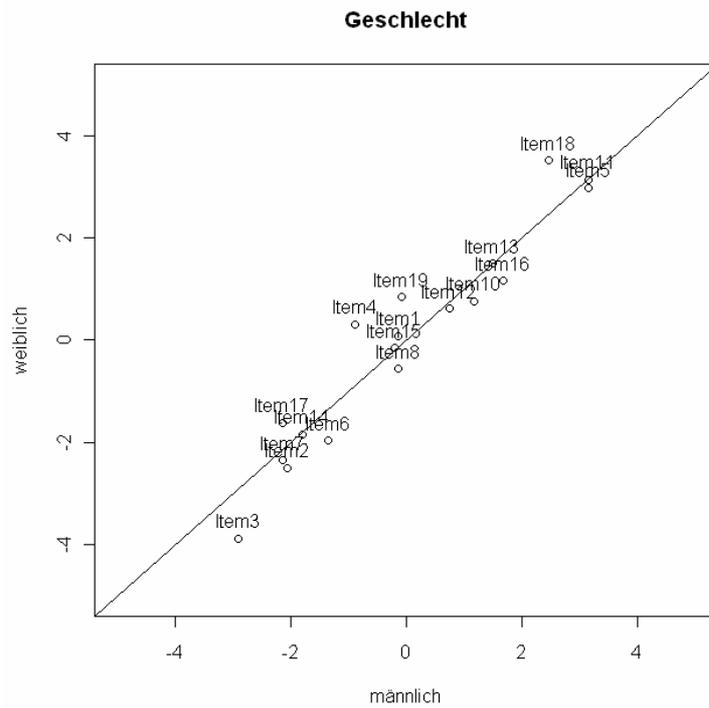
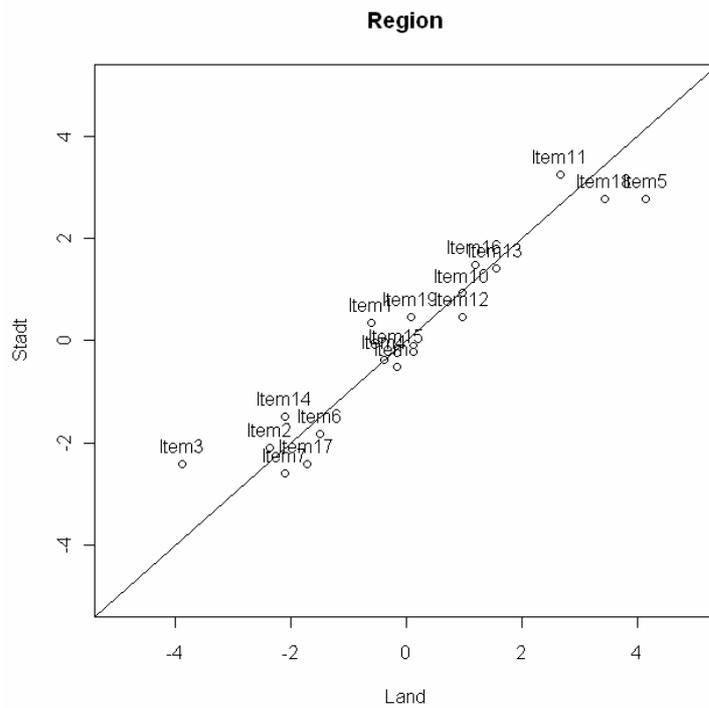


Abb.3: Grafischer Modelltest „Wortschatz“ Testform B, Teilungskriterium „Region“



Wortschatz Testform C

Für den Itempool der Testform C, welcher 20 Items umfasste, musste die H0 „Das Rasch-Modell gilt“ verworfen werden, da im Teilungskriterium Intern und Region auf dem 1% Niveau ein signifikantes Ergebnis auftrat. Eine Übersicht der Resultate zeigt die folgende Tabelle (Tab. 11):

Tab. 11: Erste Parameterschätzung, Test Wortschatz C, nicht Rasch-Modell-konform

Andersen LQT	Intern	Geschlecht	Region	Kritischer Wert df=19
LQ-Wert	37,731	30,375	37,298	35,26
Chi-square df	19	19	19	19
p-Wert	0,006	0,047	0,007	0,013

Wortschatz Testform D

Für den Itempool der Testform D konnte die H0 „Das Rasch-Modell gilt“ beibehalten werden, da die Andersen Likelihood-Quotienten-Tests auf dem 1% Niveau nicht signifikant wurden.

Der Itempool umfasste beim Likelihood-Quotienten-Test mit dem Teilungskriterium Geschlecht bzw. Region 19 Items. Beim internen Teilungskriterium beinhaltet dieser nur 17 Items, da die Items D07 und D11 wegen ungünstigen Antwortmustern in den Subgruppen von den Berechnungen ausgeschlossen wurden. Eine Übersicht sowie Veranschaulichung der Resultate zeigen die folgende Tabelle (Tab. 12) und Grafiken (Abb. 4- Abb. 6):

Tab. 12: Erste Parameterschätzung, Test Wortschatz D, Rasch-Modell-konform

Andersen LQT	Intern	Kritischer Wert df=16	Geschlecht	Region	Kritischer Wert df=18
LQ-Wert	28,363	31,12	30,377	26,636	33,89
Chi-square df	16	16	18	18	18
p-Wert	0,029	0,013	0,034	0,086	0,013

Abb.4: Grafischer Modelltest „Wortschatz“ Testform D, Teilungskriterium „Höhe des Rohwerts“

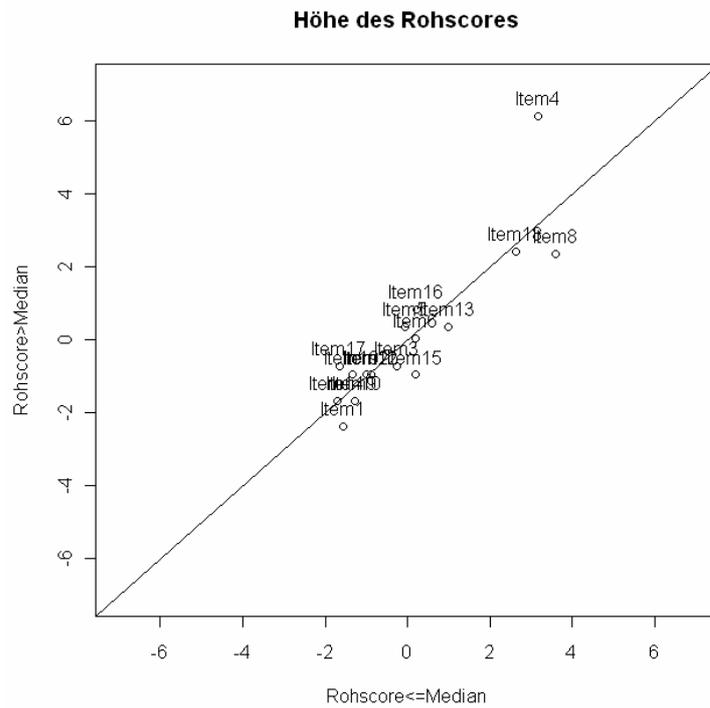


Abb.5: Grafischer Modelltest „Wortschatz“ Testform D, Teilungskriterium „Geschlecht“

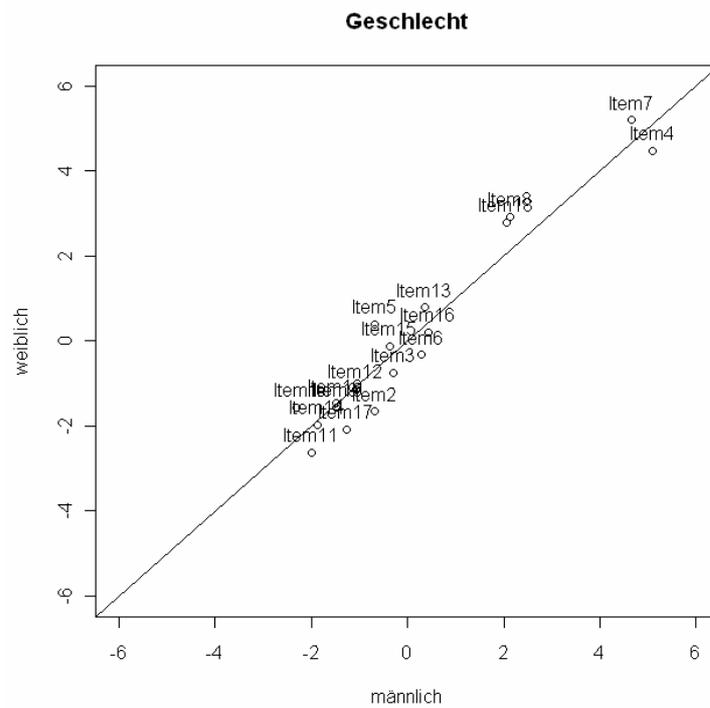
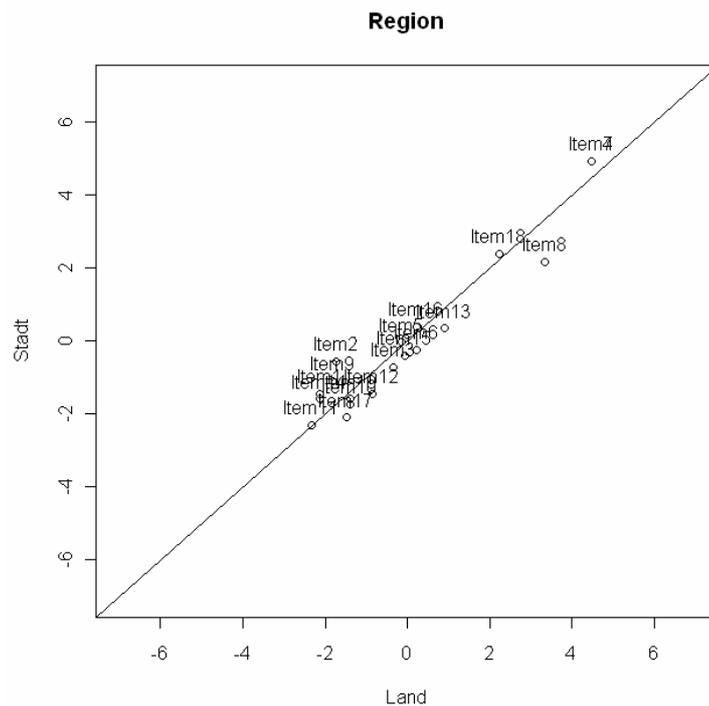


Abb.6: Grafischer Modelltest „Wortschatz“ Testform D, Teilungskriterium „Region“



Funktionen abstrahieren Testform A

Für den Itempool der Testform A musste die H_0 „Das Rasch-Modell gilt“ verworfen werden, da in den Teilungskriterien Intern und Geschlecht auf dem 1% Niveau ein signifikantes Ergebnis auftrat.

Der Itempool umfasste beim Likelihood-Quotienten-Test mit dem Teilungskriterium Region 15 Items, bei den Teilungskriterien Intern und Geschlecht nur 14 Items, da das Item E01 wegen ungünstigen Antwortmustern in den Subgruppen von den Berechnungen ausgeschlossen wurde.

Eine Übersicht der Resultate zeigt die folgende Tabelle (Tab. 13).

Tab. 13: Erste Parameterschätzung, Test Funktionen Abstrahieren Testform A, nicht Rasch-Modell-konform

Andersen LQT	Intern	Geschlecht	Kritischer Wert df=13	Region	Kritischer Wert df=14
LQ-Wert	27,652	38,382	26,87	22,234	28,29
Chi-square df	13	13	13	14	14
p-Wert	0,01	0	0,013	0,074	0,013

Untertest Funktionen Abstrahieren Testform B

Für den Itempool der Testform B, welcher 15 Items umfasste, musste die H0 „Das Rasch-Modell gilt“ verworfen werden, da im Teilungskriterium Intern auf dem 1% Niveau ein signifikantes Ergebnis auftrat.

Eine Übersicht der Resultate zeigt die folgende Tabelle (Tab. 14):

Tab. 14: Erste Parameterschätzung, Test Funktionen Abstrahieren Testform B, nicht Rasch-Modell-konform

Andersen LQT	Intern	Geschlecht	Region	Kritischer Wert df=14
LQ-Wert	62,358	20,699	16,37	28,29
Chi-square df	14	14	14	14
p-Wert	0	0,11	0,291	0,013

Untertest Funktionen Abstrahieren Testform C

Für den Itempool der Testform C konnte die H0 „Das Rasch-Modell gilt“ beibehalten werden, da die Andersen Likelihood-Quotienten-Tests auf dem 1% Niveau nicht signifikant waren. Der Itempool umfasste beim Likelihood-Quotienten-Test mit dem Teilungskriterium Geschlecht bzw. Region 15 Items, beim internen Teilungskriterium nur 14 Items, da das Item G03 wegen ungünstigen Antwortmustern in den Subgruppen von den Berechnungen ausgeschlossen wurde.

Eine Übersicht sowie Veranschaulichung der Resultate zeigen die folgende Tabelle (Tab. 15) und Grafiken (Abb. 7-9):

Tab. 15: Erste Parameterschätzung, Test Funktionen Abstrahieren Testform C, Rasch-Modell-konform

Andersen LQT	Intern	Kritischer Wert df=13	Geschlecht	Region	Kritischer Wert df=14
LQ-Wert	9,874	26,87	17,463	17,394	28,29
Chi-square df	13	13	14	14	14
p-Wert	0,704	0,013	0,232	0,236	0,013

Abb.7: Grafischer Modelltest „Funktionen Abstrahieren“ Testform C, Teilungskriterium „Höhe des Rohwerts“

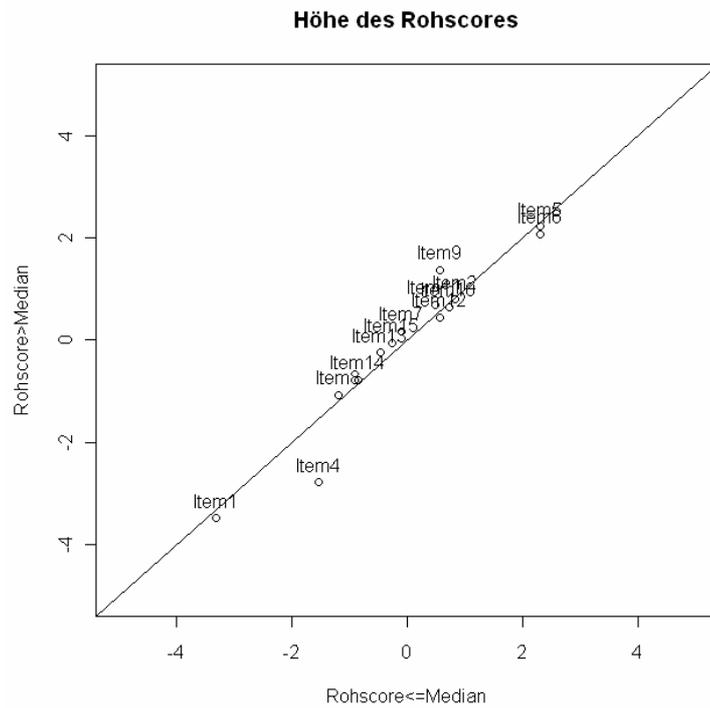


Abb.8: Grafischer Modelltest „Funktionen Abstrahieren“ Testform C, Teilungskriterium „Geschlecht“

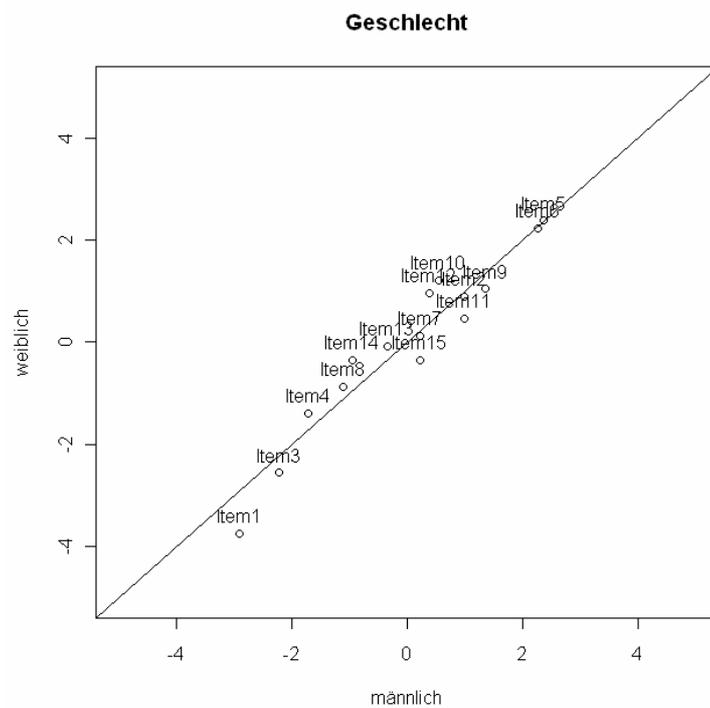
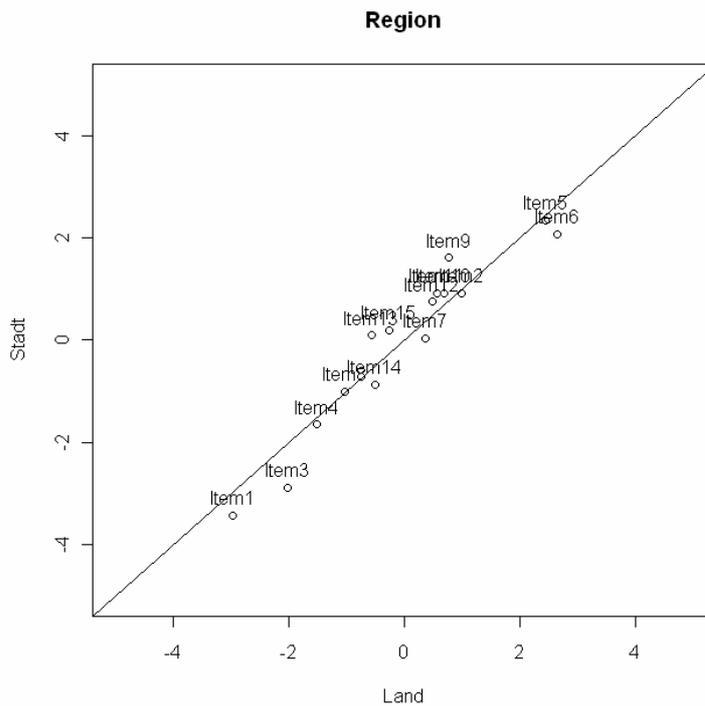


Abb.9: Grafischer Modelltest „Funktionen Abstrahieren“ Testform C, Teilungskriterium „Region“



Untertest Funktionen Abstrahieren Testform D

Für den Itempool der Testform D musste die H0 „Das Rasch-Modell gilt“ verworfen werden, da im Teilungskriterium Intern und Geschlecht auf dem 1% Niveau ein signifikantes Ergebnis auftrat.

Der Itempool umfasste beim Likelihood-Quotienten-Test mit dem Teilungskriterium Geschlecht bzw. Region 15 Items, beim internen Teilungskriterium nur 14 Items, da das Item H01 wegen ungünstigen Antwortmustern in den Subgruppen von den Berechnungen ausgeschlossen wurde.

Eine Übersicht der Resultate zeigt die folgende Tabelle (Tab. 16):

Tab. 16: Erste Parameterschätzung, Test Funktionen Abstrahieren Testform D, nicht Rasch-Modell-konform

Andersen LQT	Intern	Kritischer Wert df=13	Geschlecht	Region	Kritischer Wert df=14
LQ-Wert	47,903	26,86	30,27	16,149	28,29
Chi-square df	13	13	14	14	14
p-Wert	0	0,013	0,007	0,304	0,013

Conclusio der Likelihood-Quotienten-Tests der ersten Schätzung

Die Likelihood-Quotienten-Tests ergaben, dass für die Tests „Wortschatz“ Testform B und D sowie „Funktionen abstrahieren“ Testform C die H_0 „Das Rasch-Modell gilt“ beibehalten werden konnte.

Bei den übrigen Tests („Wortschatz“ Testform A und C, „Funktionen abstrahieren“ Testform A, B und D) trat bei zumindest einem Teilungskriterium ein signifikantes Ergebnis auf. Die Items dieser Tests verhielten sich somit nicht Rasch-Modell-konform.

9.3. Parameterschätzungen und Modelltests nach Itemselektion

Ausscheiden nicht Rasch-Modell-konformer Items

Um zu entscheiden welche Items aus dem Pool der oben genannten nicht Rasch-Modell-konformen Testformen entfernt werden müssen, wurden der graphische Modelltest, die Itemfitstatistik und der Z-Test herangezogen. Diese ermöglichen im Gegensatz zum Likelihood-Quotienten-Test eine itemweise Überprüfung des Pools (Kubinger, 1989). Die Itemselektion wurde sukzessive vorgenommen; d.h. ein Item wurde ausgeschieden und der Likelihood-Quotienten-Test für alle Kriterien neu berechnet.

Im Anschluss erfolgt für jede der fünf betroffenen Testformen eine Darstellung der graphischen Modelltests, der Itemfitstatistiken nach Hatzinger und Mair sowie der Ergebnisse im Z-Test, zu dem findet sich eine qualitative Analyse der auszuscheidenden Items und eine Übersicht der abermaligen Überprüfung des bereinigten Itempools auf seine Rasch-Modell-Konformität anhand des LQT.

Wortschatz Testform A

Z-Test

Beim Z-Test (Tab. 17) mit dem Teilungskriterium „Intern“ wurden die Items A05 ($p=0,008$), A07 ($p=0,031$) und A18 ($p=0,046$) signifikant. Beim Teilungskriterium „Geschlecht“ unterschritten die p -Werte der Items A03, A07, A15 und A16 den Wert von 0,05. Die beiden letztgenannten Items (A15 und A16) zeigten auch beim Teilungskriterium „Region“ signifikante p -Werte.

Das Item A19 welches mit dem Teilungskriterium „Intern“ nicht geschätzt werden konnte, verhielt sich im Z-Test sowohl beim Teilungskriterium „Geschlecht“ ($p=0,524$) als auch beim Teilungskriterium „Region“ ($p=0,915$) Rasch-Modell-konform.

Tab. 17: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Wortschatz Testform A

Item	z-Wert INTERN	p-Wert INTERN	z-Wert GESCHL.	p-Wert GESCHL.	z-Wert REGION	p-Wert REGION
ItemA01	1,401	0,161	-1,795	0,073	0,028	0,978
ItemA02	1,770	0,077	-1,674	0,094	0,653	0,514
ItemA03	0,109	0,914	2,606	0,009	-0,891	0,373
ItemA04	0,775	0,438	-0,017	0,986	1,697	0,090
ItemA05	2,640	0,008	-0,434	0,664	1,437	0,151
ItemA06	0,973	0,331	-0,187	0,852	-0,323	0,746
ItemA07	2,161	0,031	3,942	0,000	1,614	0,106
ItemA08	-0,931	0,352	0,411	0,681	0,651	0,515
ItemA09	-0,900	0,368	0,407	0,684	-1,100	0,271
ItemA10	0,227	0,821	0,030	0,976	1,458	0,145
ItemA11	-0,286	0,775	-0,061	0,951	-0,430	0,667
ItemA12	-1,136	0,256	0,007	0,994	1,156	0,248
ItemA13	1,114	0,265	0,076	0,940	0,333	0,739
ItemA14	-0,977	0,328	0,694	0,487	1,013	0,311
ItemA15	-1,842	0,065	-2,497	0,013	-2,652	0,008
ItemA16	-1,464	0,143	-2,043	0,041	-2,446	0,014
ItemA17	-0,884	0,377	0,643	0,520	-0,278	0,781
ItemA18	-1,997	0,046	0,108	0,914	-1,199	0,231
ItemA19	-	-	0,638	0,524	0,107	0,915
ItemA20	-0,178	0,859	-1,696	0,090	-1,165	0,244

Itemfitstatistik

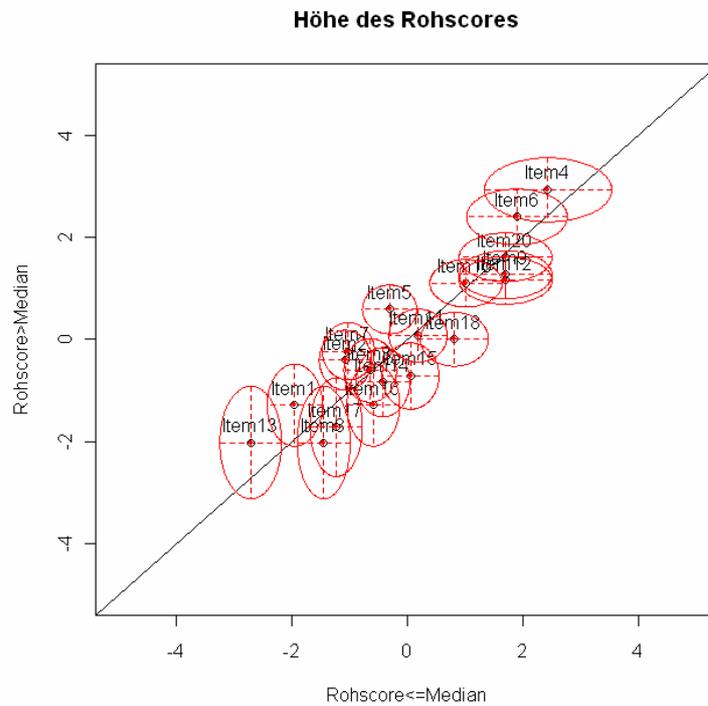
Tab. 18: Itemfitstatistik Wortschatz Testform A

Item	Chisquare	Df	p-Wert
ItemA01	157,359	167	0,692
ItemA02	301,180	167	0,000
ItemA03	136,020	167	0,962
ItemA04	213,658	167	0,009
ItemA05	305,376	167	0,000
ItemA06	234,015	167	0,000
ItemA07	214,234	167	0,008
ItemA08	144,189	167	0,899
ItemA09	170,858	167	0,403
ItemA10	153,233	167	0,770
ItemA11	119,819	167	0,998
ItemA12	133,564	167	0,973
ItemA13	164,894	167	0,532
ItemA14	137,071	167	0,956
ItemA15	106,742	167	1,000
ItemA16	119,271	167	0,998
ItemA17	140,934	167	0,929
ItemA18	112,064	167	1,000
ItemA19	47,952	167	1,000
ItemA20	197,620	167	0,053

In der Itemfitstatistik (Tab. 18) wiesen die Item A02, A04, A05, A06 und A07 signifikante p -Werte auf.

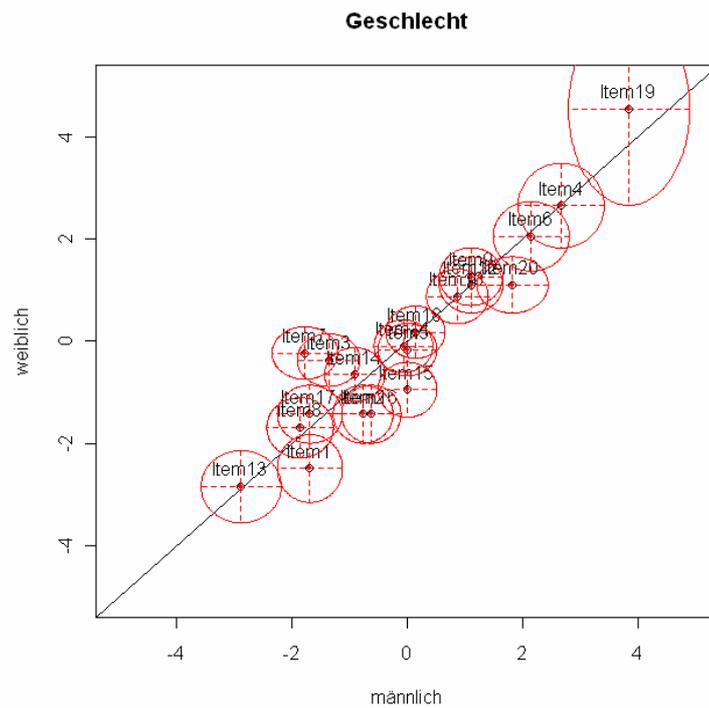
Grafische Modelltests

Abb.10: Grafischer Modelltest „Wortschatz“ Testform A, TK „Höhe des Rohwerts“



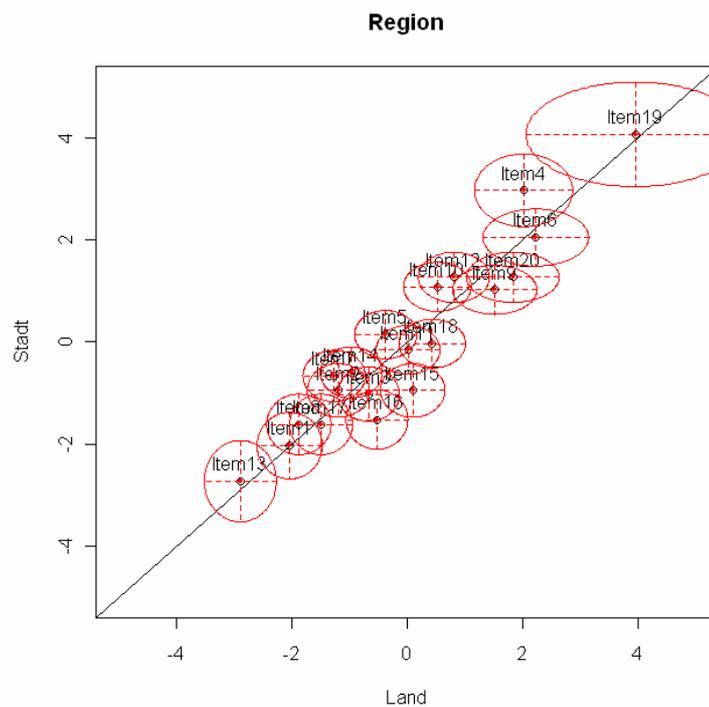
Der grafische Modelltest mit dem Teilungskriterium “Rohwert” (Abb. 10) zeigt, dass die Konfidenz-Ellipsen (diese repräsentieren grafisch die Konfidenzintervalle der Itemparameter in den beiden Subgruppen bei einem $p=0,95$) der Items 5, 7 und 18 (entspricht Items A05, A07 und A18) die Gerade nicht schneiden.

Abb.11: Grafischer Modelltest „Wortschatz“ Testform A, TK „Geschlecht“



Der grafische Modelltest mit dem Teilungskriterium “Geschlecht” (Abb. 11) veranschaulichte, dass die Konfidenz-Ellipsen der Items 3, 7, 15 und 16 (entspricht Items A03, A07, A15 und A16) die Gerade nicht schneiden.

Abb.12: Grafischer Modelltest „Wortschatz“ Testform A, TK „Region“



Der grafische Modelltest mit dem Teilungskriterium "Region" (Abb. 12) zeigte, dass die Konfidenz-Ellipsen der Items 15 und 16 (entspricht Items A15 und A16) die Gerade nicht schneiden.

Qualitative Analyse

Item A05:

A05) Artist	
Ein Artist ist ein	a) <i>Adeliger</i> der zumeist in/im e) <i>Schulen</i> auftritt.
	b) <i>Philosoph</i> f) <i>Universitäten</i>
	c) <i>Schausteller</i> g) <i>Zirkus</i>
	d) <i>Schauspieler</i> h) <i>Restaurants</i>

Veralteter Begriff: Das Wort „Schausteller“ welches als richtige Antwort galt, wird in der heutigen Zeit kaum mehr verwendet.

Item A07:

A07) Helikopter	
Ein Helikopter ist ein/e	a) <i>Ballon</i> , das/der/die durch eine/n motorisch
	b) <i>Einsatzwagen</i>
	c) <i>Unterwasserboot</i>
	d) <i>Luftfahrzeug</i>
betriebene	e) <i>Reifenachse</i> angetrieben wird.
	f) <i>Rotor</i>
	g) <i>Heliumspender</i>
	h) <i>Schraube</i>

Einfluss von Fachwissen: Das Item A07 birgt die Gefahr, nicht nur die Dimension „Wortschatz“ zu messen, sondern auch Fachwissen im technischen Bereich. So wussten vermutlich einige Testpersonen, um was es sich bei einem Helikopter handelt – er befindet sich damit in ihrem Wortschatz- jedoch nicht ob dieser mit einem Rotor, einer Reifenachse, einem Heliumspender oder einer Schraube betrieben wird.

Zusammenfassende Schlussfolgerung

Die Items A05 und A07 erwiesen sich in den Z-Tests und grafischen Modelltests mit dem Teilungskriterium „Rohwert“ sowie in der Itemfitstatistik als nicht Rasch-Modell-

konform. Auch in der qualitativen Analyse konnten Schwachstellen dieser Items aufgedeckt werden. Der zweite LQT verlief demnach ohne den Items A05 und A07.

Likelihood-Quotienten-Test der vierten Schätzung

Nachdem die Likelihood-Quotienten-Tests bei getrennten Ausscheiden der Items A05 bzw. A07 in zumindest einem der drei Teilungskriterien signifikante Ergebnisse aufwiesen, wurden in der vierten Parameterschätzung mitsamt Modelltests beide Items gleichzeitig aus dem ursprünglichen Itempool genommen.

Für den Itempool „Wortschatz Testform A“ konnte die H_0 „Das Rasch-Modell gilt“ nach Ausschluss der Items A05 und A07 a posteriori beibehalten werden, da die Andersen Likelihood-Quotienten-Tests auf dem 1% Niveau nicht signifikant wurden. Der Itempool umfasste beim Likelihood-Quotienten-Test mit dem Teilungskriterien Geschlecht und Region 18 Items, bei dem Teilungskriterium Intern nur 17 Items, da das Item A19 wegen ungünstigen Antwortmustern in den Subgruppen von den Berechnungen ausgeschlossen wurde.

Eine Übersicht der Resultate zeigt die folgende Tabelle (Tab. 19):

Tab. 19: Vierte Parameterschätzung, Test Wortschatz A, Rasch-Modell-konform

Andersen LQT	Intern	Kritischer Wert df=16	Geschlecht	Region	Kritischer Wert df=17
LQ-Wert	26,94	31,12	27,14	26,27	32,51
Chi-square df	16	16	17	17	17
p-Wert	0,042	0,013	0,056	0,07	0,013

Wortschatz Testform C

Z-Test

Der Z-Test (Tab. 20) wies mit dem Teilungskriterium „Intern“ bei den Items C02 ($p=0,00$) und C18 ($p=0,012$) einen p -Wert kleiner dem kritischen Wert von $p=0,05$ auf. Beim Teilungskriterium „Geschlecht“ unterschritten die Items C01 und C15 den p -Wert von 0,05. Die Items C02 und C20 erwiesen sich beim Teilungskriterium „Region“ als signifikant.

Tab. 20: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Wortschatz
Testform C

Item	z-Wert INTERN	p-Wert INTERN	z-Wert GESCHL.	p-Wert GESCHL.	z-Wert REGION	p-Wert REGION
ItemC01	-0,194	0,846	-3,260	0,001	0,925	0,355
ItemC02	4,417	0,000	-0,398	0,691	2,893	0,004
ItemC03	0,881	0,378	0,311	0,755	-1,017	0,309
ItemC04	-1,157	0,247	1,593	0,111	0,556	0,578
ItemC05	-1,840	0,066	-0,162	0,872	-1,923	0,055
ItemC06	0,937	0,349	-0,848	0,397	1,935	0,053
ItemC07	0,138	0,890	-1,562	0,118	-0,624	0,533
ItemC08	-1,129	0,259	-0,198	0,843	-1,350	0,177
ItemC09	0,140	0,888	0,427	0,669	-0,735	0,462
ItemC10	0,726	0,468	1,802	0,072	1,498	0,134
ItemC11	0,574	0,566	0,649	0,516	-1,443	0,149
ItemC12	1,259	0,208	-0,993	0,321	-1,138	0,255
ItemC13	1,380	0,168	-0,100	0,920	1,878	0,060
ItemC14	-0,673	0,501	-0,317	0,751	-0,391	0,696
ItemC15	-1,511	0,131	2,115	0,034	1,066	0,287
ItemC16	-0,366	0,714	-1,810	0,070	-0,662	0,508
ItemC17	-0,510	0,610	0,905	0,366	-0,082	0,935
ItemC18	-2,519	0,012	0,732	0,464	-1,534	0,125
ItemC19	0,015	0,988	1,521	0,128	-0,101	0,919
ItemC20	-0,372	0,710	0,141	0,888	2,039	0,041

Itemfitstatistik

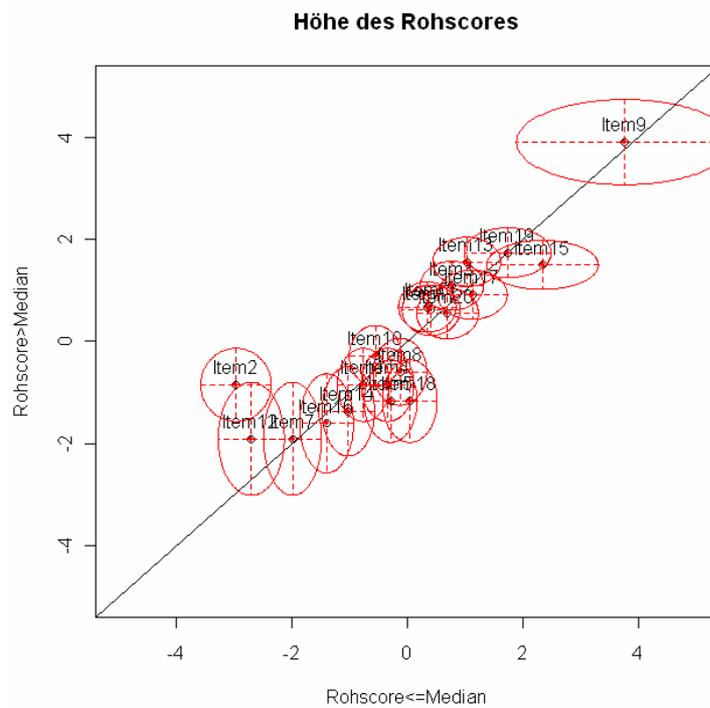
Tab. 21: Itemfitstatistik Wortschatz Testform C

Item	Chisqare	Df	p-Wert
ItemC01	230,672	165	0,001
ItemC02	451,638	165	0,000
ItemC03	162,006	165	0,551
ItemC04	129,793	165	0,980
ItemC05	159,085	165	0,615
ItemC06	140,483	165	0,917
ItemC07	140,742	165	0,915
ItemC08	131,624	165	0,974
ItemC09	119,206	165	0,997
ItemC10	221,021	165	0,002
ItemC11	141,573	165	0,906
ItemC12	179,165	165	0,213
ItemC13	283,810	165	0,000
ItemC14	116,982	165	0,998
ItemC15	115,017	165	0,999
ItemC16	338,897	165	0,000
ItemC17	141,838	165	0,904
ItemC18	98,848	165	1,000
ItemC19	178,483	165	0,224
ItemC20	262,707	165	0,000

In der Itemfitstatistik (Tab. 21) wiesen die Items C01, C02, C10, C13, C16 und C20 einen signifikanten P-Wert auf.

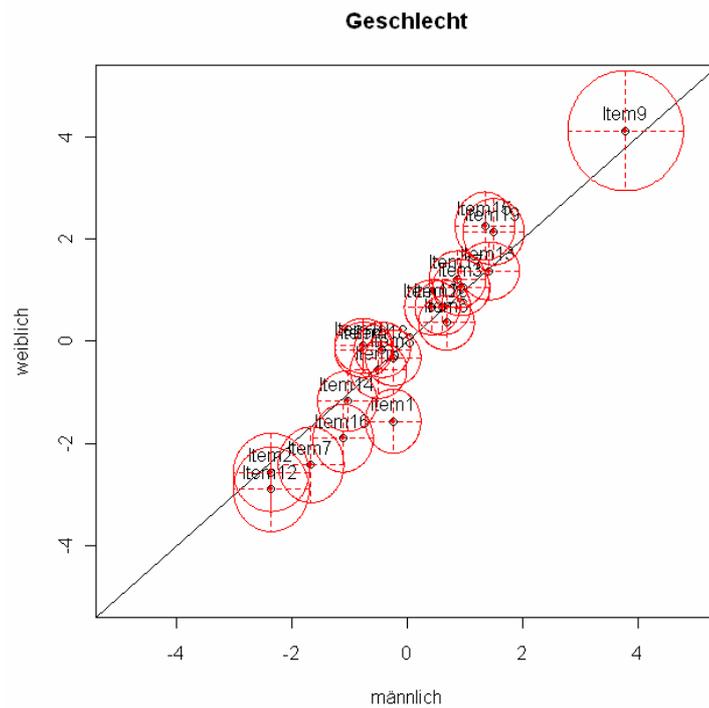
Grafische Modelltests

Abb.13: Grafischer Modelltest „Wortschatz“ Testform C, TK „Höhe des Rohwerts“



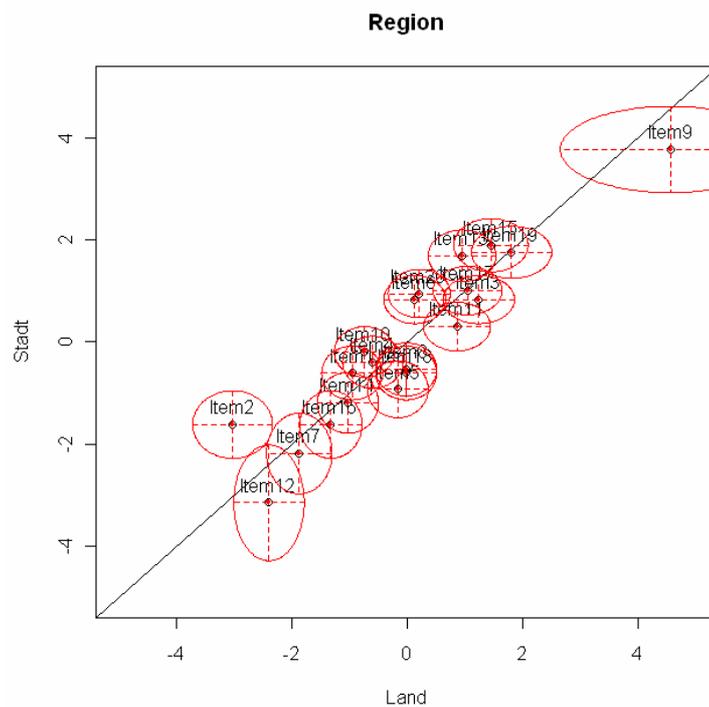
Der grafische Modelltest mit dem Teilungskriterium "Rohwert"(Abb. 13) zeigte, dass die Konfidenz-Ellipsen der Items 2 und 18 (entspricht Items C02 und C18) die Gerade nicht schneiden.

Abb.14: Grafischer Modelltest „Wortschatz“ Testform C, TK „Geschlecht“



Der grafische Modelltest mit dem Teilungskriterium “Geschlecht” (Abb. 14) zeigte, dass die Konfidenz-Ellipsen der Items 1 und 15(entspricht Items C01 und C15) die Gerade nicht schneiden.

Abb.15: Grafischer Modelltest „Wortschatz“ Testform C, TK „Region“



Der grafische Modelltest mit dem Teilungskriterium "Region" (Abb. 15) zeigte, dass die Konfidenz-Ellipse des Items 2 (entspricht Items C02) die Gerade nicht schneidet.

Qualitative Analyse

ItemC02 und C20:

C02) Leopard
Der Leopard gehört zur Familie der <i>a)Edelsteine</i> und ist ein
<i>b)Menschenaffen</i>
<i>c)Grünpflanzen</i>
<i>d)Katzen</i>
<i>e)Beutetier.</i>
<i>f)Raubtier</i>
<i>g)Heilmittel</i>
<i>h)Schutzsymbol</i>

C20) Frustration
Frustration entsteht unter anderem wenn ein/e <i>a)Frage</i> nicht
<i>b)Bedingung</i>
<i>c)Wunsch</i>
<i>d)These</i>
<i>e)verboten</i> wird.
<i>f)gestellt</i>
<i>g)aufbereitet</i>
<i>h)erfüllt</i>

Nicht eindeutig lösbar: Sowohl bei Item C02 als auch bei Item C20 ergab eine genauere Analyse, dass mehrere Antwortkombinationen als richtig gewertet werden müssten. So ist bei der in Item C20 unglücklich gewählten Definition „Frustration entsteht unter anderem wenn ...“ nicht nur die Fortsetzung „ein Wunsch nicht erfüllt wird“ sondern auch „eine Bedingung nicht erfüllt wird“ möglich. Durch die vage Formulierung „unter anderem“ sind auch weitere Antwortkombinationen wie „eine Frage nicht gestellt wird“ möglich – man denke zum Beispiel an einen Schüler, der frustriert ist, weil ihm bei einer Prüfung genau zu dem von ihm bestgelernten Kapitel keine Frage gestellt wurde. Betreffend Item C02 ist zu sagen, dass einige Schüler angaben, dass es sich bei einem Leopard um einen Edelstein handle. Internetrecherchen ergaben, dass es tatsächlich einen Edelstein namens Leopard gibt, der zur Familie der Jaspis Edelsteine zählt und dem sowohl heilende als auch schützende Wirkung zugesprochen wird. Somit ist auch Item C02 nicht eindeutig lösbar, da die Antworten „Katzen“ in Kombination mit

„Raubtier“ sowie „Edelsteine“ in Kombination mit „Heilmittel“ oder „Schutzsymbol“ als richtig gewertet werden müssen.

Zusammenfassende Schlussfolgerung

Aufgrund der in der qualitativen Analyse aufgedeckten Konstruktionsfehler wurden die Items C02 und C20 in der zweiten LQT ausgeschlossen. Auch die Ergebnisse der Z-Tests, Itemfitstatistik und der grafischen Modelltests sprachen für diese Entscheidung.

Likelihood-Quotienten-Test der zweiten Schätzung

Für den Itempool „Wortschatz Testform C“ konnte die H_0 „Das Rasch-Modell gilt“ nach Ausschluss der Items C02 und C20 a posteriori beibehalten werden, da die Andersen Likelihood-Quotienten-Tests auf dem 1% Niveau nicht signifikant wurden. Eine Übersicht sowie Veranschaulichung der Resultate zeigt die folgende Tabelle (Tab. 22):

Tab. 22: Zweite Parameterschätzung, Test Wortschatz C, Rasch-Modell-konform

Andersen LQT	Intern	Geschlecht	Region	Kritischer Wert df=17
LQ-Wert	19,191	30,089	25,333	32,51
Chi-square df	17	17	17	17
P-Wert	0,318	0,026	0,088	0,013

Funktionen Abstrahieren Testform A

Z-Test

Der Z-Test (Tab. 23) wies mit dem Teilungskriterium „Intern“ bei den Items E02, E10 und E14 einen p -Wert kleiner dem vorher festgelegten kritischen Wert von $p=0,05$ auf. Beim Teilungskriterium „Geschlecht“ unterschritten die p -Werte der Items E02, E03, E07 und E09 den Wert von 0,05. Die Items E15 und E04 wiesen beim Teilungskriterium Region einen signifikanten p -Wert auf. Das Item E01, welches bei den Teilungskriterien „Intern“ und „Geschlecht“ nicht geschätzt werden konnte, verhielt sich im Z-Test mit dem Teilungskriterium „Region“ ($p = 0,359$) Rasch-Modell-konform.

Tab. 23: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Funktionen Abstrahieren Testform A

Item	z-Wert INTERN	p-Wert INTERN	z-Wert GESCHL.	p-Wert GESCHL.	z-Wert REGION	p-Wert REGION
ItemE01	-	-	-	-	0,918	0,359
ItemE02	2,628	0,009	-3,739	0,000	-0,650	0,516
ItemE03	0,459	0,646	-2,129	0,033	-1,562	0,118
ItemE04	0,762	0,446	0,162	0,871	-2,029	0,042
ItemE05	0,154	0,878	-0,189	0,850	1,294	0,196
ItemE06	-1,205	0,228	-1,841	0,066	-0,978	0,328
ItemE07	1,236	0,217	2,044	0,041	0,490	0,624
ItemE08	-1,551	0,121	0,552	0,581	0,164	0,870
ItemE09	-1,325	0,185	2,350	0,019	-0,435	0,664
ItemE10	2,205	0,027	1,161	0,246	1,487	0,137
ItemE11	0,138	0,890	0,552	0,581	0,603	0,546
ItemE12	-0,724	0,469	0,720	0,471	-1,282	0,200
ItemE13	0,208	0,835	1,779	0,075	-1,345	0,179
ItemE14	2,558	0,011	-1,377	0,169	0,640	0,522
ItemE15	1,951	0,051	-0,860	0,390	2,284	0,022

Itemfitstatistik

In der Itemfitstatistik (Tab. 24) wiesen die Items E02 ($p = 0,001$) und E03 ($p = 0,003$) ein signifikantes Ergebnis auf.

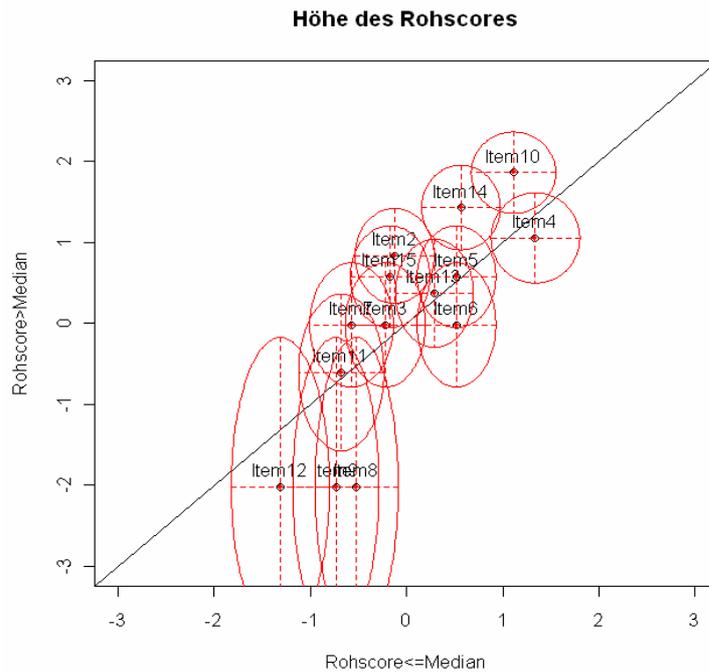
Tab. 24: Itemfitstatistik Funktionen Abstrahieren Testform A

Item	Chisqare	Df	p-Wert
ItemE01	158,037	163	0,595
ItemE02	225,495	163	0,001
ItemE03	138,761	163	0,916
ItemE04	216,794	163	0,003
ItemE05	157,132	163	0,615
ItemE06	133,334	163	0,957
ItemE07	181,003	163	0,159
ItemE08	113,134	163	0,999
ItemE09	120,163	163	0,995
ItemE10	162,393	163	0,499
ItemE11	162,669	163	0,493
ItemE12	145,313	163	0,837
ItemE13	159,823	163	0,556
ItemE14	180,042	163	0,171
ItemE15	182,996	163	0,135

Grafische Modelltests

Der grafische Modelltest mit dem Teilungskriterium "Rohwert" (Abb. 16) zeigte, dass die Konfidenz-Ellipsen der Items E02, E10 und E14 die Gerade nicht schneiden.

Abb.16: Grafischer Modelltest „Funktionen Abstrahieren“ Testform A, TK „Höhe des Rohwerts“



Der grafische Modelltest mit dem Teilungskriterium "Geschlecht" (Abb. 17) veranschaulichte, dass die Konfidenz-Ellipsen der Items E02, E03, E07 und E09 die Gerade nicht schneiden.

Abb.17: Grafischer Modelltest „Funktionen Abstrahieren“ Testform A, Teilungskriterium „Geschlecht“

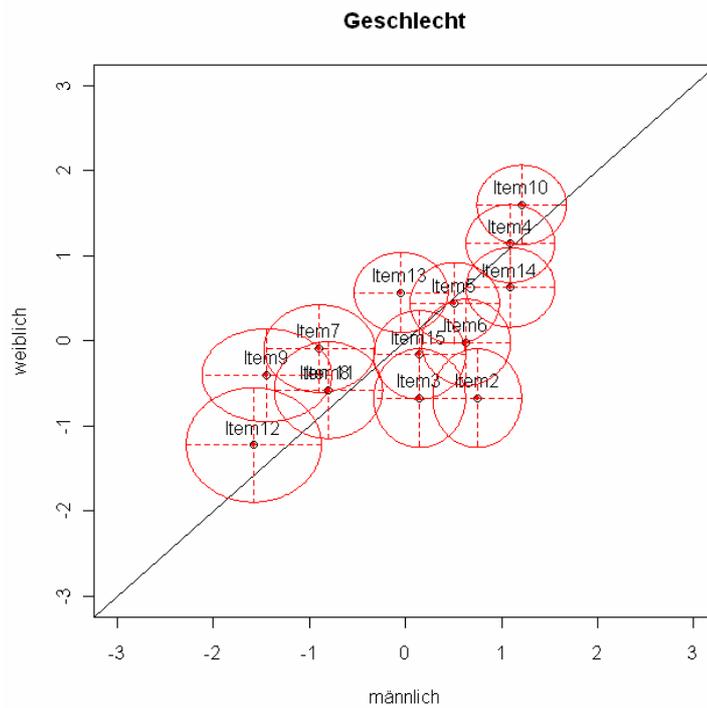
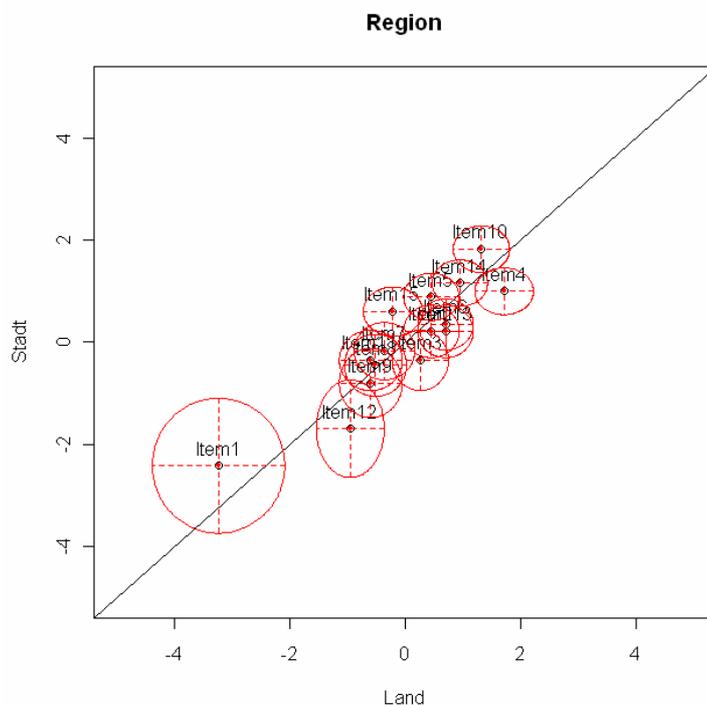


Abb.18: Grafischer Modelltest „Funktionen Abstrahieren“ Testform A, Teilungskriterium „Region“



Der grafische Modelltest mit dem Teilungskriterium “Region” (Abb. 18) zeigte, dass die Konfidenz-Ellipsen der Items E04 und E15 die Gerade nicht schneiden.

Qualitative Analyse

Item E02:

E02)Lockenstab	
a)Kamm	d)Lockenwickler
b)Schere	e)Pinzette
c)Rundbürste	

Geschlechtsunterschiede: Bei diesem Item ist gut denkbar, dass Mädchen bei der Lösung einen Vorteil hatten, da sie mit höherer Wahrscheinlichkeit als ein Bube einen Lockenstab kennen. So wurde während der Testung besonders von Buben oft gefragt, um was es sich bei einem „Lockenstab“ handle. Das Item läuft somit Gefahr nicht geschlechtsneutral zu sein, zudem steht seine Eindimensionalität in Frage: Die Lösungswahrscheinlichkeit hängt scheinbar nicht nur von der Fähigkeit abstrakt zu denken, sondern auch vom Geschlecht der Testperson und seinem Fachwissen im Bereich Haarpflege ab.

Zusammenfassende Schlussfolgerung

Mit Ausnahme des grafischen Modelltests und des Waldetsts mit dem Teilungskriterium „Region“ verhielt sich das Item E02 in allen Tests nicht Rasch-Modell-konform, zudem ergab eine qualitative Analyse, dass das Item E02 mehrere Schwachstellen aufweist. Es wurde daher von der abermaligen Überprüfung des Itempools „Funktionen Abstrahieren Testform A“ ausgeschlossen.

Likelihood-Quotienten-Test der zweiten Schätzung

Für den Itempool „Funktionen Abstrahieren Testform A“ konnte die H_0 „Das Rasch-Modell gilt“ nach Ausschluss des Items E02 a posteriori beibehalten werden, da die Andersen Likelihood-Quotienten-Tests auf dem 1% Niveau nicht signifikant wurden. Der Itempool umfasste beim Likelihood-Quotienten-Test mit dem Teilungskriterium Region 14 Items, bei den Teilungskriterien Intern und Geschlecht nur 13 Items, da das Item E01 wegen ungünstigen Antwortmustern in den Subgruppen von den Berechnungen ausgeschlossen wurde.

Eine Übersicht sowie Veranschaulichung der Resultate zeigt die folgende Tabelle (Tab. 25):

Tab. 25: Zweite Parameterschätzung, Funktionen Abstrahieren Testform A, Rasch-Modell-konform

Andersen LQT	Intern	Geschlecht	Kritischer Wert df=12	Region	Kritischer Wert df=13
LQ-Wert	19,222	25,106	25,41	22,205	26,86
Chi-square df	12	12	12	13	13
p-Wert	0,083	0,014	0,013	0,052	0,013

Funktionen Abstrahieren Testform B

Z-Test

Tab. 26: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Funktionen Abstrahieren Testform B

Item	z-Wert INTERN	p-Wert INTERN	z-Wert GESCHL.	p-Wert GESCHL.	z-Wert REGION	p-Wert REGION
ItemF01	-0,011	0,991	-2,118	0,034	-1,239	0,215
ItemF02	24,398	0,000	0,887	0,375	1,511	0,131
ItemF03	1,003	0,316	0,664	0,507	0,577	0,564
ItemF04	-1,307	0,191	1,081	0,280	-1,164	0,245
ItemF05	1,735	0,083	1,624	0,104	-0,248	0,804
ItemF06	-1,298	0,194	2,309	0,021	0,530	0,596
ItemF07	-1,049	0,294	1,353	0,176	-1,220	0,222
ItemF08	0,779	0,436	-0,130	0,897	0,592	0,554
ItemF09	-1,332	0,183	0,215	0,829	1,893	0,058
ItemF10	4,542	0,000	-1,303	0,192	1,611	0,107
ItemF11	0,313	0,754	0,473	0,636	-0,182	0,856
ItemF12	-1,743	0,081	-1,355	0,175	-1,713	0,087
ItemF13	-1,375	0,169	-1,077	0,281	0,084	0,933
ItemF14	-2,820	0,005	0,784	0,433	-0,147	0,883
ItemF15	-1,904	0,057	-0,751	0,453	-0,105	0,916

Der Z-Test (Tab. 26) wies mit dem Teilungskriterium „Intern“ bei den Items F02, F10 und F14 einen p -Wert kleiner dem kritischen Wert von $p = 0,05$ auf. Beim

Teilungskriterium „Geschlecht“ unterschritten die p -Werte der Items F01 und F06 den Wert von 0,05. Beim Teilungskriterium Region trat kein einziger signifikanter p -Wert auf.

Itemfitstatistik

Tab. 27: Itemfitstatistik Funktionen Abstrahieren Testform B

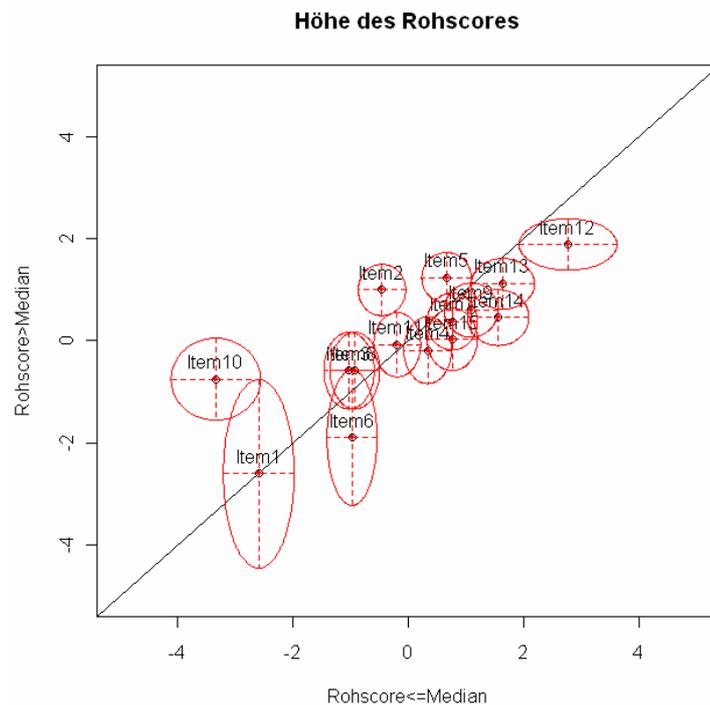
Item	Chisqare	Df	p-Wert
ItemF01	158,842	169	0,701
ItemF02	276,980	169	0,000
ItemF03	186,177	169	0,174
ItemF04	151,260	169	0,833
ItemF05	168,837	169	0,489
ItemF06	104,087	169	1,000
ItemF07	147,221	169	0,886
ItemF08	156,441	169	0,747
ItemF09	166,181	169	0,547
ItemF10	357,257	169	0,000
ItemF11	146,253	169	0,896
ItemF12	112,674	169	1,000
ItemF13	152,542	169	0,813
ItemF14	131,797	169	0,985
ItemF15	122,209	169	0,997

In der Itemfitstatistik (Tab. 27) wiesen die Items F02 und F10 mit einem $p=0,000$ ein signifikantes Ergebnis auf

Grafische Modelltests

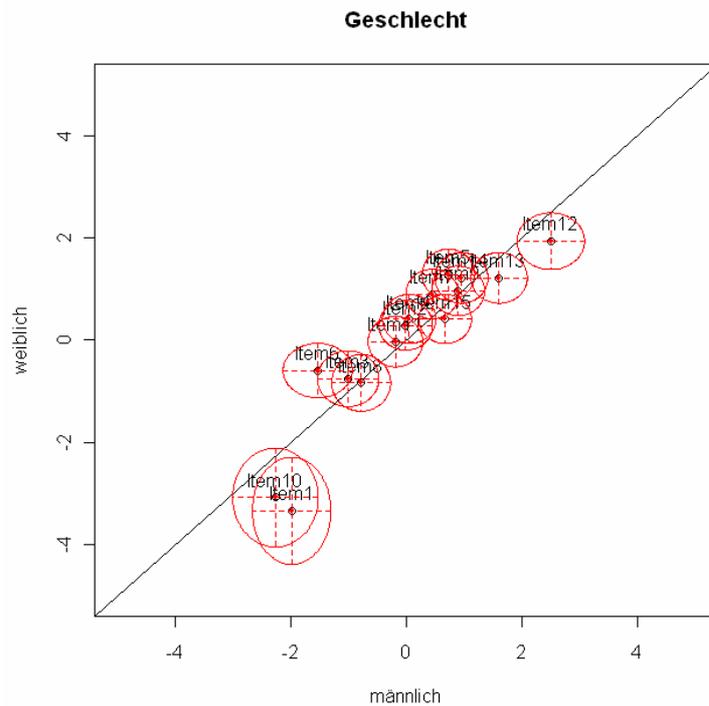
Der grafische Modelltest (Abb. 19) mit dem Teilungskriterium “Höhe des Rohwerts” zeigte, dass die Konfidenz-Ellipsen der Items 2, 10 und 14 (entspricht Items F02, F10 und F14) die Gerade nicht schneiden.

Abb.19: Grafischer Modelltest „Funktionen Abstrahieren“ Testform B, Teilungskriterium „Höhe des Rohwerts“



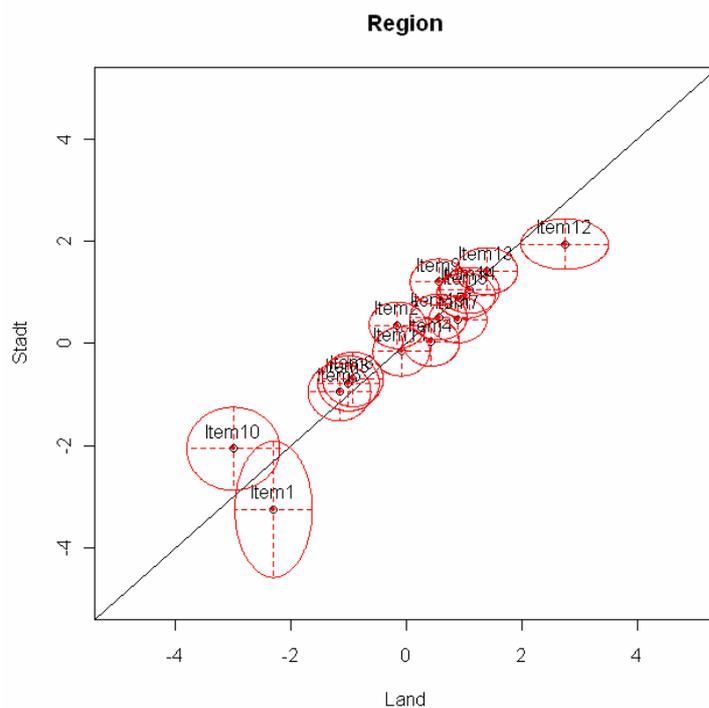
Der grafische Modelltest mit dem Teilungskriterium “Geschlecht” (Abb. 20) zeigte, dass die Konfidenz-Ellipsen der Items 1 und 6 (entspricht Items F01 und F06) die Gerade nicht schneiden.

Abb. 20: grafischer Modelltest „Funktionen Abstrahieren“ Testform B, Teilungskriterium „Geschlecht“



Im grafischen Modelltest mit dem Teilungskriterium “Region” (Abb. 21) schnitten sämtliche Konfidenz-Ellipsen die Gerade.

Abb.21: Grafischer Modelltest „Funktionen Abstrahieren“ Testform B, Teilungskriterium „Region“



Qualitative Analyse

Item F02

F02)Mikrofon

- a)Videorekorder
- b)Lautsprecher
- c)Hörgerät

- d)Antenne
- e)Scheinwerfer

Einfluss von Fachwissen: Bei der Lösung dieses Items spielten neben der Fähigkeit zu abstrahieren das Fachwissen einer Testperson im technischen Bereich eine Rolle. Besonders Kinder die ansonsten gute Leistungen erzielten, hatten bei der Lösung dieses Items Schwierigkeiten. Grund dafür könnte sein, dass sie komplizierter und weiter dachten als andere Testpersonen und nicht die Erhöhung der Lautstärke (Mikrofon, Lautsprecher, Hörgerät), sondern z.B. die Funktion des Aufnehmens von (akustischen) Signalen, wie es ein Mikrofon, ein Hörgerät und ein Videorekorder nicht aber ein Lautsprecher vermag, im Vordergrund sahen. So nehmen ein Mikrofon, ein Videorekorder und ein Hörgerät (akustische) Signale auf, während ein Lautsprecher diese ausstrahlt.

Andere Testpersonen wählten Lautsprecher, Mikrofon und Videorekorder; sie sahen vermutlich die Funktion und Möglichkeit des sich Repräsentieren Könnens. So kann eine Person mit Hilfe eines Mikrofons, Lautsprechers oder eine Videoaufnahme auf sich aufmerksam machen und sich präsentieren. Die Schwäche dieses Items lag also nicht nur im Einfluss des Fachwissens auf die Lösungswahrscheinlichkeit sondern auch in seiner fehlenden Eindeutigkeit.

Item F10

F10)Springschnur

- a)Murmeln
- b)Trampolin
- c)Ball

- d)Sprungbrett
- e)Sandschaufel

Problem der Sichtweise: Die Lösung dieses Items hing von der Sichtweise der Testperson ab. So sahen manche Testpersonen die Springschnur als Hilfsmittel und die spielende Person im Zentrum. Dies läge die Funktion des „Springens“ nahe und somit die Wahl der Alternativen Trampolin und Sprungbrett. Andere Testpersonen sahen die Springschnur im Zentrum und dachten aus diesem Blickwinkel über ihre Funktion nach,

die jene des „Kreisens“ wäre. Bei diesem Gedankengang, den vor allem Testpersonen mit einem hohen Rohwert hatten, waren die Antworten Murmeln und Ball naheliegend.

Zusammenfassende Schlussfolgerung

Im grafischen Modelltest und im Waldtest mit dem Teilungskriterium „Intern“, welches als einziges Teilungskriterium ein signifikantes Ergebnis im LQT aufwies, verhielten sich die Items F02, F10 und F14 nicht Rasch-Modell-konform. Auch in der Itemfitstatistik zeigte sich, dass die Items F02 und F10 nicht dem Rasch Modell entsprachen. Die vorgenommene qualitative Analyse stützte die Ergebnisse der Modelltests.

Likelihood-Quotienten-Test der letzten Schätzung

Nach dem die Likelihood-Quotienten-Tests bei getrennten Ausscheiden des Items F02 bzw. F10 signifikante Ergebnisse im Teilungskriterium „Höhe des Rohwerts“ aufwiesen, wurden in der vierten Parameterschätzung beide Items gleichzeitig aus dem ursprünglichen Itempool genommen.

Schließlich konnte für den Test „Funktionen Abstrahieren Testform B“, welcher nun 13 Items umfasste, die H_0 „Das Rasch-Modell gilt“ a posteriori beibehalten werden. Die Andersen Likelihood-Quotienten-Tests wurden auf dem 1% Niveau nicht mehr signifikant.

Eine Übersicht sowie Veranschaulichung der Resultate zeigt die folgende Tabelle (Tab. 28):

Tab. 28: Vierte Parameterschätzung, Test Funktionen abstrahieren Testform B, Rasch-Modell-konform

Andersen LQT	Intern	Geschlecht	Region	Kritischer Wert df=12
LR-value	18,612	19,516	12,057	25,41
Chi-square df	12	12	12	12
p-value	0,098	0,077	0,441	0,013

Funktionen Abstrahieren Testform D

Z-Test

Tab. 29: Z-Test mit Teilungskriterium Intern, Geschlecht und Region, Funktionen Abstrahieren Testform D

Item	z-Wert INTERN	p-Wert INTERN	z-Wert GESCHL.	p-Wert GESCHL.	z-Wert REGION	p-Wert REGION
ItemH01	-	-	0,593	0,553	-0,329	0,742
ItemH02	-1,232	0,218	-0,056	0,955	-0,275	0,784
ItemH03	5,362	0,000	-3,421	0,001	1,644	0,100
ItemH04	-0,323	0,747	-0,802	0,423	-1,019	0,308
ItemH05	-0,079	0,937	0,126	0,900	1,144	0,253
ItemH06	-0,837	0,402	0,229	0,819	0,187	0,851
ItemH07	-0,690	0,490	-1,388	0,165	-1,849	0,064
ItemH08	1,813	0,070	-1,013	0,311	0,546	0,585
ItemH09	0,763	0,445	-1,225	0,221	0,099	0,921
ItemH10	-0,107	0,915	1,615	0,106	-0,982	0,326
ItemH11	-1,670	0,095	-0,371	0,711	-0,827	0,408
ItemH12	-2,501	0,012	-0,664	0,507	0,636	0,524
ItemH13	0,316	0,752	2,801	0,005	-0,506	0,613
ItemH14	1,388	0,165	0,807	0,419	1,961	0,050
ItemH15	1,374	0,170	-1,701	0,089	1,275	0,202

Der Z-Test (Tab. 29) wies mit dem Teilungskriterium „Intern“ bei den Items H03 ($p = 0,00$) und H12 ($p = 0,012$) einen p -Wert kleiner dem kritischen Wert von $p=0,05$ auf. Beim Teilungskriterium „Geschlecht“ unterschritten die p -Werte der Items H03 und H13 den Wert von 0,05. Kein einziger signifikanter p -Wert trat hingegen beim Teilungskriterium Region auf.

Das Item H01, welches mit dem Teilungskriterium „Intern“ nicht geschätzt werden konnte, verhielt sich im Z-Test sowohl beim Teilungskriterium „Geschlecht“ ($p = 0,553$) als auch beim Teilungskriterium „Region“ ($p = 0,742$) Rasch-Modell-konform.

Itemfitstatistik

In der Itemfitstatistik (Tab. 30) wies das Items H03 mit einem $p = 0,000$ ein signifikantes Ergebnis auf.

Tab. 30: Itemfitstatistik Funktionen Abstrahieren Testform D

Item	Chisqare	Df	p-Wert
ItemH01	115,738	162	0,998
ItemH02	141,796	162	0,872
ItemH03	262,719	162	0,000
ItemH04	130,107	162	0,969
ItemH05	149,058	162	0,759
ItemH06	139,967	162	0,894
ItemH07	150,891	162	0,724
ItemH08	146,585	162	0,802
ItemH09	153,440	162	0,673
ItemH10	99,477	162	1,000
ItemH11	129,802	162	0,970
ItemH12	124,099	162	0,988
ItemH13	159,457	162	0,542
ItemH14	180,499	162	0,152
ItemH15	172,781	162	0,267

Grafische Modelltests

Der grafische Modelltest mit dem Teilungskriterium “Höhe des Rohwerts” (Abb. 22) zeigte, dass die Konfidenz-Ellipsen der Items 3 und 12(entspricht Items H03 und H12) die Gerade nicht schneiden.

Abb.22: Grafischer Modelltest „Funktionen Abstrahieren“ Testform D,
 Teilungskriterium „Höhe des Rohwerts“

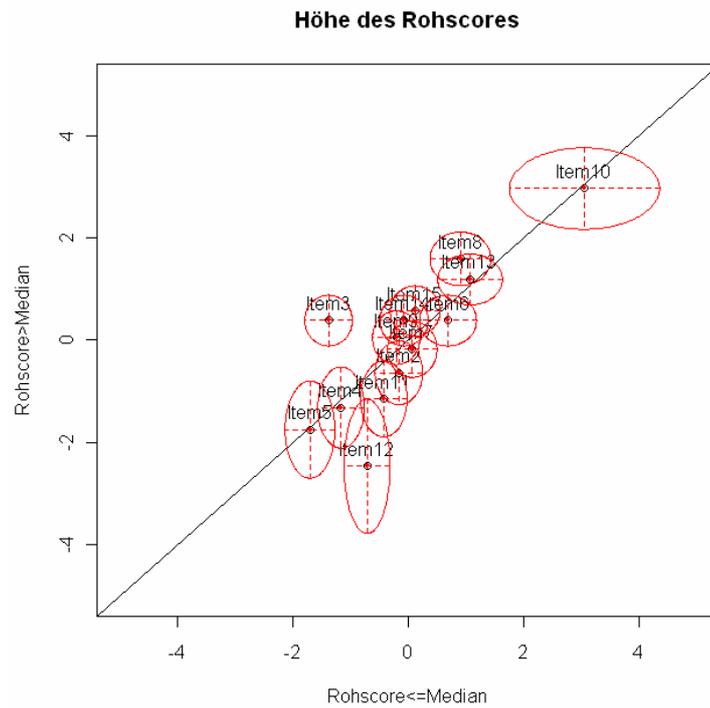
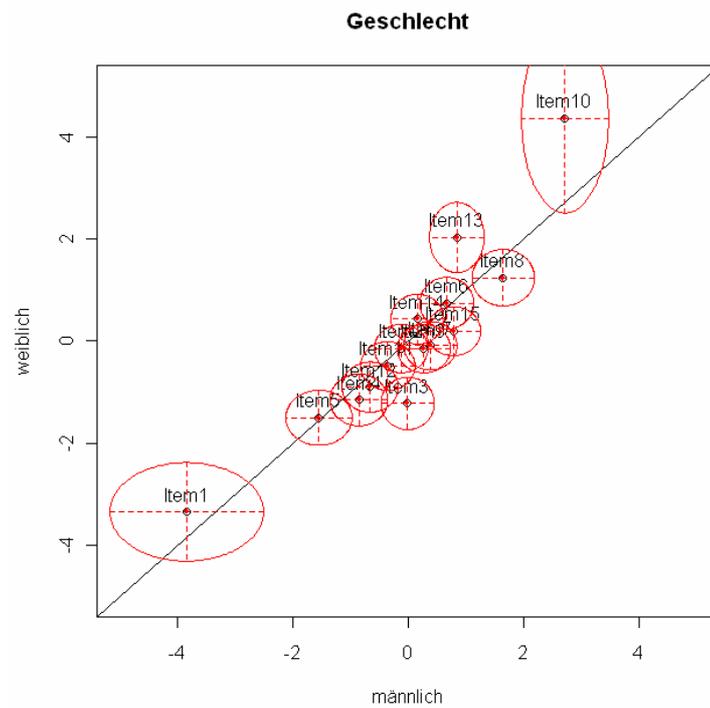
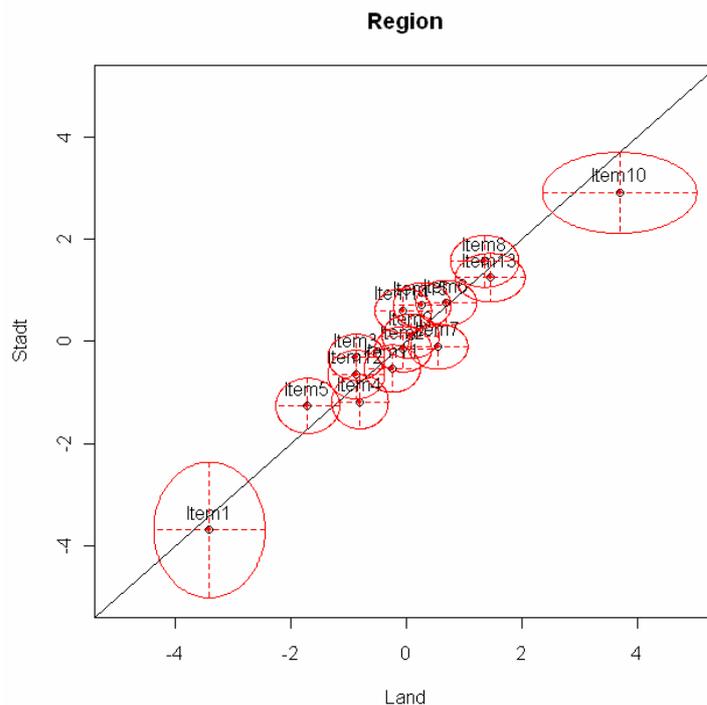


Abb.23: Grafischer Modelltest „Funktionen Abstrahieren“ Testform D,
 Teilungskriterium „Geschlecht“



Der grafische Modelltest mit dem Teilungskriterium “Geschlecht” (Abb. 23) zeigte, dass die Konfidenz-Ellipsen der Items 3 und 13 (entspricht Items H03 und H13) die Gerade nicht schneiden

Abb.24: Grafischer Modelltest „Funktionen Abstrahieren“ Testform D, Teilungskriterium „Region“



Der grafische Modelltest mit dem Teilungskriterium “Region” (Abb. 24) zeigte, dass sämtliche Konfidenz-Ellipsen die Gerade schneiden

Qualitative Analyse

Item H03:

H03) Socken

a) Handschuhe

b) Schuhe

c) Strumpfhose

d) Bademantel

e) Mütze

Konstruktionsfehler: Bei der Entwicklung dieses Items hatte sich ein Denkfehler eingeschlichen. Als zentrale Funktion der Socken wurde das Warmhalten von Füßen gesehen und daher die Schuhe und Strumpfhose als die richtig geltenden Antworten. Dieser Denkvorgang ist in mehrerer Hinsicht problematisch; so dienen Schuhe oder

auch eine Strumpfhose nicht immer zum warmhalten – man denke an Sandalen, Flip Flops oder eine Seidenstrumpfhose, eine Mütze oder Handschuhe haben hingegen eindeutig eine wärmende Funktion und wurden so auch von vielen Testpersonen gewählt. Zudem ist der Fokus auf die räumliche bzw. körperliche Nähe dieses Items kritisch zu sehen. Sollen die Testpersonen doch angehalten werden abstrakt zu denken und die Antworten aufgrund der gemeinsamen grundlegenden Funktion (in diesem Beispiel: spendet Wärme) und nicht aufgrund optischer Ähnlichkeiten oder räumlicher Nähe (trägt man am Fuß) zu wählen.

Item H12:

H12) Internat	
a) Hort	d) Kindergarten
b) Schwimmbad	e) Park
c) Schlafzimmer	

Zentralität der Funktion: Bei diesem Item stellte sich die Frage, ob die zentrale Funktion eines Internats darin besteht, dort schlafen und übernachten zu können, oder in der Betreuung von Kindern und Jugendlichen. Letzteres war die Ausgangsidee der Itemkonstruktion und somit „Hort“ und „Kindergarten“ die als richtig gewerteten Antworten. Für viele Testpersonen stand aber scheinbar die Funktion „des dort Schlafens“ im Vordergrund, so wählten sie „Schlafzimmer“ und „Kindergarten“ (auch hier wird ein Mittagsschlaf gehalten).

Kritisch ist außerdem, dass ein Hort nicht nur eine Tagesstätte für Kinder darstellt, sondern auch die Bedeutung „Schatz“ trägt.

Zusammenfassende Schlussfolgerung

Aufgrund des in der qualitativen Analyse offenbarten Konstruktionsfehlers im Item H03, wird dieses aus dem Itempool des Tests „Funktionen Abstrahieren Testform D“ ausgeschlossen. Zudem erwies sich dieses Item in sämtlichen Modelltests als nicht Rasch-Modell-konform. Auch für das Item H12 wurden Schwachstellen in der qualitativen Analyse festgestellt. Nicht Rasch-Modell-konform zeigte sich dieses Item in den Modelltests mit dem internen Teilungskriterium. Der nachfolgende LQT wurde ohne diese Items durchgeführt.

Likelihood-Quotienten-Test der vierten Schätzung

Nach dem die Likelihood-Quotienten-Tests bei getrennten Ausscheiden des Items H03 bzw. H12 signifikante Ergebnisse in zumindest einem der drei Teilungskriterien aufwiesen, wurden in der vierten Parameterschätzung mitsamt Modelltests beide Items gleichzeitig aus dem ursprünglichen Itempool entnommen.

Für den Itempool „Funktionen Abstrahieren Testform D“ konnte die H0 „Das Rasch-Modell gilt“ nach Ausschluss des Items H03 und H12 a posteriori beibehalten werden, da die Andersen Likelihood-Quotienten-Tests auf dem 1% Niveau nicht signifikant wurden. Der Itempool umfasste beim Likelihood-Quotienten-Test mit dem Teilungskriterien Geschlecht und Region 13 Items, bei dem Teilungskriterium Intern nur 12 Items, da das Item H01 wegen ungünstigen Antwortmustern in den Subgruppen von den Berechnungen ausgeschlossen wurde.

Eine Übersicht sowie Veranschaulichung der Resultate zeigt die folgende Tabelle (Tab. 31):

Tab. 31: Vierte Parameterschätzung, Test Funktionen abstrahieren Testform D, Rasch-Modell-konform

Andersen LQT	Intern	Kritischer Wert df=11	Geschlecht	Region	Kritischer Wert df=12
LQ-Wert	13,364	23,94	20,297	13,791	25,41
Chi-square df	11	11	12	12	12
p-Wert	0,27	0,013	0,062	0,314	0,013

9.4. Mittelwertvergleiche

Um den Einfluss der festen Faktoren Geschlecht, Schultyp und Schulstufe auf die Höhe des Rohwerts festzustellen, wurde pro Untertest – sofern die Voraussetzungen dafür als erfüllt angesehen werden konnten – eine dreifache Varianzanalyse durchgeführt.

Die VolksschülerInnen und SchülerInnen der 5. Klasse AHS wurden dabei von der Analyse ausgeschlossen, da in diesen Schulstufen zu wenige Daten pro Testform vorlagen, um aussagekräftige Schlüsse zu ziehen. So wurden in der 3. Klasse Volksschule nur 3 SchülerInnen im Durchschnitt pro Testform getestet, in der 4. Klasse Volksschule zwischen 5 und 6 SchülerInnen und in den 5. Klassen AHS durchschnittlich 13 Personen.

Die Faktorstufen der durchgeführten Varianzanalysen wurden gemäß einer Kreuzklassifikation (Geschlecht x Schultyp x Schulstufe) mit vollständiger Klassenbesetzung kombiniert (der interessierte Leser siehe Rasch & Kubinger, 2006).

Im Anschluss werden die zu testenden Hypothesen vorgestellt, die Überprüfung der Voraussetzungen für eine Varianzanalyse (an dieser Stelle sei erwähnt, dass die Voraussetzung des geforderten Skalenniveaus - nämlich zumindest intervallskaliert - für die Rohwerte aller Testformen als gegeben angesehen werden kann) sowie die Ergebnisse der Varianzanalyse präsentiert.

Hypothesen

Da für jeden Untertest und jede seiner Testformen dieselben Hypothesen überprüft wurden, werden diese einmalig hier angeführt. Die Variable x steht für den Untertest („Wortschatz“ oder „Funktionen Abstrahieren“), die Variable y für die Testform (A, B, C, D).

H01: „Die Testwerte im Untertest X Testform Y unterscheiden sich nicht zwischen den einzelnen Stufen des Faktors Geschlecht“

H02: „Die Testwerte im Untertest X Testform Y unterscheiden sich nicht zwischen den einzelnen Stufen des Faktors Schulstufe“

H03: „Die Testwerte im Untertest X Testform Y unterscheiden sich nicht zwischen den einzelnen Stufen des Faktors Schultyp“

H04: „Es gibt keine Wechselwirkungen zwischen Geschlecht und Schulstufe im Bezug auf die Testwerte im Untertest X Testform Y “

H05: „Es gibt keine Wechselwirkungen zwischen Geschlecht und Schultyp im Bezug auf die Testwerte im Untertest X Testform Y“

H06: „Es gibt keine Wechselwirkungen zwischen Schulstufe und Schultyp im Bezug auf die Testwerte im Untertest X Testform Y“

H07: „Es gibt keine Wechselwirkungen zwischen Geschlecht, Schulstufe und Schultyp im Bezug auf die Testwerte im Untertest X Testform Y“

Dreifache Varianzanalyse

Der Levene Test zur Überprüfung der Homogenität der Varianzen über die Gruppen hinweg wurde in allen Testformen der Untertests nicht signifikant, die Berechnung der Varianzanalysen war damit gerechtfertigt.

Der Faktor Geschlecht ergab in keiner der Testformen ein signifikantes Ergebnis. Somit war die H01 für alle Untertests und deren Testformen beizubehalten. Gleichfalls verhielt es sich mit den Wechselwirkungen zwischen je zwei Faktoren, wodurch die H04, H05, H06 und H07 beibehalten wurden.

Hingegen zeigten die Faktoren Schulstufe und Schultyp signifikante Haupteffekte in nahezu allen Testformen. Ausnahme bildete die Testform C im Untertest „Funktionen Abstrahieren“, in der nur der Faktor „Schulstufe“ einen signifikanten Einfluss zeigte. Für diese Testform musste die H02 verworfen werden, in allen anderen Testformen wurde neben dieser auch die H03 verworfen.

Wechselwirkungen zwischen allen drei Faktoren wurden in der Testform A des Untertests „Funktionen Abstrahieren“ festgestellt, womit die H08 hier verworfen wurde. In Tabelle 32 sind jene Faktoren zusammengefasst, die einen signifikanten Anteil der Varianz der Ergebnisse in einer Testform erklären. Zudem ist auch „Eta-Quadrat“ als Maß für die Effektgröße der Faktoren abzulesen. Die Faktoren weisen in den Testformen des Untertests „Wortschatz“ durchwegs höhere Effektgrößen als im Untertest „Funktionen Abstrahieren“ auf.

Tab. 32: Dreifache Varianzanalyse, Wortschatz und Funktionen Abstrahieren

Untertest	Testform	Quelle	F	Sig.	Partielles Eta-Quadrat
Wortschatz	Testform A	Schulstufe	10,750	0,000	0,193
		Schultyp	30,153	0,000	0,183
	Testform B	Schulstufe	19,203	0,000	0,304
		Schultyp	36,486	0,000	0,217
	Testform C	Schulstufe	14,823	0,000	0,253
		Schultyp	19,448	0,000	0,129
	Testform D	Schulstufe	9,243	0,000	0,176
		Schultyp	30,882	0,000	0,192
Funktionen Abstrahieren	Testform A	Schulstufe	6,128	0,001	0,121
		Schultyp	8,974	0,003	0,063
		Schulstufe x Schultyp x Geschlecht	3,732	0,013	0,077
	Testform B	Schulstufe	6,984	0,000	0,135
		Schultyp	8,494	0,004	0,060
	Testform C	Schulstufe	5,268	0,002	0,108
	Testform D	Schulstufe	4,956	0,003	0,105
Schultyp		5,307	0,023	0,040	

10. Diskussion und Ausblick

Resümierend lässt sich die Ausgangsfrage der Diplomarbeit „AID 2 als Gruppentestung?“ zumindest für die Untertests „Wortschatz“ und „Funktionen abstrahieren“ mit einem „Ja“ beantworten. So ließ sich die Testung in allen Klassen ohne Zwischenfälle durchführen, selbst in der dritten Klasse Volksschule kam es bei keinem Kind zu Verständnisschwierigkeiten und die Beispiel - Items konnten in jeder Klasse gelöst werden.

Aus testtheoretischer Sicht ist zu sagen, dass die Konformität mit dem dichotom logistischen Modell nach Rasch zumindest a posteriori für alle Testformen angenommen werden konnte. Die Betonung liegt dabei auf „angenommen“. Gemäß der wissenschaftlichen Praxis nach Popper (1976) können Theorien und Hypothesen niemals verifiziert sondern nur vorübergehend angenommen oder falsifiziert werden.

Die Konformität der acht Itempools mit dem Rasch – Modell ist daher in Zukunft anhand weiterer Teilungskriterien und vor allem anhand einer größeren Stichprobe zu überprüfen. So ist kritisch anzumerken, dass der Stichprobenumfang n pro Testform im Durchschnitt nur 170 betrug, was eine äußerst geringe Anzahl zur Überprüfung eines Datensatzes auf Rasch-Modell-Konformität darstellt und in jeden Falle durch die Testung einer größeren, repräsentativeren Stichprobe bestätigt werden muss.

So liegt eine weitere Schwachstelle der Stichprobe darin, dass keine Zufallsauswahl stattfand und lediglich vier verschiedene Schulen, welche sich alle in Oberösterreich befinden, getestet wurden. Außerdem kam es zu einer Konfundierung der Variablen „Region“ und „Schultyp“, da im ländlichen Raum ausschließlich Volksschulen und Hauptschulen getestet wurden, während die Testung in der Stadt in einem Gymnasium erfolgte. Dies könnte auch der Grund sein, warum die Unterschiede zwischen Hauptschule und Gymnasium relativ klein ausgefallen waren. Während in ländlichen Regionen viele Kinder nach der Volksschule vorerst die Hauptschule besuchen, anschließend eine Lehre antreten oder sich für eine AHS bzw. BHS entscheiden, findet in der Stadt eine Differenzierung meist schon unmittelbar nach der Volksschule statt und es ist eher die Ausnahme denn die Regel, dass ein/e SchülerIn nach der Hauptschule eine weiterbildende Schule besucht. Es ist gut denkbar, dass bei einem Vergleich zwischen städtischen Hauptschulen und höheren Bildungsanstalten größere Unterschiede in den Testleistungen, vor allem im Untertest „Wortschatz“, auftreten. Der stärkere Einfluss der Variable „Schultyp“ auf die Ergebnisse in diesem Untertest

verglichen mit dem Untertest „Funktionen abstrahieren“ zeichnete sich auch in der vorliegenden Testung ab. Dies geht mit den Ergebnissen aus diversen Studien konform, in denen sich die Förderungsabhängigkeit des Wortschatzes zeigte (Angerer, 1991; Mayer-Frühwirth, 1988; McElvany et al., 2009; Schneider & Stefanek, 2004).

Der Einfluss des Faktors Region wurde nicht untersucht, da sich wie weiter oben bereits thematisiert die Variablen „Schultyp“ und „Region“ zu stark überschneiden. Es war daher obsolet neben dem Einfluss des Faktors Schultyp auch jenen des Faktors Region zu untersuchen. Bei der Interpretation der Ergebnisse ist jedoch zu berücksichtigen, dass die Variablen Region und Schultyp konfundiert sind. Daher können festgestellte Unterschiede in den Rohwerten auf die Region und/oder den Schultyp zurückgeführt werden. Eindeutigkeit bezüglich der Quelle der Rohwertunterschiede liefert nur eine Untersuchung, in der sämtliche Schultypen sowohl in ländlichen Regionen als auch in der Stadt getestet werden.

Im Zuge einer größeren Testung wäre es auch von Interesse, pro Untertest die vier verschiedenen Testformen zu verbinden, um festzustellen, ob sich die Itempools der verschiedenen Testformen auch zueinander Rasch-Modell-konform verhalten. Diese Vorgehensweise würde ebenso die Überprüfung der Gültigkeit der postulierten *Radicals* mittels Linearem Logistischem Testmodell von Fischer (LLTM) ermöglichen.

Weiters ist zu überprüfen, ob eine Vorgabe der generierten Items im angestrebten Computerformat zu denselben Ergebnissen wie im Zuge dieser Diplomarbeit vorgegebenem Paper-Pencil-Format führt. Nicht selten zeigen sich nämlich bei identischem Testmaterial aber verschiedenen Vorgabemodi Unterschiede in diversen psychometrischen Variablen (vgl. Troche, Rammstedt & Rammsayer, 2002).

Trotz der beschriebenen Verbesserungswürdigkeiten konnte im Zuge dieser Diplomarbeit die Wichtigkeit eines theorie- und regelgeleiteten Vorgehens bei der Itemgenerierung aufgezeigt werden. Die Durchsicht von Literatur und das Erstellen von Regeln kostet zwar viel Zeit und Mühe, macht sich aber am Ende bezahlt. So mussten im Untertest „Wortschatz“ lediglich 6,5% und im Untertest „Funktionen Abstrahieren“ rund 8% der gesamten Items ausgeschlossen werden, um die Rasch-Modell-Konformität der Itempools zu gewährleisten. Die Mängel der ausgeschiedenen Items

bestanden vor allem darin, dass eine der vorher festgelegten Konstruktionsregeln verletzt worden war.

Ließe sich in Zukunft anhand des LLTM für beide Untertests ein verbindlicher Regelkatalog für die Erstellung von Items entwickeln, würde diese um ein Vielfaches erleichtert werden. Der/die Testentwickler/in wüsste, auf welche Element er/sie zu achten hat und wie diese zu variieren sind, um mehr oder weniger schwierige Items zu erhalten.

Mit zunehmender Entwicklung der Computerhochsprachen ist auch ein Schritt in automatische Itemgenerierung, wie sie schon bei der Erstellung von Matrizenaufgaben eingesetzt wurde (vgl. Arendasy & Gittler, 2003), denkbar. Da sich aber das verbale Testmaterial grundlegend von Matrizenaufgaben unterscheidet, in vielen Bereichen komplexer und facettenreicher ist und die abgeleiteten Regeln anderer Natur sind, bleibt eine vollständige automatische Itemgenerierung wohl unrealistisch.

Im Bereich des Möglichen befindet sich aber eine teilautomatisierte Generierung in der ein ständiger Austausch zwischen Computer und PsychologInnen stattfindet. Anhand des entwickelten Programms werden Items generiert, durch TestentwicklerInnen kontrolliert, analysiert und verbessert, und das Itemgenerierungsprogramm auf Basis neuer Erkenntnisse weiterentwickelt.

11. Zusammenfassung

Ziel dieser Diplomarbeit war es eine Gruppentestversion für die Untertests „Soziale und Sachliche Folgerichtigkeit“, „Synonyme finden“ und „Funktionen Abstrahieren“ des Adaptiven Intelligenz Diagnostikum 2 (Version 2.2) von Kubinger (2009a) zu entwickeln.

Die Itemgenerierung erfolgte auf Basis einer eingehenden Literaturrecherche und einer Analyse der Originalitems des AID 2 hinsichtlich der Unterschiede zwischen den verschiedenen Schwierigkeitsstufen.

Für den Untertest „Soziale und Sachliche Folgerichtigkeit“ wurde die Grundannahme getroffen, dass die Testperson bei der Lösung eines Items vor einem Problem steht – sie muss von einem Ausgangszustand (ungeordnete Bilder) zu einem gewünschten Zielzustand (in sich logische Bildergeschichte) finden. Nach Kotovsky und Simon (1990) gestaltet sich ein Problem umso schwieriger, je mehr Dimensionen und Merkmale bei seiner Lösung zu beachten sind. Auch die Größe von Zeitabständen spielt beim Erkennen von Ursache-Wirkungs-Zusammenhängen eine Rolle (Goswami, 2001). Erschwerend für die Lösungsfindung ist schließlich, wenn ein wesentlicher Unterschied zwischen aktueller und vergangener Situation besteht und somit analoges Problemlösen ineffizient und behindernd wirkt (Chen & Honomichl, 2004; Mayer, 1979; Wessels, 1994). Diese Erkenntnisse aus der Literatur wurden verwendet um *Radicals* zu entwickeln (vgl. Arendasy et al., 2006) und das Schwierigkeitsniveau der Items zu variieren.

Bei der Entwicklung des Untertests „Wortschatz“ wurden die aus der Literatur begründeten Annahmen getroffen, dass die Wortart (Busemann, 1965; Neuhaus, 1962; Oerter & Montada, 2008), das Abstraktionsniveau (Gerrig & Zimbardo, 2008; Neuhaus, 1962) und die Verwendungshäufigkeit im Alltag (Oerter & Montada, 2008) einen Einfluss auf den Schwierigkeitsgrad eines Items haben.

Bei der Bearbeitung von Items im Untertest „Funktionen Abstrahieren“ sind Prozesse der Kategorienbildung beteiligt. Ausgangspunkt für das Bilden von Kategorien sind zentrale Merkmale eines Stimulus; dies können unter anderem seine äußere Gestalt oder seine Funktionalität sein (Goswami, 2001; Neuhaus, 1962). Je nachdem ob die Funktion

eines Stimulus bestimmend für die Kategorisierung ist oder nicht, fällt die Lösung des Items mehr oder weniger schwer. Einfluss auf die Schwierigkeit eines Items hat auch hier wieder die Relevanz im Alltag des präsentierten Stimulus.

Eine Vorgabe der entwickelten Items fand für die beiden letzt genannten Untertests statt, wobei 679 SchülerInnen aus vier verschiedenen Schulen getestet wurden.

Im Anschluss an die Testung wurde die Konformität der Itempools mit dem dichotom-logistischen Modell nach Rasch mittels Likelihood-Quotienten-Test nach Andersen (1973) anhand der Teilungskriterien Geschlecht, Region (Stadt-Land) und Median überprüft. Diese konnte zumindest a posteriori nach Ausschluss von maximal zwei Items für alle vier pro Untertest entwickelten Testformen angenommen werden.

In weiteren Untersuchungen sollte die Rasch-Modell-Konformität anhand einer größeren Stichprobe überprüft, sowie ein LLTM pro Untertest für die miteinander verbundenen Testformen durchgeführt werden.

Lassen sich daraus verbindliche Regeln für die Itemkonstruktion ableiten, wäre für die Zukunft eine teil-automatische Itemgenerierung denkbar.

12. Literatur

Andersen, E.B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123-140.

Angerer, C. (1991). Elterliches Anregungsmilieu und Intelligenzprofil im AID. Dipl. Arb. Univ. Wien, Wien.

Arendasy, M.E. & Gittler, G. (2003). IRT-basierter Vergleich zweier Varianten automatisiert erstellter Matrizentestaufgaben. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24(4), 261-257.

Arendasy, M.E., Sommer, M., Gittler, G. & Hergovich, A. (2006). Automatic Generation of Quantitative Reasoning Items: A Pilot Study. *Journal of Individual Differences*, 27(1), 2-14.

Busemann, A. (1965). *Kindheit und Reifezeit*. Stuttgart: Union Druckerei GmbH.

Case, R. (1992). *The mind's staircase. Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale NJ: Erlbaum.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1-22.

Chen, Z., Mo, L., & Honomichl, R. (2004). Having the memory of an elephant: Long-term retrieval and the use of analogues in problem solving. *Journal of Experimental Psychology: General*, 133(3), 415-433.

Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.

Gerrig, R.J. & Zimbardo, P.G. (2008). *Psychologie* (18., akt. Aufl.). München: Pearson Education Deutschland GmbH.

Goswami, U. (2001). *So denken Kinder - Einführung in die Psychologie der kognitiven Entwicklung*. Bern: Verlag Hans Huber.

Greeno, J.G., Moore, J.L. & Smith, D.R. (1993). Transfer of situated learning. In D.K. Dettermann & R.J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition and instruction* (pp. 99-167). Westport, CT: Ablex.

Häcker, H.O. & Stapf, K.H. (Hrsg.). (2009). *Dorsch Psychologisches Wörterbuch* (15. überarb. u. erw. Aufl.). Bern: Verlag Hans Huber.

Hornke, L. (2001). Benötigte Itemanzahlen beim meß- und entscheidungsorientierten adaptiven Testen. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 22(3), 185-193.

Keil, F.C. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234-254). New York: Cambridge University Press.

Kubinger, K.D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. In K.D. Kubinger (Hrsg.), *Moderne Testtheorie - Ein Abriss samt neuesten Beiträgen* (2. Aufl., S. 19-83). Weinheim: Beltz.

Kubinger, K.D. (2003). Adaptives Testen. In K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik*. Weinheim: Beltz.

Kubinger, K.D. (2009a). *Adaptives Intelligenz Diagnostikum – Version 2.2 (AID 2) samt AID 2-Türkisch*. Göttingen: Beltz.

Kubinger, K.D. (2009b). *Psychologische Diagnostik – Theorie und Praxis psychologischen Diagnostizierens* (2., überarb. u. erw. Aufl.). Göttingen: Hogrefe.

Kubinger, K.D. & Wild, B. (1989). Die Optimierung der Meßgenauigkeit beim „branched“-adaptiven Testen. In K.D. Kubinger (Hrsg.), *Moderne Testtheorie - Ein Abriss samt neuesten Beiträgen* (2. Aufl., S. 187-218). Weinheim: Beltz.

Kotovsky, K. & Simon, H. A. (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology*, 22(2), 143-183.

Lewin, K. (1963). *Feldtheorie in den Sozialwissenschaften*. Bern: Huber.

Lienert, G.A. & Raatz, U. (1994). *Testaufbau und Testanalyse* (5., vollst. überarb. u. erw. Auflage). Weinheim: Beltz.

Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

Mayer-Frühwirth, G. (1998). Differentielle Intelligenzdiagnostik von besonders geförderten und nicht geförderten Kindern. Dipl. Arb. Univ. Wien, Wien

McElvany, N., Becker, M. & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten. *Zeitschrift für Entwicklungspsychologische und Pädagogische Psychologie*, 41(3), 121-131.

Neuhaus, W. (1962). *Der Aufbau der geistigen Welt des Kindes*. (2., überarb. u. erw. Auflage). München: E. Reinhardt.

Oerter, R. & Montada, L. (Hrsg.). (2008). *Entwicklungspsychologie*. Weinheim: Beltz PVU.

Popper, K.R. (1976). *Logik der Forschung*. Tübingen: Mohr.

Rasch, D. & Kubinger, K.D. (2006). *Statistik - Für das Psychologiestudium*. München: Elsevier GmbH.

Schmidt, F.L. & Hunter, J.E. (1999). Comparison of Three Meta-Analysis Methods Revisited: An Analysis of Johnson, Mullen, and Salas (1995). *Journal of Applied Psychology*, 84(1), 144-148.

Schneider, W. & Stefanek, J. (2004). Entwicklungsveränderungen allgemeiner kognitiver Fähigkeiten und schulbezogener Fertigkeiten im Kindes- und Jugendalter. *Zeitschrift für Entwicklungspsychologische und Pädagogische Psychologie*, 36(3), 147-159.

Schröder, E. (1989). *Vom konkreten zum formalen Denken – Individuelle Entwicklungsverläufe von der Kindheit zum Jugendalter*. Bern: Verlag Hans Huber.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Troche, S., Rammstedt, B. & Rammsayer T. (2002). Vergleich einer Papier-Bleistift- und einer computergestützten Version des Leistungsprüfsystems (LPS). *Diagnostica*, 48(3), 115-120.

Wagner-Menghin, M.M. (2004). *Lexikon-Wissen-Test (LEWITE)* [Software und Manual]. Mödling: Schuhfried.

Wechsler, D. (1944). *Measurement of adult intelligence*. Baltimore: Williams & Wilkins.

Wessels, M.G. (1994). *Kognitive Psychologie* (3., verb. Aufl.). München: E. Reinhardt.

Lösungsvorschlag

C und F

Testleiter:

*Diese Antwort ist richtig. Gut gemacht.
Lest nun die erste Seite leise fertig und
versucht die nächsten Aufgaben alleine zu
lösen*

C aber nicht F

*Die Antwort ist leider nicht ganz richtig,
„Hören“ bedeutet zwar sehr wohl ein
Geräusch aufzunehmen, jedoch nicht mit
den *** sondern mit den Ohren.
Lest nun die erste Seite leise fertig und
versucht die nächsten Aufgaben alleine zu
lösen*

F aber nicht C

*Diese Antwort ist leider nicht ganz richtig,
„Hören“ bedeutet zwar etwas mit den
Ohren aufzunehmen, jedoch kein/en ***
sondern ein Geräusch.
Lest nun die erste Seite leise fertig und
versucht die nächsten Aufgaben alleine zu
lösen*

Nicht F und nicht C

*Diese Antwort ist leider nicht richtig. Lest
euch den Satz bitte noch einmal genau
durch, vielleicht findet ihr die richtige
Lösung. (Falls die richtige Lösung nicht
gefunden wird, erfolgt die Nennung der
richtigen Lösung durch den Testleiter)
Lest nun die erste Seite leise fertig und
versucht die nächsten Aufgaben alleine zu
lösen*

Fortsetzung schriftliche Instruktion:

Versuche nun die anderen Aufgaben alleine zu lösen.

Falls du ein Wort nicht kennst oder dir eine Beschreibung zu schwer fällt, kannst du dieses überspringen und zum nächsten gehen.

Testbogen A

A01) Kleidung

Ein *a) Gebäude* der/das zur Bedeckung des menschlichen *e) Auges* dient.
b) Textil *f) Besitzes*
c) Schmuck *g) Eigentums*
d) Gefäß *h) Körpers*

A02) Fragen

Jemanden um/über eine *a) Ermahnung* über/für/auf etwas *e) berichten*.
b) Antwort *f) zwingen*
c) Verzeihung *g) bitten*
d) Durchsicht *h) belügen*.

A03) Manager

Ein Manager ist ein/e *a) Computerfachmann* in einem *e) christlichen*
b) Spion *f) kriminellen*
c) Führungskraft *g) wirtschaftlichen*
d) Netzwerk *h) riskanten*

Unternehmen arbeitet.

A04) Gram

Gram bezeichnet eine *a) Sorge über* etwas und einen Zustand der/des
b) Aggression gegen
c) Freude über
d) Neugierde für

e) Zornes.
f) Bedrückung
g) Überschwenglichkeit
h) Unruhe

A05) Artist

Ein Artist ist ein *a) Adelliger* der zumeist in/im *e) Schulen* auftritt.
b) Philosoph *f) Universitäten*
c) Schausteller *g) Zirkus*
d) Schauspieler *h) Restaurants*

A06) Vergeltung

Vergeltung bedeutet ein/e *a)Wiederholung* im Sinne einer *e)Weltanschauung*
b)Verbieten *f)Übung*
c)Vorhersehung *g)Pflicht*
d)Reaktion *h)Gegenleistung*

von eine/r vorhergegangenen Handlung.

A07) Helikopter

Ein Helikopter ist ein/e *a)Ballon*, das/der/die durch eine/n motorisch
b)Einsatzwagen
c)Unterwasserboot
d)Luftfahrzeug

betriebene *e)Reifenachse* angetrieben wird.

f)Rotor
g)Heliumspender
h)Schraube

A08) Sanierung

Eine Sanierung bezeichnet die *a)Verbreitung* einer/s *e)Person*.

b)Heilung *f)Gegenstandes*
c)Bekämpfung *g)Krankheit*
d)Erneuerung *h)Bevölkerung*

A09) Argwöhnisch

Eine argwöhnische Person *a)verwöhnt* andere Personen und denkt

b)ärgert
c)ignoriert
d)verdächtigt

e)Schönes von ihnen.

f) Gutes
g)nichts
h)Übles

A10) Symbiose

Symbiose bezeichnet das *a)Kennenlernen* unterschiedlicher Arten, das für

b)sich Bekriegen
c)Zusammenleben
d)Auslöschen

alle Beteiligten *e)bedenklich* ist.
f)vorteilhaft
g)negativ
h)ungewohnt

A11) Greis

Greis ist die Bezeichnung für eine sehr *a)alte* *e)Pflanzenart.*
b)helle *f)Behausung*
c)romantische *g)Kunstrichtung*
d)bekannte *h)Person*

A12) Besonnen

Eine besonnene Person handelt *a)beschwingt* und vermeidet *e)langweilige*
b)überlegt *f)anstrengende*
c)emotional *g)vorschnelle*
d)kaum *h)ernste*

Taten.

A13) Keller

Ist ein zumindest zum Teil unter der/m *a)Dachstuhl* liegender *e)Ziegel*
b)Horizont *f)Gegenstand*
c)Garten *g)Kasten*
d)Erde *h)Raum*

A14) Auszeichnung

Eine Auszeichnung ist eine Form der öffentlichen *a)Verleugnung* besonderer
b>Ehrung
c)Bestrafung
d)Bekundung

e)Ereignisse.
f)Verdienste
g)Fehlritte
h)Umstände

A15) Dämmerung

Als Dämmerung bezeichnet man den *a) unfassbaren* Übergang zwischen *e) Gut*
b) ruckartigen *f) Tag*
c) plötzlichen *g) Wald*
d) fließenden *h) Genie*

und Böse.
und Nacht
und Wiese
und Wahnsinn

A16) Diktieren

Diktieren bedeutet etwas *a) anzudeuten* mit dem Sinn, dass es jemand anderer
b) auszubessern
c) vorzuzeigen
d) anzusagen

e) aufschreiben muss.
f) erkennen
g) lernen
h) verstehen

A17) Ehrgeizig

Eine ehrgeizige Person gilt als *a) ehrenvoll* und bemüht ihr/e *e) Minimum*
b) bodenständig *f) Freunde*
c) fleißig *g) Pension*
d) treu *h) Ziele*

zu erreichen.

A18) Vermiesen

Vermiesen bedeutet jemand anderen die/den *a) Furcht* von/über/an etwas
b) Ärger
c) Erinnerung
d) Freude

zu *h) erhöhen.*
e) nehmen.
f) wecken
g) geben

A19) Apathisch

Eine apathische Person zeigt sich gegenüber *a)zweideutigen* Reizen
b)chemischen
c)äußeren
d)inneren

e)beeinträchtigt.
f)teilnahmslos
g)verängstigt
h)aggressiv

A20) Wehmütig

Bezeichnet ein Gefühl zarter(n) *a)Angst,* hervorgerufen durch
b)Traurigkeit
c)Mutes
d)Ärgers

e)Erwartungen.
f)Verletzungen
g)Erinnerungen
h)Mitmenschen

B07) Wettkampf

Beim sportlichen Wettkampf messen *a)mehrere* Teilnehmer ihre *e)Beinlängen*
b)große *f)Leistungen*
c)fremdsprachige *g)Wohnungsgrößen*
d)unerlaubte *h)Beliebtheit*

gegeneinander.

B08) Misstrauisch

Eine misstrauische Person *a)glaubt* eine(r) andere(n) Person *e)gerne*.
b)verhöhnt *f)häufig*
c)belügt *g)nicht*
d)hintergeht *h)immer*

B09) Massage

Die Massage ist eine *a)unbewusste* Beeinflussung von Haut, Bindegewebe und
b)chemische
c)mechanische
d)psychische

Muskel und dient zur Lösung von *e)Verspannungen*.
f)Schadstoffen.
g)Ängsten.
h)Juckreizen.

B10) Emsig

Eine emsige Person ist *a)glücklich* und *e)lustig*.
b)faul *f)arbeitswillig*
c)fleißig *g)arbeitsunwillig*
d)aggressiv *h)ausgelassen*

B11) Despotisch

Ein despotischer Mensch verhält sich *a)penibel* und *e)belehrt* seine
b)herrisch *f)beträgt*
c)hinterlistig *g)dominiert*
d)scheinheilig *h)belauscht*

Mitmenschen in ungerechtfertigtem Ausmaß.

B12) Traditionell

Traditionelles Handeln zeichnet sich durch die/den/das *a)Gegensatz* von *e)Einfällen*
b)Konkurrenz *f)Mitgliedern*
c>Weitergabe *g)Gesetztexten*
d)Wählen *h)Handlungsmustern*

innerhalb einer Gruppe aus

B13) Jahreszeit

Die Jahreszeit unterteilt das Jahr in *a)bestimmte Jahreskreise* und zeichnet sich
b)mehrere Stunden
c)verschiedene Perioden
d)gewissen Sonnenwenden

durch bestimmte *e)wetterunabhängige* Eigenschaften aus.
f)tagesabhängige
g)unscheinbare
h)klimatische

B14) Triumph

Triumph bedeutet ein/e/n *a)Verlust* im Kampf zu haben und eine/n *e)Gleichstand*
b)Nachteil *f)Niederlage*
c)Erfahrung *g)Sieg über*
d)Erfolg *h)Friedensschluss*

mit den/m Gegner.
gegen

mit

B15) Journalist

Ein Journalist ist für die/das *a)Verbreitung* von Informationen durch *e)Führungen*
b)Erfinden *f)Vorträge*
c)Bewertung *g)Massenmedien*
d)Unterdrücken *h)Tagungen*

zuständig.

B16) Möbel

Bezeichnet alle *a) günstigen* *e) Elektrogegenstände* in einem Raum.
b) dunklen *f) Holzgegenstände*
c) beweglichen *g) Dekorationen*
d) schweren *h) Einrichtungsgegenstände*

B17) Professor

Ein Professor ist meist in einer *a) Hausverwaltung* tätig und ist für die
b) Justizanstalt
c) Kirche
d) Hochschule

e) Bekehrung zuständig.
f) Straftätigen
g) Lehre
h) Vereinigung

B18) Melancholie

Melancholie bezeichnet einen Zustand von *a) Misstrauen* und/aber *e) Frustration*.
b) Zorn *f) Traurigkeit*
c) Angst *g) Getriebenheit*
d) Schwermut *h) Faulheit*

B19) Vorhersehung

Vorhersehung bezeichnet eine/n höhere/n *a) Macht* die/der/ das *e) Kapital* der
b) Wissenschaft *f) Schicksal*
c) Gesetzgebung *g) Gesundheit*
d) Staatsbeschluss *h) Grundrecht*

Menschen beeinflusst.

Testbogen C

C01) Spielen

Eine *a) langweilige* Tätigkeit die besonders *e) Kinder* oft durchführen.

- | | |
|------------------------|----------------------|
| <i>b) erzwungene</i> | <i>f) Lehrer</i> |
| <i>c) angenehme</i> | <i>g) Könige</i> |
| <i>d) anstrengende</i> | <i>h) Erwachsene</i> |

C02) Leopard

Der Leopard gehört zur Familie der *a) Edelsteine* und ist ein *e) Beutetier*.
b) Menschenaffen *f) Raubtier*
c) Grünpflanzen *g) Heilmittel*
d) Katzen *h) Schutzsymbol*

C03) Jähzornig

Eine jähzornige Person wird *a) langsam* oder aus *e) außerordentlichem* Anlass
b) vorhersehbar *f) wichtigem*
c) unvermittelt *g) besonderem*
d) begründet *h) kleinem*

zornig gegenüber einer anderen Person oder Sache.

C04) Schwerkraft

Die Schwerkraft wird auch *a) Übergewicht* genannt und bewirkt, dass Gegenstände
b) Boxkategorie
c) Kraftreserve
d) Erdanziehung

- e) verändert werden.*
- f) beschädigt werden.*
- g) getragen werden.*
- h) zu Boden fallen.*

C05) Egoistisch

Egoistisch bedeutet sich *a) unüberlegt* und *e) rücksichtslos* zu verhalten
b) selbstständig *f) sozial*
c) eigennützig *g) kriminell*
d) freundschaftlich *h) gefährlich*

C06) Symptom

Ein Symptom ist ein/e *a) Art* einer *e) Geheimsprache*.
b) Fortschritt *f) Krankheit*
c) Zeichen *g) Gattung*
d) Regel *h) Religion*

C07) Hort

Ein Hort stellt einen *a) sicheren* Ort dar, in der Schule werden hier Kinder von
b) geheimnisvollen
c) gefährlichen
d) verwunschenen

Erziehen *e) verzaubert*.
f) therapiert.
g) betreut.
h) versteckt.

C08) Gelassen

Ein gelassener Mensch hat eine/n innere/n *a) Überzeugung* und die Fähigkeit auch in
b) Ruhe
c) Vorstellungskraft
d) Unsicherheit

schwierigen Situationen die/den *e) Freundschaft* zu bewahren.
f) Fassung
g) Arbeitsplanung
h) Rhythmus

C09) Nostalgie

Nostalgie ist eine *a) wehmütige* Hinwendung an vergangene Zeiten, welche oftmals
b) widerwillige
c) ängstliche
d) erzwungene

e) vergessen werden.
f) verleugnet
g) verhöhnt
h) idealisiert

C10) Plantage

Eine Plantage ist ein *a)adeliger* Großgrundbesitz und dient zur/m
b)landwirtschaftlicher
c)naturgeschützter
d)industrieller

e)Erforschung von Produkten.
f)Vertreibung
g)Erzeugung
h)Schutz

C11) Verhöhn

Verhöhn bedeutet jemand anderen zu *a)verwunden* und/aber zu *e)vergöttern*.
b)ehren *f)bewundern*.
c)erniedrigen *g)erpressen*
d)fürchten *h)verspotten*

C12) Laufen

Bedeutet sich mit *a)kleinen* Schritten zu *e) verletzen*.
b) bedachten *f) verrenken*
c) schweren *g) bewegen*
d) schnellen *h) verbessern*

C13) Verleugnen

Verleugnen bedeutet nicht zu einer Person *a)schreiten*, obwohl man diese *e)fürchtet*.
b)stehen *f)kennt*
c)kommen *g) achtet*
d)laufen *h) liebt*

C14) Traurig

Ist ein Gefühl, das meist wegen eines *a)unerwünschten* Ereignisses ausgelöst wird und
b)schönen
c)wunderbaren
d)sensationellen
durch *e) Jähzorn* gekennzeichnet ist.
f) Langeweile
g) Bedrücktheit
h)Eifer

C15) Kredit

Ein Kredit ist die/das/der *a)Wert* von Geld auf/durch *e)ewig.*
b)Schenken *f)Zwang.*
c)Verlust *g)Zeit*
d)Überlassung *h)Börsenentwicklungen*

C16) Reporter

Ein Reporter ist ein/e *a)Rundfunk*, der/die aktuelle Ereignisse *e)erstellt.*
b)Maschine *f)überträgt*
c)Computer *g)berichtet*
d)Person *h) sendet*

C17) Urkunde

Eine Urkunde ist ein/e *a)Telegramm*, das/die einen bestimmten Tatbestand
b)Dokumentationsfilm
c)Schriftstück
d)Fundstück

oder Sachverhalt fixiert und zumeist ihren *e)Erfinder* erkennen lässt.
f)Aussteller
g)Entdecker
h)Regisseur

C18) Testament

Ein Testament ist eine Regelung einer Person über ihre *a)Vermögen*, die im Falle ihres/r
b)Rechte
c)Ausgaben
d)Schulden

e)Heirat in Kraft tritt.
f)Scheidung
g)Todes
h)Verurteilung

C19) Radikal

Radikal bedeutet an etwas *a)Giftigem* festhalten und das nicht *e)Kennen* anderer
b)Schönem *f)Fürchten*
c)Vergessenem *g)Akzeptieren*
d)Extremen *h)Scheuen*

Denkweisen.

C20) Frustration

Frustration entsteht unter anderem wenn ein/e *a)Frage* nicht *e)verboten*
b)Bedingung *f)gestellt*
c)Wunsch *g)aufbereitet*
d)These *h)erfüllt*

wird.

Testbogen D

D01) Schwach

Schwach bedeutet über *a)manche* Kraft und *e)Energie* zu verfügen.
b)viel *f)Wohlwollen*
c)einige *g) Vertrauen*
d)wenig *h) Zeit*

D02) Singen

a) Leise Laute mit der/dem *e) Herzen* produzieren.
b) Laute *f) Stimme*
c) Musikalische *g) Instrument*
d) Lange *h) Saite*

D03) Erstaunt

Eine Emotion, die *a) sich reduziert* wenn etwas *e)Schönes* geschieht.
b)verschwindet *f) Trauriges*
c)sich verändert *g) Unerwartetes*
d) auftritt *h) Ärgerliches*

D04) Protektion

Bedeutet das *a)Bewundern* oder *e)Beobachten* einer Person.
b)Imitieren *f)Isolieren*
c)Schützen *g)Begünstigen*
d)Beschatten *h)Verehren*

D05) Weltall

Mit Weltall wird der Raum *a)westlich* der/des *e)Erdatmosphäre* bezeichnet.
b)innerhalb *f)Antarktis*
c)außerhalb *g)Atlantiks*
d)nördlich *h) Erdkugel*

D06) Anständig

Eine anständige Person verhält sich den herrschenden *a)Idolen* entsprechend und
b)Trends
c)Politikern
d)Sitten

behandelt seine Mitmenschen *e)respektvoll* .
f)kritisch
g)erträglich
h)neutral

D07) Irreversibel

Irreversibel bedeutet, dass ein Zustand *a)dynamisch* und nicht mehr *e)unscheinbar*
b)wechselhaft *f)erkennbar*
c)dauerhaft *g)umkehrbar*
d)schädlich *h)einschätzbar*

ist.

D08) Euphorie

Die Euphorie ist eine *a)bedrückte* Gemütsverfassung mit allgemeinen/m/r
b)überschwängliche
c)nervöse
d)ruhige

e)Gedanken
f)Traurigkeit
g)Hektik
h)Hochgefühl

D09) Planet

Ein Planet ist ein *a)Kontinent*, der/das eine näherungsweise *e)quaderförmige*
b)Himmelskörper *f)rechteckigen*
c)Bauwerk *g)kugelförmigen*
d)Gewächs *h)zylindrischen*

Gestalt hat.

D10) Ruine

Eine Ruine bezeichnet ein *a)vergessene/s* *e)Schriftstück* und dessen Überreste.
b)zerfallene/s *f)Bauwerk*
c)verweste/s *g)Erdreich*
d)steinzeitliches *h)Schmuckstück*

D11) Fohlen

Fohlen ist die Bezeichnung für ein/e *a)Züchtung* der Familie der *e)Hunde*.
b)Leittier *f)Katzen*
c)Jungtier *g)Pferde*
d)Krankheit *h)Schmetterlinge*

D12) Thermoskanne

Sind verschließbare *a)Heizkesseln* die Suppen oder Getränke *e)erhitzen* können.
b)Maschinen *f)warmhalten*
c)Gefäße *g)mixen*
d)Elektrogeräte *h)verdicken*

D13) Burn Out

Burn Out bezeichnet eine/n besonders ausgeprägte/n berufliche/n und/oder familiäre/n

- a) *Verletzung* aufgrund von e) *Unachtsamkeit*.
b) *Streit* f) *Gerüchten*
c) *Veränderung* g) *Überbelastung*
d) *Erschöpfung* h) *Missverständnissen*

D14) Autogramm

Ein Autogramm ist ein/e a) *eigenhändige/r/s* e) *Gemälde*.
b) *gefälschte/r/s* f) *Grafik*
c) *kopierte/r/s* g) *Unterschrift*
d) *veraltete/r/s* h) *Vertrag*

D15) Gedenktag

Ein Gedenktag ist ein Kalendertag an dem ein bestimmte/s a) *Essen* e) *erfunden*
b) *Ereignis* f) *aufgedeckt*
c) *Tier* g) *gegessen*
d) *Gedanke* h) *erinner*

wird.

D16) Vertrag

Ein Vertrag a) *unterbindet* das soziale Verhalten durch eine gegenseitige
b) *verletzt*
c) *regelt*
d) *erschafft*

- e) *Inspiration* zweier oder mehrer Parteien.
f) *Beleidigung*
g) *Bekriegung*
h) *Abmachung*

D17) Beet

Ein Beet ist eine a) *felsige* Fläche und dient dem (An)bau von e) *Sozialbauten*.
b) *betonierte* f) *Häusern*.
c) *erdige* g) *Stützstationen*.
d) *ungenützte* h) *Pflanzen*

D18) Neutral

Neutral bedeutet a) *formlos* beziehungsweise e) *neu* zu sein.
b) *unparteiisch* f) *ungenau*
c) *unökonomisch* g) *ungeladen*
d) *patriotisch* h) *positiv*

D19) Silo

Das Silo ist ein großer/s *a) Gefängnis*, der/das unter anderem zur Aufbewahrung von
b) Speicher
c) Schiff
d) Stall

e) Menschen dient.

f) Tieren

g) Futtermitteln

h) Metallwaren

13.2. Funktionen Abstrahieren: Instruktion & Items

Schriftliche Instruktion

In dieser Aufgabe werden dir einige Wörter in folgender Form vorgegeben:

Einfamilienhaus

- | | |
|--------------|------------|
| a)Reihenhaus | d)Fuchsbau |
| b)Vogelnest | e)Miethaus |
| c)Hundehütte | |

Du sollst dir pro Aufgabe überlegen, welche 2 der 5 angegebenen Wörter (a-e) eine gemeinsame grundlegende Funktion erfüllen oder einen gemeinsamen Verwendungszweck wie das dick geschriebene Wort haben.

Wenn sich mehrere Wörter eine sehr allgemeine und zwei mit dem dick geschriebenen Wort eine speziellere, besondere Funktion teilen, wähle die besondere.

In dem oben genannten Beispiel ist die gemeinsame Verwendung, dass man darin wohnen kann. Das Reihenhaus und Miethaus haben aber noch mit dem Einfamilienhaus gemeinsam, dass darin Menschen wohnen können. Also ist die richtige Antwort Einfamilienhaus und Reihenhaus.

Das nächste Beispiel darfst du nun alleine lösen. Überlege dir, welche zwei Wörter dieselbe Verwendung wie das dick geschriebene Wort „Ofen“ haben. Vergiss nicht: Es sind immer genau zwei Wörter! Wenn du nur eines oder mehr als zwei Wörter ankreuzt, ist die Antwort leider falsch.

Wenn du dich für zwei Wörter entschieden hast, kreuze diese bitte an und zeige auf.

Ofen

- | | |
|-------------|-----------|
| a)Rauchfang | d)Lampe |
| b)Kamin | e)Heizung |
| c)Holz | |

Nachdem alle Kinder aufgezeigt haben, wird eine Testperson gebeten ihre Lösung zu nennen. Bevor der/die TestleiterIn anmerkt, ob die Antwort richtig oder falsch war, fragt er/sie ob jemand eine andere Lösung hat (*„Hat jemand etwas anderes angekreuzt?“*). Sobald keine anderen Lösungen mehr kommen, löst der/die TestleiterIn das Übungsbeispiel auf:

B und E wurde gewählt: *„Das ist Richtig. Gut gemacht. Kannst du mir/Könnt ihr mir auch sagen, warum du/ihr euch für diese Antwort entschieden hast/habt? Welche Funktion teilen sich ein Ofen, ein Kamin und eine Heizung?“*

Abwarten ob die Funktion des Erwärmens von einem Kind genannt wird. Falls nicht.

„Weiß niemand, was ein Ofen, ein Kamin und eine Heizung machen? Für was brauchen wir sie?“

Falls an dieser Stelle noch immer niemand die richtige Antwort nennt, erfolgt die Auflösung durch den Testleiter.

Andere Kombinationen wurden gewählt: *„Du hast dich/Ihr habt euch für X und Y entschieden. Kannst du mir/Könnt ihr mir sagen warum du/ihr euch für diese Antwort entschieden habt?“*

Antwort der Testperson –

- a) Wenn bei der Antwort eine Funktion berücksichtigt wird, jedoch nicht die richtige, grundlegende z.B. Ofen, Lampe, Kamin (weil sie Licht geben):
„Deine/Eure Überlegung ist verständlich, aber fällt dir/euch noch eine grundlegendere, wichtigere Funktion des Ofens ein, die auch zwei der anderen Wörter erfüllen?“
- b) Wenn bei der Antwort nicht auf die Funktion geachtet wird, sondern auf die räumliche Nähe, etc. z.B. Ofen, Rauchfang, Holz (man gibt Holz in den Ofen und über den Rauchfang kann der entstehende Rauch entweichen):
„Deine/Eure Überlegung ist verständlich, aber gib/gebt Acht: Aufgabe ist es jene Wörter zu wählen die eine gleiche Funktion erfüllen, die denselben Zweck haben. Denke/Denkt noch einmal gut nach.“

Nennung der richtigen Lösung, falls Testpersonen diese nicht finden

„Nun wollen wir noch eine Aufgabe gemeinsam lösen. Lest bitte dazu die Angaben auf dem Zettel“

Fortsetzung der schriftlichen Instruktion:

Das nächste Beispiel ist etwas schwieriger, vielleicht schaffst du trotzdem es zu lösen. Welche zwei der fünf angegebenen Wörter erfüllen eine gleiche grundlegende Funktion wie ein „Einkaufswagen“?

Einkaufswagen

- | | |
|---------------|------------|
| a)Regal | d)Geschäft |
| b)Straßenbahn | e)Zug |
| c)Rechnung | |

Wenn du dich entschieden hast und insgesamt zwei Wörter angekreuzt hast, zeige bitte auf.

Nachdem alle Kinder aufgezeigt haben, wird eine Testperson gebeten ihre Lösung zu nennen. Bevor der/die TestleiterIn anmerkt, ob die Antwort richtig oder falsch war, fragt er/sie ob jemand eine andere Lösung hat (*„Hat jemand etwas anderes angekreuzt?“*). Sobald keine anderen Lösungen mehr kommen löst der/die TestleiterIn das Übungsbeispiel auf:

B und E wurden gewählt: *„Das ist Richtig. Gut gemacht. Kannst du mir/Könnt ihr mir auch sagen, warum du/ihr euch für diese Antwort entschieden hast/habt? Welche Funktion teilen sich ein Einkaufswagen, Straßenbahn und ein Zug?“*

Abwarten ob die Funktion der Beförderung von einem Kind genannt wird. Falls nicht.

„Weiß niemand, für was man einen Einkaufswagen, eine Straßenbahn und einen Zug verwendet? Für was brauchen wir sie?“

Falls an dieser Stelle noch immer niemand die richtige Antwort nennt, erfolgt die Auflösung durch den Testleiter.

Andere Kombinationen wurden gewählt: „*Du hast dich/Ihr habt euch für X und Y entschieden. Kannst du mir/Könnt ihr mir sagen warum du/ihr euch für diese Antwort entschieden habt?*“ – Antwort der Testperson –

- a) Wenn bei der Antwort eine Funktion berücksichtigt wird, jedoch nicht die richtige, grundlegende z.B. Regal, Geschäft, Einkaufswagen (weil sie Lebensmittel aufbewahren): „*Deine/Eure Überlegung ist verständlich, aber fällt dir/euch noch eine grundlegendere Funktion des Einkaufswagens ein, die auch zwei der anderen Angaben erfüllen?*“
- b) Wenn bei der Antwort nicht auf die Funktion geachtet wird, sondern auf die räumliche Nähe, etc. z.B. Rechnung, Einkaufswagen, Regale (alles befindet sich bzw. erhält man in einem Kaufhaus): „*Deine/Eure Überlegung ist verständlich, aber gib/gebt Acht: Aufgabe ist es jene Wörter zu wählen die eine gleiche Funktion erfüllen, die denselben Zweck haben. Denke/Denkt noch einmal gut nach.*“

Nennung der richtigen Lösung, falls Testpersonen diese nicht finden.

„*Lest nun die erste Seite leise fertig und versucht die nächsten Aufgaben alleine zu lösen*“.

Fortsetzung der schriftlichen Instruktion:

Versuche nun die nächsten Aufgaben alleine zu lösen. Kreuze dabei die von dir gewählten Antworten an. Falls dir eine Aufgabe zu schwer fällt, kannst du diese überspringen und zur Nächsten gehen.
--

Testbogen A

E01)Apfel

- a)Schokolade
- b)Birne
- c)Pommes

- d)Zwetschke
- e)Eis

E02)Lockenstab

- a)Kamm
- b)Schere
- c)Rundbürste

- d)Lockenwickler
- e)Pinzette

E03)Bikini

- a)Taucherbrille
- b)Strand
- c)Badehaube

- d)Wasserball
- e)Meer

E04)Autobahn

- a)Flugschneise
- b)Bahnhof
- c)Schienen

- d)Parkplatz
- e)Tankstelle

E05)Frauenhaus

- a)Museum
- b)Waisenhaus
- c)Gefängnis

- d>Burg
- e)Flüchtlingsheim

E06)Räucherstäbchen

- a)Zigarette
- b)Duftöl
- c)Rauchmelder

- d)Raumspray
- e)Feuerzeug

E07)Mikrowelle

- a)Messer
- b)Dampfgarer
- c)Wasserkocher

- d)Mixer
- e)Kühlschrank

E08)Erdöl

- a)Gas
- b)Essig
- c)Kohle

- d)Geld
- e)Gewürz

E09)Brief

- a)E-Mail
- b)Buch
- c)Fax

- d)Handy
- e)Film

E10)Uhr

- a)Kalender
- b)Armband
- c)Sonnenstand

- d)Schlüsselanhänger
- e)Sternwarte

E11)Geld

- a)Kreditkarte
- b)Führerschein
- c)Glückspfennig

- d)Pass
- e)Scheck

E12)Leiter

- a)Boden
- b)Aufzug
- c)Stiege

- d)Dach
- e)Gebäude

E13)Taschentuch

- a)Ohrenstäbchen
- b>Tischtuch
- c)Wattepad

- d)Karton
- e)Zeitung

E14)Eishockey

- a)Tennis
- b)Golf
- c)Basketball

- d)Polo
- e)Volleyball

E15)Sauerstoff

- a)Zitrone
- b)Wasser
- c)Kleidung

- d)Nahrung
- e>Süßigkeit

Testbogen B

F01) Weste

- a) Tasche
- b) Uhr
- c) Schuhe

- d) Mantel
- e) Jacke

F02) Mikrophon

- a) Videorekorder
- b) Lautsprecher
- c) Hörgerät

- d) Antenne
- e) Scheinwerfer

F03) Regenschirm

- a) Badehaube
- b) Schal
- c) Gummistiefel

- d) Stirnband
- e) Stöckelschuhe

F04) Klobürste

- a) Nagellack
- b) Gesichtscreme
- c) Zahnseide

- d) Ohrenstäbchen
- e) Kamm

F05) Förster

- a) Schriftsteller
- b) Polizist
- c) Reinigungskraft

- d) Maler
- e) Maurer

F06) Streichholz

- a) Pinsel
- b) Feuerzeug
- c) Bleistift

- d) Bunsenbrenner
- e) Zahnstocher

F07) Blindenhund

- a) Zebrastreifen
- b) Hauskatze
- c) Hörgerät

- d) Rollstuhl
- e) Auto

F08) Zeitung

- a) Rundfunk
- b) Roman
- c) Oper

- e) TV-Nachrichten
- f) Spielfilm

F09) Thermoskanne

- a) Kühlbox
- b) Gefrierschrank
- c) Tasse

- d) Schwimmbad
- e) Wasserkocher

F10) Springschnur

- a) Murmeln
- b) Trampolin
- c) Ball

- d) Sprungbrett
- e) Sandschaufel

F11) Wasserschlauch

- a) Regentonne
- b) Kanal
- c) Badewanne

- d) Wasserbecken
- e) Regenrinne

F12) Videothek

- a) Fleischerei
- b) Bibliothek
- c) Elektrohandel

- d) Trafik
- e) Bank

F13) Ventilator

- a) Hubschrauber
- b) Klimaanlage
- c) Heizung

- d) Kreisel
- e) Stromkreis

F14) Ampel

- a) Lampe
- b) Bodenmarkierung
- c) Auto

- d) Straßenschild
- e) Verkehr

F15) Flasche

- a) Dose
- b) Flaschenöffner
- c) Tetrapack

- d) Kochtopf
- e) Mülleimer

Testbogen C

G01)Rucksack

- | | |
|----------|--------------|
| a)Koffer | d)Schal |
| b)Hose | e)Turnschuhe |
| c)Tasche | |

G02)Docht

- | | |
|----------------|-------------|
| a)Kerzenhalter | d)Glühfaden |
| b)Heizstab | e)Zündholz |
| c)Lampenschirm | |

G03)Präsident

- | | |
|-------------------|-----------------|
| a)Briefträger | d)Koch |
| b)Popstar | e)Bürgermeister |
| c)Klassensprecher | |

G04)Rauchfang

- | | |
|---------------|--------------|
| a)Dach | d)Ruß |
| b)Glockenturm | e)Dunstabzug |
| c)Auspuff | |

G05)Gemälde

- | | |
|----------------|----------|
| a)Lied | d)Pinsel |
| b)Papier | e)Modell |
| c)Theaterstück | |

G06)Wetterbericht

- | | |
|---------------------|-------------------|
| a)Todesanzeigen | d)Kreuzworträtsel |
| b)Horoskop | e)Reportage |
| c>Börsenspekulation | |

G07)Altenheim

- | | |
|----------------|---------------|
| a)Kaufhaus | d)Krankenhaus |
| b)Kinderkrippe | e) Spielplatz |
| c)Arbeitsplatz | |

G08)Trafik

- a)Arbeitsamt
- b)Kiosk
- c)Bibliothek

- d)Supermarkt
- e)Schule

G09)Blut Spenden

- a)Wiederbelebung
- b)Krankenhaus
- c)Operation

- d)Krankheit
- e)Erste Hilfe

G10)Fotoapparat

- a)Album
- b)Videokamera
- c)Tonband

- d)Comic
- e)Fernsehapparat

G11)Löffel

- a)Rechen
- b)Schaufel
- c)Heugabel

- d)Reibe
- e)Messer

G12)Archiv

- a)Kiosk
- b>Weihnachtsmarkt
- c)Silo

- d)Weinkeller
- e)Bäckerei

G13)Klettverschluss

- a)Sohle
- b)Schuhband
- c)Schnalle

- d)Lasche
- e)Einlage

G14)Inlineskates

- a)Helm
- b)Rad
- c)Skateboard

- d)Schweißband
- e)Flugzeug

G15)Alarmanlage

- a)Gefängnis
- b)Sicherheitsschloss
- c)Schlüsselbund

- d)Hund
- e)Hauskatze

Testbogen D

H01) Besen

- a) Mixer
- b) Schwamm
- c) Nähmaschine

- d) Staubsauger
- e) Ofen

H02) Schlüssel

- a) Bleistift
- b) Nussknacker
- c) Korkenzieher

- d) Alarmanlage
- e) Tür

H03) Socken

- a) Handschuhe
- b) Schuhe
- c) Strumpfhose

- d) Bademantel
- e) Mütze

H04) Haarföhn

- a) Waschmittel
- b) Warmwasser
- c) Wäschespinne

- d) Waschmaschine
- e) Wäschetrockner

H05) Radfahren

- a) Joggen
- b) Inlineskaten
- c) Snowboarden

- d) Bergsteigen
- e) Skateboarden

H06) Schaufenster

- a) Werbung
- b) Möbel
- c) Vertreter

- d) Gartentor
- e) Tür

H07) Podest

- a) Vorhang
- b) Kran
- c) Teppich

- d) Stelzen
- e) Bürgermeister

H08)Biene

- a)Heuschrecke
- b)Kuh
- c)Henne

- d)Stachelschwein
- e)Tiger

H09)Kakao

- a)Banane
- b)Kaffee
- c)Cola

- d)Wasser
- e)Pommes

H10)Hass

- a)Liebe
- b)Gleichgültigkeit
- c)Vernunft

- d)Freude
- e)Gefühl

H11)Fingerabdruck

- a)Passfoto
- b)Zeichnung
- c)Schuh

- d)DNA
- e)Karikatur

H12)Internat

- a)Hort
- b)Schwimmbad
- c)Schlafzimmer

- d)Kindergarten
- e)Park

H13)Dolmetscher

- a)Diplomat
- b)Tourist
- c)Arzt

- d)Mediator
- e)Schriftsteller

H14)Schaumstoff

- a)Oropax
- b>Metallsplitter
- c)Spannfilz

- d)Seifenblasen
- e)Seide

H15)Schmuck

- a)Geld
- b)Brille
- c)Geschenk

- d)Schminke
- e)Tattoo

13.3. Itemleichtigkeitsparameter

Wortschatz Testform A

Item	Item- parameter	Standard- schätzfehler	Konfidenzintervall (p=0,95)	
beta Item1	2,175	0,222	1,740	2,610
beta Item2	1,179	0,197	0,793	1,566
beta Item3	0,948	0,193	0,569	1,328
beta Item4	-2,675	0,288	-3,240	-2,110
beta Item6	-2,100	0,246	-2,582	-1,617
beta Item8	1,893	0,213	1,475	2,311
beta Item9	-1,151	0,204	-1,551	-0,751
beta Item10	-0,829	0,196	-1,213	-0,444
beta Item11	0,151	0,187	-0,215	0,518
beta Item12	-1,068	0,202	-1,463	-0,673
beta Item13	2,989	0,255	2,490	3,488
beta Item14	0,873	0,192	0,496	1,250
beta Item15	0,543	0,189	0,173	0,913
beta Item16	1,101	0,196	0,718	1,485
beta Item17	1,672	0,208	1,265	2,079
beta Item18	-0,096	0,187	-0,463	0,271
beta Item19	-4,192	0,500	-5,172	-3,213
beta Item20	-1,415	0,213	-1,832	-0,998

Wortschatz Testform B

Item	Item- parameter	Standard- schätzfehler	Konfidenzintervall (p=0,95)	
beta Item1	0,049	0,183	-0,310	0,409
beta Item2	2,215	0,234	1,757	2,673
beta Item3	3,251	0,301	2,661	3,842
beta Item4	0,358	0,185	-0,004	0,721
beta Item5	-3,019	0,297	-3,602	-2,436
beta Item6	1,580	0,208	1,172	1,989
beta Item7	2,215	0,234	1,757	2,673
beta Item8	0,324	0,185	-0,038	0,686
beta Item10	-0,960	0,192	-1,337	-0,584
beta Item11	-3,112	0,306	-3,711	-2,512
beta Item12	-0,671	0,187	-1,038	-0,304
beta Item13	-1,479	0,207	-1,884	-1,075
beta Item14	1,808	0,216	1,384	2,232
beta Item15	0,186	0,184	-0,174	0,546
beta Item16	-1,394	0,204	-1,793	-0,995
beta Item17	1,904	0,220	1,473	2,335
beta Item18	-2,932	0,290	-3,500	-2,364
beta Item19	-0,325	0,184	-0,685	0,036

Wortschatz Testform C

Item	Item- parameter	Standard schätzfehler	Konfidenzintervall (p=0,95)	
beta Item1	0,898	0,192	0,521	1,275
beta Item3	-0,890	0,188	-1,258	-0,521
beta Item4	0,602	0,187	0,235	0,968
beta Item5	0,638	0,188	0,270	1,006
beta Item6	-0,433	0,183	-0,791	-0,074
beta Item7	2,111	0,233	1,655	2,567
beta Item8	0,389	0,184	0,027	0,750
beta Item9	-3,832	0,385	-4,587	-3,076
beta Item10	0,566	0,187	0,200	0,931
beta Item11	-0,433	0,183	-0,791	-0,074
beta Item12	2,738	0,269	2,212	3,264
beta Item13	-1,306	0,197	-1,692	-0,920
beta Item14	1,215	0,200	0,823	1,607
beta Item15	-1,639	0,207	-2,046	-1,233
beta Item16	1,562	0,211	1,149	1,975
beta Item17	-0,926	0,189	-1,296	-0,556
beta Item18	0,424	0,185	0,062	0,786
beta Item19	-1,684	0,209	-2,093	-1,274

Wortschatz Testform D

Item	Item- parameter	Standard- schätzfehler	Konfidenzintervall (p=0,95)	
beta Item1	1,848	0,242	1,374	2,322
beta Item2	1,177	0,213	0,760	1,594
beta Item3	0,522	0,194	0,141	0,903
beta Item4	-4,843	0,514	-5,851	-3,835
beta Item5	0,092	0,187	-0,275	0,460
beta Item6	0,024	0,187	-0,342	0,390
beta Item7	-4,843	0,514	-5,851	-3,835
beta Item8	-2,453	0,235	-2,913	-1,993
beta Item9	1,516	0,226	1,073	1,960
beta Item10	1,516	0,226	1,073	1,960
beta Item11	2,308	0,268	1,783	2,834
beta Item12	1,088	0,210	0,676	1,499
beta Item13	-0,575	0,184	-0,937	-0,214
beta Item14	1,909	0,245	1,428	2,389
beta Item15	0,232	0,189	-0,139	0,603
beta Item16	-0,311	0,185	-0,673	0,050
beta Item17	1,677	0,233	1,219	2,134
beta Item18	-2,348	0,229	-2,797	-1,899
beta Item19	1,465	0,224	1,026	1,904

Funktionen Abstrahieren Testform A

Item	Item- parameter	Standard- schätzfehler	Konfidenzintervall (p=0,95)	
beta Item1	2,979	0,452	2,092	3,866
beta Item3	-0,010	0,188	-0,378	0,359
beta Item4	-1,351	0,171	-1,687	-1,015
beta Item5	-0,703	0,173	-1,043	-0,363
beta Item6	-0,549	0,176	-0,893	-0,205
beta Item7	0,262	0,198	-0,125	0,649
beta Item8	0,478	0,207	0,072	0,883
beta Item9	0,668	0,216	0,244	1,092
beta Item10	-1,644	0,174	-1,985	-1,302
beta Item11	0,478	0,207	0,072	0,883
beta Item12	1,188	0,249	0,700	1,676
beta Item13	-0,485	0,177	-0,831	-0,139
beta Item14	-1,089	0,171	-1,424	-0,754
beta Item15	-0,222	0,182	-0,579	0,135

Funktionen Abstrahieren Testform B

Item	Item- parameter	Standard- schätzfehler	Konfidenzintervall (p=0,95)	
beta Item1	2,858	0,302	2,265	3,451
beta Item3	1,148	0,198	0,759	1,537
beta Item4	-0,016	0,173	-0,354	0,323
beta Item5	-0,819	0,173	-1,159	-0,480
beta Item6	1,322	0,205	0,920	1,724
beta Item7	-0,479	0,171	-0,815	-0,144
beta Item8	1,066	0,196	0,682	1,449
beta Item9	-0,757	0,173	-1,095	-0,418
beta Item11	0,338	0,177	-0,009	0,685
beta Item12	-2,119	0,210	-2,530	-1,707
beta Item13	-1,272	0,180	-1,625	-0,918
beta Item14	-0,914	0,174	-1,255	-0,572
beta Item15	-0,356	0,171	-0,692	-0,021

Funktionen Abstrahieren Testform C

Item	Item- parameter	Standard- schätzfehler	Konfidenzintervall (p=0,95)	
beta Item1	3,155	0,322	2,524	3,785
beta Item2	-0,932	0,178	-1,281	-0,583
beta Item3	2,342	0,244	1,863	2,821
beta Item4	1,586	0,202	1,191	1,981
beta Item5	-2,357	0,247	-2,840	-1,873
beta Item6	-2,232	0,237	-2,698	-1,767
beta Item7	-0,172	0,168	-0,502	0,157
beta Item8	1,019	0,182	0,662	1,376
beta Item9	-1,205	0,186	-1,569	-0,841
beta Item10	-0,803	0,175	-1,147	-0,460
beta Item11	-0,740	0,174	-1,081	-0,399
beta Item12	-0,617	0,172	-0,954	-0,280
beta Item13	0,235	0,169	-0,096	0,567
beta Item14	0,691	0,175	0,348	1,033
beta Item15	0,031	0,168	-0,299	0,360

Funktionen Abstrahieren Testform D

Item	Item- parameter	Standard- schätzfehler	Konfidenzintervall (p=0,95)	
beta Item1	3,676	0,396	2,900	4,453
beta Item2	0,212	0,169	-0,119	0,544
beta Item4	1,109	0,183	0,750	1,468
beta Item5	1,635	0,202	1,240	2,031
beta Item6	-0,641	0,175	-0,983	-0,298
beta Item7	-0,076	0,169	-0,407	0,255
beta Item8	-1,409	0,198	-1,798	-1,021
beta Item9	0,010	0,169	-0,320	0,341
beta Item10	-3,121	0,346	-3,799	-2,444
beta Item11	0,506	0,171	0,170	0,842
beta Item13	-1,252	0,192	-1,628	-0,877
beta Item14	-0,221	0,170	-0,554	0,111
beta Item15	-0,428	0,172	-0,764	-0,091

13.4. Personenparameter

Wortschatz Testform A

Rohwert	Personenparameter	Standardschätzfehler
0	-4,8077169	NA
1	-3,8698435	1,0842520
2	-3,0017687	0,8212782
3	-2,4213274	0,7135637
4	-1,9559544	0,6557227
5	-1,5496005	0,6220087
6	-1,1757698	0,6026921
7	-0,8190480	0,5931879
8	-0,4691171	0,5910144
9	-0,1181449	0,5947232
10	0,2403424	0,6035976
11	0,6127072	0,6177346
12	1,0063949	0,6384195
13	1,4323242	0,6687117
14	1,9085258	0,7145061
15	2,4683085	0,7873307
16	3,1819292	0,9134319
17	4,2428236	1,1840143
18	5,3871112	NA

Wortschatz Testform B

Rohwert	Personenparameter	Standardschätzfehler
0	-4,99923655	NA
1	-4,03733640	1,0948611
2	-3,14441336	0,8373883
3	-2,53512499	0,7349927
4	-2,03667381	0,6818303
5	-1,59419773	0,6509740
6	-1,18374651	0,6317009
7	-0,79307709	0,6193103
8	-0,41443095	0,6121910
9	-0,04135718	0,6103294
10	0,33296247	0,6143557
11	0,71631526	0,6251444
12	1,11797693	0,6437920
13	1,54958843	0,6717415
14	2,02632527	0,7112565
15	2,57091964	0,7682587
16	3,23006866	0,8653452
17	4,16355800	1,1094669
18	5,16344020	NA

Wortschatz Testform C

Rohwert	Personenparameter	Standardschätzfehler
0	-4,57887200	NA
1	-3,64571788	1,0823190
2	-2,78304410	0,8170538
3	-2,21142429	0,7059257
4	-1,75912354	0,6438907
5	-1,37053844	0,6055496
6	-1,01939925	0,5814044
7	-0,69044703	0,5670355
8	-0,37344517	0,5601305
9	-0,06060881	0,5594893
10	0,25478560	0,5646853
11	0,57949265	0,5760706
12	0,92153224	0,5950193
13	1,29213519	0,6245724
14	1,70952748	0,6708708
15	2,20769746	0,7467568
16	2,86085539	0,8819642
17	3,87637153	1,1713054
18	4,97920031	NA

Wortschatz Testform D

Rohwert	Personenparameter	Standardschätzfehler
0	-4,7696418	NA
1	-3,8981192	1,0520535
2	-3,0948208	0,7832709
3	-2,5720362	0,6742182
4	-2,1587495	0,6165814
5	-1,8003647	0,5835602
6	-1,4715926	0,5650854
7	-1,1578431	0,5566048
8	-0,8490654	0,5559663
9	-0,5370406	0,5623695
10	-0,2137660	0,5760347
11	0,1300428	0,5981897
12	0,5067167	0,6312250
13	0,9340739	0,6787548
14	1,4387309	0,7449401
15	2,0586029	0,8327014
16	2,8457623	0,9450553
17	3,8654734	1,0723178
18	5,1751169	1,2465182
19	6,5484663	NA

Funktionen Abstrahieren Testform A

Rohwert	Personenparameter	Standardschätzfehler
0	-4,25205670	NA
1	-3,21899655	1,1362172
2	-2,26722875	0,8540564
3	-1,65263366	0,7254410
4	-1,18157390	0,6528948
5	-0,78539165	0,6093177
6	-0,43108109	0,5834518
7	-0,09945961	0,5700709
8	0,22279425	0,5668921
9	0,54696456	0,5735071
10	0,88494888	0,5912745
11	1,25249528	0,6241768
12	1,67558167	0,6818200
13	2,20913124	0,7905511
14	3,02432407	1,0581156
15	3,90798603	NA

Funktionen Abstrahieren Testform B

Rohwert	Personenparameter	Standardschätzfehler
0	-4,2231334	NA
1	-3,1755837	1,1382719
2	-2,2041326	0,8730766
3	-1,5475803	0,7587186
4	-1,0221468	0,6962764
5	-0,5643338	0,6601214
6	-0,1428001	0,6407207
7	0,2625874	0,6348349
8	0,6690164	0,6426982
9	1,0961944	0,6677855
10	1,5734375	0,7192825
11	2,1589051	0,8227019
12	3,0269434	1,0834641
13	3,9637876	NA

Funktionen Abstrahieren Testform C

Rohwert	Personenparameter	Standardschätzfehler
0	-4,6516235	NA
1	-3,5982229	1,1370227
2	-2,6177662	0,8832306
3	-1,9406694	0,7723282
4	-1,3972796	0,7063876
5	-0,9300411	0,6634825
6	-0,5093756	0,6357079
7	-0,1165185	0,6195981
8	0,2627412	0,6137634
9	0,6411074	0,6182538
10	1,0321591	0,6345990
11	1,4534846	0,6665963
12	1,9328415	0,7230860
13	2,5262726	0,8287070
14	3,4047270	1,0881049
15	4,3523404	NA

Funktionen Abstrahieren Testform D

Rohwert	Personenparameter	Standardschätzfehler
0	-4,6941442	NA
1	-3,4669966	1,2293314
2	-2,3370420	0,9310514
3	-1,6104254	0,7872859
4	-1,0551620	0,7099106
5	-0,5838274	0,6673380
6	-0,1540034	0,6469589
7	0,2608773	0,6440210
8	0,6827983	0,6580545
9	1,1362571	0,6925786
10	1,6576655	0,7576421
11	2,3178256	0,8792799
12	3,0108442	NA
13	3,7038628	NA

13.5. Ansuchen an Landesschulrat

Johanna Böck
Uferstraße 2
4191 Vorderweißenbach

Landesschulrat Linz
Sonnensternstraße 20
4020 Linz

Ansuchen um Unterstützung eines universitären Forschungsprojekts

Sehr geehrte Frau Wagner!

Mit diesem Schreiben möchten wir Sie um die Unterstützung bei der Durchführung eines Forschungsprojekts im Rahmen unserer Diplomarbeit (Machbarkeitsstudie zum Thema AID2 als PC Version) an der Fakultät für Psychologie an der Universität Wien bitten.

Wir bräuchten dazu Ihre Genehmigung eine Testung im Umfang von einer Schulstunde an folgenden Schulen durchzuführen:

- Volksschule Vorderweißenbach (3. und 4. Schulstufe)
- Hauptschule Vorderweißenbach (5.-8. Schulstufe)
- Hauptschule Bad Leonfelden (5.-8. Schulstufe)
- Petrinum Linz (5.-9. Schulstufe)

Die Beschreibung des Forschungsprojekts und des eingesetzten Tests sowie ein Elternbrief liegen bei.

Mit freundlichen Grüßen

Böck Johanna

13.6. Beschreibung des Forschungsprojekts

Ziel dieses Forschungsprojekts ist eine Abklärung der Machbarkeit einer Computerversion des AID2.

Im Zuge unserer Diplomarbeit, welche als eine Voruntersuchung dient, soll für vier der elf Untertests abgeklärt werden, ob eine Vorgabe dieser in Form einer Gruppentestung ohne wesentliche Qualitätseinbußen z.B. Aufrechterhaltung der Rasch- Modell - Konformität machbar ist.

Bearbeitet wurden dabei folgende Untertests:

UT 1 Alltagswissen

UT 3 Angewandtes Rechnen

UT 6 Synonyme finden

UT 9 Funktionen abstrahieren

Die erstellten Items, welche jenen des Original AID2 nachempfunden sind, würden den Testpersonen in Form einer Paper-Pencil Testung vorgegeben werden.

Um möglichst generalisierbare Aussagen treffen zu können, benötigen wir für unsere Testung SchülerInnen der 3. und 4. VS, sowie der 1.-4. HS bzw. AHS und der 5. AHS bzw. BHS. Der Stichprobenumfang sollte dabei mindestens 800 betragen.

Sinn dieses Forschungsprojektes und einer Computerversion des AID2 ist es, ökonomische (Vor)-untersuchungen zu ermöglichen. Durch die Vorgabe der PC-Version in Gruppen wäre es auch möglich, im Bedarfsfall ganze Schulklassen zu testen.

Eingesetzte Verfahren

Bei dem eingesetzten Verfahren handelt es sich um eine Gruppentestung im Paper-Pencil Format.

Wobei die vorgegebenen Aufgaben jenen des AID2 nachempfunden, jedoch nicht mit diesem identisch, sind.

Den Testpersonen werden vier verschiedene Aufgabenstellungen vorgegeben:

- 1) Fragen zum Thema Alltagswissen werden im Multiple-Choice-Format vorgegeben und sollen von den Testpersonen beantwortet werden
- 2) Vorgabe von Textrechnungen, freies Antwortformat
- 3) Vorgabe eines Wortes von dem die richtige Definition erarbeitet werden muss
- 4) Vorgabe eines Wortes und fünf Antwortmöglichkeiten, die Testperson soll angeben, welche zwei der fünf möglichen Antworten eine zentrale Gemeinsamkeit mit dem vorgegeben Wort teilen

Zu Beginn der Testung erfolgt eine Instruktion durch die Testleiterin. Anschließend sollen die Kinder selbstständig arbeiten. Falls sich Fragen ergeben, ist die Testleiterin natürlich zu Stelle und verhält sich je nach Art der Frage entsprechend.

Die Testung wird eine Schulstunde in Anspruch nehmen.

13.7. Elternbrief

Sehr geehrte Damen und Herren, Liebe Eltern.

Wir wenden uns durch diesen Brief mit der Bitte an Sie, uns bei der Umsetzung eines Forschungsprojektes im Zuge unserer Diplomarbeit zu unterstützen.

Thema unserer Arbeit ist die Umsetzung des Adaptiven Intelligenzdiagnostikum 2 (AID2 – eine in der Schulleistungsdiagnostik häufig eingesetzte Testbatterie zur Erfassung kognitiver Teilfertigkeiten) in eine Computerversion desselbigen.

Eine PC-Version des AID2 würde eine zeitsparende und ökonomische Voruntersuchung bei Kindern und die Durchführung von Gruppenuntersuchungen in Klassen ermöglichen. Außerdem eröffnet eine Vorgabe mittels Computer neue Möglichkeiten der Aufgabengestaltung.

Um die Machbarkeit einer Computerversion des AID2 zu überprüfen, ist es notwendig durch Testungen in verschiedenen Schulen Daten zu erhalten und diese auszuwerten, wobei die Testungen im Zuge einer Schuleinheit durchgeführt werden.

Die Testung beinhaltet verschiedene Aufgaben, wie das Durchführen von Rechnungen, das Definieren von Wörtern etc.

Sie können uns und die Forschungsarbeit an der Universität Wien unterstützen, wenn sie sich als bereit erklären, Ihr Kind an dieser Testung teilnehmen zu lassen.

Vielen Dank im Voraus.

Mit freundlichen Grüßen

Böck Johanna und Eiter Angelika

Abschnitt zum Unterzeichnen

Ich erkläre mich bereit, dass mein/e Sohn/Tochter im Zuge einer
Schuleinheit an der von Studentinnen der Universität Wien durchgeführten Testung
teilnimmt.

Unterschrift

13.8. Abstract - Deutsch

In der heutigen Zeit ist die Psychologische Diagnostik aus vielen und sehr verschiedenen Bereichen nicht mehr wegzudenken. Mit den Anwendungsfeldern sind aber auch die Anforderungen an die Psychologische Diagnostik gewachsen. Sie soll einerseits Basis für eine objektive und valide Entscheidung sein, andererseits eine zeitsparende und ökonomische Vorgehensweise ermöglichen. Aus diesen Überlegungen wurde die Idee geboren, eine Gruppentestversion des Adaptiven Intelligenz Diagnostikums 2 (Version 2.2) von Kubinger (2009a) zu entwickeln und in dieser Diplomarbeit für die Untertests „Soziale und Sachliche Folgerichtigkeit“, „Synonyme finden“ und „Funktionen Abstrahieren“ realisiert. Eine Vorgabe der entwickelten Items fand für die beiden letzt genannten Untertests statt, wobei 679 SchülerInnen in Österreich getestet wurden. Im Anschluss an die Testung wurde eine Überprüfung der Konformität der Itempools mit dem dichotom-logistischen Modell nach Rasch mittels Likelihood-Quotienten-Test nach Andersen (1973) anhand der Teilungskriterien Median, Geschlecht und Region (Stadt-Land) durchgeführt. Diese konnte zumindest a posteriori nach Ausschluss von maximal zwei Items für alle vier pro Untertest entwickelten Testformen angenommen werden. In weiteren Untersuchungen sollte die Rasch-Modell-Konformität anhand einer größeren Stichprobe überprüft, sowie ein LLTM pro Untertest für die miteinander verbundenen Testformen durchgeführt werden. Lassen sich daraus verbindliche Regeln für die Itemkonstruktion ableiten, wäre für die Zukunft eine teil-automatische Itemgenerierung denkbar.

Schlüsselwörter: AID 2, Gruppentest, Itemgenerierung, Rasch-Modell

13.9. Abstract – English

These days psychological diagnostics has become essential in various scientific and societal areas. Parallel to its expansion in applications the requirements for psychological diagnostics have increased; on the one hand acting as a basis for objective and valide judgement and on the other hand enabling a timesaving and economical approach. To seriously deliberate these thoughts the aim of this study was to develop a test procedure of the AID 2 (version 2.2) by Kubinger (2009a) for the subtests “social and factual consistency” (“Soziale und Sachliche Folgerichtigkeit“), “discovering synonyms” (“Synonyme finden”) and “abstracting functionalities” (“Funktionen abstrahieren”) applicable on groups. The two last-mentioned subtests have been assessed by testing 679 students in Austria. A likelihood-quotient statistic (Anderson, 1973) with the splitting criteria median, sex and region has been conducted in order to verify the conformance of the item pools with the Rasch model. After the a posteriori elimination of a maximum of two items in the corresponding test versions of the subtests the conformance according to the Rasch model has been accepted. Future research should focus on Rasch model conformance by investigating larger sample sizes as well as the application of LLTM for associated test versions in the corresponding subtests. The potential of setting up binding rules in the phase of item development can have a beneficial influence on semi-automated item-generation.

Key words: AID 2, group setting, item-generation, Rasch model

13.10. Lebenslauf

Johanna Böck
Erdbergstraße 164/18
1030 Wien
0699 10123953
johanna.boeck@univie.ac.at

PERSÖNLICHE INFORMATIONEN

Nationalität	Österreich
Alter	23 Jahre (29/03/1986)
Geburtsort	Rohrbach
Eltern	Karl Hans Böck, Mag. Margareta Böck (geb. Mitterhofer)

AUSBILDUNG

2009-2010 Wien	Hauptuniversität Wien Studium der Psychologie
2008-2009 Paris	Université René Descartes, Paris 5 Auslandsaufenthalt, Studium der Psychologie Erhalt des Förderstipendiums des Landes Oberösterreich
2004-2008 Wien	Hauptuniversität Wien Studium der Psychologie Erhalt des Leistungsstipendiums 2004/05, 2005/06, 2006/07 und 2007/08
1995-2004 Linz	Gymnasium der Kreuzschwestern

PSYCHOLOGISCHE PRAKTIKA

Ab 2010 Wien	Kinderschutzzentrum
2009 3 Monate, Wien	Markt- und Meinungsforschung Telemark Marketing
2007 2 Monate, Wien	Justizanstalt Favoriten