



universität  
wien

# DIPLOMARBEIT

Titel der Arbeit

## Pilotstudie zu vier Untertests des AID als Gruppenversion

Verfasserin

Petra Nicola Hofmayer

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2012

Studienkennzahl lt. Studienblatt: A 298

Studienrichtung lt. Studienblatt: Psychologie

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

## Danksagung

Wenn man Jahre seines Lebens dem Studium widmet, sich von einem kleinen Ziel zum nächsten arbeiten, und dann plötzlich der Zeitpunkt gekommen ist an dem keine Prüfung mehr zu machen und auch kein Seminar mehr übrig geblieben ist, wird schlagartig klar:

Eine Diplomarbeit ist zu schreiben.

Am Entstehungsprozess dieser Arbeit waren verschiedenste Personen beteiligt, die an dieser Stelle genannt werden sollen.

Ich möchte an erster Stelle meinem Diplomarbeitsbetreuer, Univ. Prof. Mag. Dr. Kubinger, für die Vergabe dieses interessanten und, was mir besonders wichtig war, forschungs- und praxisrelevanten Themas. Ich danke für seine Aufmerksamkeit, seine Anregungen und Kritik. Außerdem danke ich Mag.<sup>a</sup> Bettina Hagenmüller für die Durchsicht der Arbeit.

Weiterer Dank gilt meinen Studienkollegen und Studienkolleginnen, auf deren Kollegialität und Unterstützung ich bei der Datenerhebung in den oberösterreichischen Schulen und auch bei der Auseinandersetzung mit dem Programm „R“ zählen konnte. Auch meinen Freundinnen und Studienkolleginnen Daniela und Johanna sei an dieser Stelle mein Dank für die jahrelange Wegbegleitung im Studium und die gemeinsame Bewältigung von studiumsbedingten fachlichen und persönlichen Hürden ausgesprochen.

Auch meiner Familie sei für die persönliche Unterstützung und ihren Glauben an meinen Erfolg gedankt. Spezieller Dank gilt meinem Vater, der die Bürde des Korrekturlesens übernommen hat. Diesbezüglich danke ich auch meiner Freundin Julia, die sich auf Fehlersuche begeben hat und mich auf inhaltliche und logische Ungereimtheiten hingewiesen hat.

Für die unkomplizierte organisatorische Abwicklung möchte ich auch dem Landesschulrat Oberösterreich, dem Petrinum Linz, der Hauptschulen Bad Leonfelden und Vorderweißbach und der Volksschule Vorderweißbach danken und auch Johanna Böck, die mir die Kontaktaufnahme erleichtert hat, soll nicht ungenannt bleiben.

Ganz speziell danke ich auch meinem Mann, Andreas Hofmayer, der den weitesten Weg des Studiums mit mir bewältigt hat, in schwierigen Zeiten ein offenes Ohr hatte und sich bei Erfolgen mit mir gefreut hat.

## **Abstract - deutsch**

Die vorliegende Arbeit ist als Weiterführung der Diplomarbeitsstudien von Böck (2010) und Eiter (2011) zu sehen, die es sich zum Ziel gesetzt haben für ausgewählte Subtests des AID 2 nach itemgenerierenden Regeln ein Itempool für eine Gruppen- beziehungsweise PC-Version des AID 2 zu entwickeln. Zu diesem Zweck wurden sowohl in den Vorgängerstudien, als auch in der aktuellen Studie, Kinder und Jugendliche zwischen 8;0 und 15;11 Jahren in vier oberösterreichischen Schulen mit dem neu entwickelten psychologisch-diagnostischen Verfahren getestet.

In der vorliegenden Studie wurden die Daten von Böck (2010) und Eiter (2011) nach Verdopplung der Stichprobe auf etwa 1400 Testpersonen und unter Verwendung eines verlinkten Designs einer Reanalyse hinsichtlich der Konformität mit dem dichotomen logistischen Modell nach *Rasch* unterzogen. Die Modellgültigkeit wurde mit dem Likelihood-Ratio-Test nach Andersen (1973) hinsichtlich der Teilungskriterien Score, Alter und Geschlecht überprüft.

Für den Subtest „Funktionen Abstrahieren“ konnte nach Ausschluss problematischer Items *Rasch*-Modell-Gültigkeit angenommen werden. Für die Untertests „Angewandtes Rechnen“ und „Wortschatz“ zeigten sich im LR-Test für das Teilungskriterium Alter signifikante Abweichungen vom Modell. Der letzte Untertest „Alltagswissen“ kann, wie bereits in der Vorgängerstudie, nicht als *Rasch*-Modell-konform bezeichnet werden. Hier ist eine Überprüfung und Überarbeitung der itemgenerierenden Regeln anzudenken. Generell legt für Folgeuntersuchungen die Überprüfung der itemgenerierenden Regeln an Hand des LLTM nahe.

Schlüsselwörter: AID 2, Gruppenverfahren, *Rasch*-Modell

## **Abstract – englisch**

This diploma thesis represents a continuation of the studies by Böck (2010) and Eiter (2011), who developed new items for selected subtests of the AID 2 using a rule-based approach. Their goal was to develop an item pool for a group or PC version of the AID 2. Their target group consisted of children and adolescents between 8;0 and 15;11 years of age who were tested in four different schools in Upper Austria.

While focusing on the same target group, this study doubled the sample to 1400 subjects and re-analyzed the data according to the *Rasch*-Model using a linked design. The validity of the *Rasch*-Model has been evaluated through the Likelihood-Ratio Test by Andersen (1973), using the split-criteria score, age and gender.

For the subtest "abstracting functionalities" ("Funktionen Abstrahieren"), *Rasch*-Model validity could be assumed after excluding problematic items. For the subtests "applied mathematics" (Angewandtes Rechnen") and "vocabulary" ("Wortschatz"), there were significant deviations from the model shown in the LR-test for the split-criterion "age". The last test, "everyday knowledge" ("Alltagswissen"), can, as in the previous study, not be called *Rasch*-Model compliant.

A review and revision of the rules for item development is necessary. For follow-up studies it is generally suggested to check the rules-generating guidelines using the LLTM (linear logistic test model).

Keywords: AID 2, group-testing, *Rasch*-model

# Inhaltsverzeichnis

<b>I. Einleitung</b> .....	<b>1</b>
1. Zielsetzung der Arbeit .....	3
<b>II. Theoretischer Teil</b> .....	<b>4</b>
2. AID 2 .....	4
3. Adaptives Testen .....	5
3.1 Branched-testing .....	6
3.2 Tailored-testing .....	6
4. Gruppenvorgabe .....	6
5. Computerverfahren.....	7
6. Das <i>Rasch</i> -Modell .....	8
<b>III. Empirischer Teil</b> .....	<b>11</b>
7. Beschreibung der Subtests.....	11
7.1 Funktionen Abstrahieren.....	11
7.2 Angewandtes Rechnen.....	12
7.3 Wortschatz .....	13
7.4 Alltagswissen .....	14
7.5 Überarbeitung der Instruktion .....	15
8. Testdesign.....	16
9. Durchführung der Studie.....	17
9.1 Stichprobenakquirierung.....	17
9.2 Testvorgabe .....	18
9.2.1 Beobachtungen bei der Datenerhebung .....	19
10. Deskriptive Statistiken.....	19
10.1 Beschreibung der aktuellen Daten .....	20
10.2 Beschreibung der Gesamtgruppe.....	20
10.2.1 Geschlecht .....	20
10.2.2 Alter .....	21
11. Analyse der <i>Rasch</i> -Modell-Konformität der Items.....	22
11.1 Funktionen Abstrahieren.....	24
11.1.2 Itemselektion.....	26
11.1.3 Itemanalyse .....	27
11.1.4 Modellschätzung nach Itemselektion.....	32
11.2 Angewandtes Rechnen.....	35

11.2.1	Itemselektion.....	38
11.2.2	Itemanalyse .....	38
11.2.3	Modellschätzung nach Itemselektion .....	41
11.3	Wortschatz .....	43
11.3.1	Itemselektion.....	46
11.3.2	Itemanalyse .....	46
11.3.3	Modellschätzung nach Itemselektion .....	52
11.4	Alltagswissen.....	55
11.4.1	Itemselektion.....	57
11.4.2	Itemanalyse .....	58
11.4.3	Zweite Modellschätzung .....	61
11.4.4	Analyse weiterer inhaltlich auffälliger Items.....	61
11.4.5	Dritte Modellschätzung.....	63
12.	Diskussion und Ausblick .....	66
13.	Zusammenfassung.....	69
14.	Literatur.....	72
<b>15.</b>	<b>Anhang</b> .....	<b>I</b>
15.1	Ansuchen an den Landesschulrat Oberösterreich .....	I
15.2	Elternbrief .....	VIII
15.3	Tabellen Funktionen Abstrahieren.....	VIII
15.4	Tabellen Angewandtes Rechnen .....	XIV
15.5	Tabellen Wortschatz .....	XIX
15.6	Tabellen Alltagswissen .....	XXIV
	Curriculum Vitae.....	XXIX

## Abbildungsverzeichnis

<b>Abbildung 1:</b> Darstellung des Testdesigns .....	16
<b>Abbildung 2:</b> Grafischer Modelltest für das Teilungskriterium Score vor Ausschluss eines Items, Subtest „Funktionen Abstrahieren“ .....	25
<b>Abbildung 3:</b> Grafischer Modelltest für das Teilungskriterium Geschlecht vor Ausschluss eines Items, Subtest „Funktionen Abstrahieren“ .....	25
<b>Abbildung 4:</b> Grafischer Modelltest für das Teilungskriterium Alter vor Ausschluss eines Items, Subtest „Funktionen Abstrahieren“ .....	26
<b>Abbildung 5:</b> Grafischer Modelltest für das interne Teilungskriterium nach Ausschluss der auffälligen Items, Subtest „Funktionen Abstrahieren“ .....	34
<b>Abbildung 6:</b> Grafischer Modelltest für das Teilungskriterium Geschlecht nach Ausschluss der auffälligen Items, Subtest „Funktionen Abstrahieren“ .....	34
<b>Abbildung 7:</b> Grafischer Modelltest für das Teilungskriterium Alter nach Ausschluss der auffälligen Items, Subtest „Funktionen Abstrahieren“ .....	35
<b>Abbildung 8:</b> Grafischer Modelltest für das interne Teilungskriterium vor Ausschluss eines Items, Subtest „Angewandtes Rechnen“ .....	36
<b>Abbildung 9:</b> Grafischer Modelltest für das Teilungskriterium Geschlecht vor Ausschluss eines Items, Subtest „Angewandtes Rechnen“ .....	37
<b>Abbildung 10:</b> Grafischer Modelltest für das Teilungskriterium Alter vor Ausschluss eines Items, Subtest „Angewandtes Rechnen“ .....	37
<b>Abbildung 11:</b> Grafischer Modelltest für das interne Teilungskriterium nach Itemselektion, Subtest „Angewandtes Rechnen“ .....	42
<b>Abbildung 12:</b> Grafischer Modelltest für das Teilungskriterium Geschlecht nach Itemselektion, Subtest „Angewandtes Rechnen“ .....	42
<b>Abbildung 13:</b> Grafischer Modelltest für das Teilungskriterium Alter nach Itemselektion, Subtest „Angewandtes Rechnen“ .....	43
<b>Abbildung 14:</b> Grafischer Modelltest für das interne Teilungskriterium vor Ausschluss eines Items, Subtest „Wortschatz“ .....	44
<b>Abbildung 15:</b> Grafischer Modelltest für das Teilungskriterium Geschlecht vor Ausschluss eines Items, Subtest „Wortschatz“ .....	45
<b>Abbildung 16:</b> Grafischer Modelltest für das Teilungskriterium Alter vor Ausschluss eines Items, Subtest „Wortschatz“ .....	45
<b>Abbildung 17:</b> Grafischer Modelltest für das Teilungskriterium Score nach Itemselektion, Subtest „Wortschatz“ .....	53

<b>Abbildung 18:</b> Grafischer Modelltest für das Teilungskriterium Geschlecht nach Itemselektion, Subtest „Wortschatz“ .....	54
<b>Abbildung 19:</b> Grafischer Modelltest für das Teilungskriterium Alter nach Itemselektion, Subtest „Wortschatz“ .....	54
<b>Abbildung 20:</b> Grafischer Modelltest für das Teilungskriterium Score vor Itemselektion, Subtest „Alltagswissen“ .....	56
<b>Abbildung 21:</b> Grafischer Modelltest für das Teilungskriterium Geschlecht vor Itemselektion, Subtest „Alltagswissen“ .....	56
<b>Abbildung 22:</b> Grafischer Modelltest für das Teilungskriterium Alter vor Itemselektion, Subtest „Alltagswissen“ .....	57
<b>Abbildung 23:</b> Grafischer Modelltest für das interne Teilungskriterium nach Itemselektion, Subtest „Alltagswissen“ .....	64
<b>Abbildung 24:</b> Grafischer Modelltest für das Teilungskriterium Geschlecht nach Itemselektion, Subtest „Alltagswissen“ .....	64
<b>Abbildung 25:</b> Grafischer Modelltest für das Teilungskriterium Alter nach Itemselektion, Subtest „Alltagswissen“ .....	65



## Tabellenverzeichnis

<b>Tabelle 1:</b> Geschlechterverteilung in der Stichprobe.....	20
<b>Tabelle 2:</b> Altersverteilung der Gesamtgruppe.....	21
<b>Tabelle 3:</b> Erste Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Funktionen Abstrahieren“ .....	24
<b>Tabelle 4:</b> Parameterschätzung des Likelihood Ratio Tests (LRT) nach Itemselektion, Subtest Funktionen Abstrahieren.....	33
<b>Tabelle 5:</b> Erste Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Angewandtes Rechnen“ .....	35
<b>Tabelle 6:</b> Parameterschätzung des Likelihood Ratio Tests (LRT) nach Itemselektion, Subtest „Angewandtes Rechnen“ .....	41
<b>Tabelle 7:</b> Erste Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Wortschatz“ .....	43
<b>Tabelle 8:</b> Parameterschätzung des Likelihood Ratio Tests (LRT) nach Itemselektion, Subtest „Wortschatz“ .....	53
<b>Tabelle 9:</b> Erste Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Alltagswissen“...	55
<b>Tabelle 10:</b> Zweite Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Alltagswissen“ .....	61
<b>Tabelle 11:</b> Dritte Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Alltagswissen“	63

## Im Anhang

<b>Tab. a:</b> Erste Parameterschätzung, z-Test, Subtest „Funktionen Abstrahieren“, ( $\alpha=0,01$ ).....	VIII
<b>Tab. b:</b> Itemfit-Statistik, vor Itemselektion, Subtests „Funktionen Abstrahieren“; ( $\alpha=0,01$ ) .....	XI
<b>Tab. c:</b> Itemparameter (Leichtigkeitsparameter), vor Itemselektion, Subtest „Funktionen abstrahieren“.....	XII
<b>Tab. d:</b> Erste Parameterschätzung, z-Test, Subtest „Angewandtes Rechnen“, ( $\alpha=0,01$ ).....	XIV
<b>Tab. e:</b> Itemfit-Statistik, vor Itemselektion, Subtest „Angewandtes Rechnen“, ( $\alpha=0,01$ ) .....	XVI
<b>Tab. f:</b> Itemparameter (Leichtigkeitsparameter), vor Itemselektion, Subtest „Angewandtes Rechnen“ .....	XVII
<b>Tab. g:</b> Erste Parameterschätzung, z-Test, Subtest „Wortschatz“, ( $\alpha=0,01$ ) .....	XIX
<b>Tab. h:</b> Itemfit-Statistik, vor Itemselektion, Subtest „Wortschatz“, ( $\alpha=0,01$ ).....	XXI
<b>Tab. i:</b> Itemparameter (Leichtigkeitsparameter), vor Itemselektion, Subtest „Wortschatz“ .....	XXII
<b>Tab. j:</b> Erste Parameterschätzung, z-Test, Subtest „Alltagswissen“, ( $\alpha=0,01$ ) .....	XXIV
<b>Tab. k:</b> Itemfit-Statistik, vor Itemselektion, Subtests „Alltagswissen“; ( $\alpha=0,01$ ) .....	XXVI
<b>Tab. l:</b> Itemparameter (Leichtigkeitsparameter), vor Itemselektion, Subtest „Alltagswissen“ .....	XXVII



## I. Einleitung

Die Psychologische Diagnostik steht im Spannungsfeld neuer Herausforderungen und neuer Anforderungen, sowie dem wachsenden Anspruch Individuen als solche zu betrachten. Der diagnostische Zugang geht längst über das normorientierte Vorgehen hinaus. Dennoch befindet sich Diagnostik immer an der Grenze zwischen Norm (im Sinne von Normierung) und normal (im Sinne der allgemeinsprachlichen Bedeutung, jemand sei „normal“) (vergleiche Kubinger, 2006). Betrachtet man den Ursprung des Begriffes Diagnostik, als „durch und durch erkennen, beurteilen“ (Drosdowski 1989, S. 124), abgeleitet vom griechischen Verb *diagignoskein*, wird deutlich, dass die Diagnostik ein umfassender Prozess sein sollte, an dessen Ende (neue) Erkenntnisse über eine Person stehen.

Folgt man dieser Prämisse, scheint es selbstverständlich, dass Diagnostik immer auf den Einzelfall ausgelegt sein sollte. Die Testperson sollte im Mittelpunkt des diagnostischen Prozesses und der Aufmerksamkeit des Testleiters stehen. Ein offenes Antwortformat, Interaktion mit dem Testleiter, ausreichend Zeit und angenehme Rahmenbedingungen sollten dafür Sorge tragen, dass die Testperson ihr Potential ohne beschränkende oder verunsichernde Faktoren zeigen kann. In unmittelbarem Kontrast dazu steht allerdings die notwendige ökonomische Ausrichtung von psychologisch-diagnostischer Verfahren und der Diagnostik im Allgemeinen. Immer mehr Augenmerk wird auf Berufs- und Bildungsdiagnostik, Personalentwicklung und vor allem Selektionsdiagnostik gelegt. Diagnostik soll effizient, ökonomisch und eindeutig sein; eine verständliche Forderung seitens der „Kunden“.

Innerhalb dieser Arbeit geht es vor allem um Diagnostik im Kinder- und Jugendbereich. Eines der populärsten Verfahren zur Intelligenzdiagnostik bei Kindern und Jugendlichen, das Adaptive Intelligenzdiagnostikum 2 (Kubinger, 2009), soll als Gruppenversion überarbeitet werden. Interessanterweise ist gerade das AID 2 eines der Verfahren, das man als klassisches Einzelverfahren bezeichnen könnte. (Nähere Erläuterungen zum AID 2 finden sich in Kapitel 2). Die adaptive Vorgabe sorgt für ausreichend Motivation bei der Testperson und vermeidet Über- beziehungsweise Unterforderung. Die meisten Untertests versuchen ohne eine Speed-Komponente auszukommen und das Antwortformat ist überwiegend frei. Es wird von der Möglichkeit Gebrauch gemacht „Intelligenz in einem mit einer anderen (d.i. der/die Testleiter/-in) interaktiven Prozess zu erfassen“ (Kubinger, 2009, S. 6). Die Vorgabe der Testaufgaben erfolgt

über die verschiedensten Sinneskanäle – bildhaft, schriftlich, verbal, taktil und akustisch. Die Vorteile einer solchen Einzeltestung liegen auf der Hand.

Nun stellt sich die Frage, warum brauchen wir eine Gruppenversion des AID 2? Eines der wesentlichsten Merkmale des AID 2 ist die Möglichkeit einer Profilinterpretation. Das heißt, dass die Leistungen der Testpersonen differenziert betrachtet werden können und sowohl persönliche Stärken, als auch Schwächen daraus hervor gehen. „Wohl ist die Anwendung von Intelligenztests mit dem Ziel einer Selektion heutzutage geradezu verpönt, doch wird die Testung letztlich oft darauf hinauslaufen; mit dem AID soll aber wenigstens versucht werden, mehr denn je aus den Testergebnissen Hinweise auf entsprechende Fördermöglichkeiten abzuleiten, indem besonderer Wert auf die Interpretation der einzelnen Untertests zueinander und die damit verbundene differentialdiagnostische Bedeutung gelegt wird“ (Kubinger & Wurst, 1991, zitiert nach Kubinger, 2009, S. 6).

In einer Zeit, in der die Generalisierung und Normierung von Leistungen immer populärer wird, man denke beispielsweise an globale Leistungsüberprüfungen wie die Pisa-Studie, oder die ab dem Schuljahr 2013/2014 umgesetzte Zentralmatura, scheint diese Ausdifferenzierung eines individuellen Leistungsprofils gerade innerhalb einer Gruppentestung als erstrebenswert. In einem defizitorientierten Schulsystem ist die Aufschlüsselung persönlicher Fähigkeiten für Schüler und Schülerinnen möglicherweise ein motivierender Faktor gerade diese Stärken zu fördern. Auch für die Lehrerschaft mag es interessant sein, das Leistungsspektrum der Schüler und Schülerinnen derart differenziert zu begutachten und über eine Art „Leistungsbarometer“ im Sinne eines Screenings zu verfügen.

Zusätzlich sei der Vorteil genannt, dass die Verfügbarkeit einer Gruppentestung auch für die wissenschaftliche Praxis dienlich sein könnte. So könnten rasch große Datenmengen erhoben werden.

Als möglicher Ausblick im Rahmen dieses Prozesses sei auch die Verwendung technologischer Hilfsmittel genannt. Kinder und Jugendliche werden immer technik-affiner und die Verwendung von Computern wird zur Selbstverständlichkeit und auch in Schulen befindet sich mittlerweile zumeist eine gute technische Infrastruktur. Vor allem was die adaptive Testung betrifft, eröffnen sich durch den Computereinsatz zusätzliche Möglichkeiten und Vereinfachungen. Auch was die Itemkonstruktion und die Vorgabe betrifft ist durch den Einsatz eines Computers ein neues, ansprechendes Feld eröffnet.

## 1. Zielsetzung der Arbeit

Die vorliegende Arbeit versteht sich als Ergänzung und Erweiterung der Arbeiten von Böck (2010) und Eiter (2011). Innerhalb dieser beiden Diplomarbeiten wurden für vier Subtests des AID 2, nämlich „Funktionen Abstrahieren“, „Synonyme finden“, „Angewandtes Rechnen“ und „Alltagswissen“, nach itemgenerierenden Regeln neue Items entwickelt. Diese Items sind in ihrer Formulierung und Form daraus ausgerichtet, dass sie innerhalb einer Gruppen- beziehungsweise Computertestung zum Einsatz kommen.

In einer Datenerhebung wurden diese Testaufgaben rund 680 Schülern und Schülerinnen in oberösterreichischen Schulen zur Bearbeitung in einer Papier-Bleistift-Version vorgegeben. Jeder der Untertests wurde in vier etwa gleich schwierige Testformen aufgeteilt. Auf Grund der notwendigen Lesefertigkeiten der Testpersonen, wurde die Erhebung nur mit Schülern und Schülerinnen durchgeführt, die bereits zumindest die 3. Schulstufe besuchten. Der Altersbereich der Testpersonen innerhalb der vorliegenden Erhebungen erstreckt sich also auf 8;0 bis 15;11 Jahre und nicht, wie beim AID 2, auf 6;0 bis 15;11 Jahre.

Anschließend wurden die Items der jeweiligen Testformen an Hand von drei Teilungskriterien (Score, Geschlecht, Region) mittels Likelihood-Quotienten-Test nach Andersen (1973) bezüglich der Gültigkeit des dichotomen logistischen Modells nach *Rasch* überprüft.

Ausgeblieben ist in diesen vorhergehenden Studien allerdings die Überprüfung der Gültigkeit des *Rasch*-Modells in Hinblick auf den jeweiligen kompletten Untertest. Jede der vier Testformen für sich wurde zwar auf *Rasch*-Modell-Konformität getestet, das gesamte Itempool des Subtests aber eben nicht. Für die weitere Testkonstruktion des AID als PC-Version, konkret um die adaptive Vorgabe zu ermöglichen, ist es allerdings notwendig, dass sämtliche Testaufgaben eines Subtests *Rasch*-Modell konform sind.

Die Zielsetzung dieser Diplomarbeit beläuft sich also darauf, eine Reanalyse der Daten von Böck (2010) und Eiter (2011) durchzuführen. Zu diesem Zweck wurde, mit leicht überarbeitetem Testmaterial, in denselben oberösterreichischen Schulen eine weitere Datenerhebung durchgeführt, um die Stichprobengröße entsprechend aufzustocken. Um die Analyse auf *Rasch*-Modell-Konformität innerhalb der Untertests durchführen zu können, wurde als Testdesign ein verlinktes Design gewählt.

Eingebettet in den Gesamtprozess der Konstruktion des AID als Gruppenversion, stellt die vorliegende Arbeit einen kleinen, wenn auch nicht zu vernachlässigenden, Teil dar.

## II. Theoretischer Teil

### 2. AID 2

Im Rahmen dieser Arbeit soll nur überblicksartig auf das psychologisch-diagnostische Verfahren AID 2 eingegangen werden. Der Einfachheit halber ist hier die Rede vom AID 2. Die Differenzierung in AID 2 und der überarbeiteten Version AID 2.2 (Kubinger 2009) wird nicht näher beleuchtet. Der darüber hinaus gehend interessierte Leser wird gebeten bei Kubinger (2009) nachzulesen.

Das AID 2 (Adaptives Intelligenz Diagnostikum 2) von Kubinger und Wurst (2001) – und in überarbeiteter Fassung als AID 2.2 (Kubinger, 2009) – ist eine Intelligenztestbatterie für die Einzeltestung von Kindern und Jugendlichen im Altersbereich 6;0 bis 15;11 Jahren und steht in der Intelligenztheoretischen Tradition von David Wechsler. Die Vorgabe im Einzelsetting dauert zwischen 30 und 75 Minuten, wobei für die optionalen Zusatztests zwischen 10 und 15 zusätzliche Minuten zu kalkulieren ist.

Das AID 2 grenzt sich von anderen Intelligenztests, die zumeist in der Gruppe vorgegeben werden, oder zumindest eine Gruppenvorgabe ermöglichen, deutlich ab; und zwar insofern, als dass größtenteils reine Power-Tests zum Einsatz kommen und das offene Antwortformat vorherrscht (Kubinger, 2009). Zusätzlich zeichnet sich das AID 2 vor allem dadurch aus, dass eine Profilinterpretation im Sinne der förderungsorientierten Diagnostik möglich ist. Der weit verbreitete gemittelte Wert des Intelligenzquotientens wird vom Testautor abgelehnt, da „ein solcher Durchschnittswert aus forderungsorientierter Sicht gar nicht zielführend wäre.“ (Kubinger, 2009, S. 7)

Das psychologisch-diagnostische Verfahren setzt sich aus elf Untertests und drei optionalen Zusatztests zusammen, die zur Erfassung komplexer sowie basaler Kognitionen dienen. Die Subtests lassen sich in zwei Gruppen differenzieren. Demnach erfassen sechs der Untertests verbal-akustische Fähigkeiten und fünf manuell-visuelle Fähigkeiten. Zusätzlich kann eine Unterscheidung in förderungsabhängige und förderungsunabhängige Subtests getroffen werden. Die Testvorgabe bedingt eine enge Interaktion zwischen Testleiter beziehungsweise Testleiterin und den Testpersonen, als da auch vielfältige Testmaterialien, wie Bildkarten, Vorlegeteile, Würfel etc. zum Einsatz kommen.

Der überwiegende Teil der Subtests kann einem von vier Faktoren zugeordnet werden: „Informationsverarbeitung in der gesellschaftlichen Umwelt“, „Informationsverarbeitung neuer Inhalte“, „Auffassungskapazität“ oder „(Re-) Produktionsfähigkeit durch Strukturierung“. Das

AID 2 zeichnet sich besonders dadurch aus, dass acht der Subtests nach dem Prinzip des *branched testing* vorgegeben werden, wodurch ermöglicht wird, „dass jeder Testperson grundsätzlich bloß diejenigen Aufgaben gestellt werden, welche ihrem Leistungsniveau entsprechen.“ (Kubinger, 2009, S. 36) Weitere Informationen zum adaptiven Testen finden sich im nachfolgenden Kapitel. Für die übrigen Subtests erfolgt die Vorgabe konventionell.

### **3. Adaptives Testen**

Die adaptive Vorgabe grenzt das AID 2 deutlich von anderen Intelligenztests in Einzelvorgabe für Kinder und Jugendliche ab. Selbstverständlich wäre es wünschenswert, wenn auch die Vorgabe in der Gruppe adaptiv umsetzbar wäre. Wenn allerdings auch in der Gruppenversion des AID 2 eine adaptive Vorgabe möglich sein soll, erscheint die Verwendung eines Computers beinahe obligat.

Nun soll an dieser Stelle kurz auf das adaptive Testen eingegangen werden. Darüber hinausgehend interessierte Leser seien beispielsweise auf Kubinger (2003 oder 2006), Kubinger und Wild (1989) und Wild (1989) verwiesen.

Die adaptive Vorgabe psychologisch-diagnostischer Verfahren zielt immer darauf ab den Informationsgewinn, bei gleichzeitig geringer Testlänge, zu maximieren (Kubinger, 2006).

Im Gegensatz zu konventioneller Vorgabe, bei der jede Testperson zumeist idente Items in der identen Reihenfolge vorgegeben werden, ergibt sich beim adaptiven Testen die jeweils nächste Testaufgabe aus der vorhergehenden. Je nachdem, ob das vorhergehende Item gelöst oder nicht gelöst wurde, wird die nächste Testaufgabe in der Schwierigkeit angepasst. So lässt sich der Informationsgewinn durch die ausschließliche Vorgabe informativer Items, also jene, die der Fähigkeit der Testperson entsprechen, erheblich erhöhen. Für die Testperson birgt diese Vorgabemethode den Vorteil, dass sie sich während der Testung weder über- noch unterfordert fühlt und die Testdauer sich deutlich verkürzt. Nachdem aber jede Testperson eine andere Kombination von Items erhält, kann der Score, beziehungsweise Rohwert nicht in herkömmlicher Art und Weise herangezogen werden, da dieser zwischen den Testpersonen nicht vergleichbar ist und in Abhängigkeit der vorgegebenen Items zu betrachten ist.

Innerhalb der probabilistischen Testtheorie stehen allerdings Modelle zur Verfügung, die es ermöglichen „die wahre, aber unbekannte Fähigkeit einer beliebigen Person (...) unter Berücksichtigung der getroffenen Itemauswahl zu schätzen.“ (Kubinger, 2003). Außerdem ermöglicht dieses Vorgehen die Ermittlung der sogenannten „Information“ eines Items in Form der Itemparameter.

Für die adaptive Vorgabe stehen zwei Möglichkeiten zur Verfügung: Branched-testing und Tailored-testing. Beide Vorgabemöglichkeiten werden im Folgenden kurz umrissen.

### **3.1 Branched-testing**

Der hier im Mittelpunkt der Arbeit stehende AID 2 wird nach dem Prinzip des Branched-testing vorgegeben. Bei dieser Vorgabemethode werden die Items in Gruppen zusammengefasst. Je nachdem, wie viele Items einer Itemgruppe eine Testperson löst, wird nachfolgend eine Itemgruppe mit leichteren, schwierigeren oder nahezu gleichschwierigen Items für die Vorgabe gewählt. Diese Verzweigung in die jeweils nächste Gruppe erfolgt durch im Vorhinein festgelegte Regeln. Auch wenn das Branched-testing den Spielraum bezüglich der Testökonomie nicht optimal ausreizt, besteht der klare Vorteil, dass der administrative Aufwand durch die festgelegte Verzweigung derart eingeschränkt ist, dass die Vorgabe auch durch einen Testleiter in einer Papier-Bleistift Version möglich ist (Kubinger & Wild, 1989; Kubinger, 2003)

### **3.2 Tailored-testing**

Die vorliegende Arbeit zielt darauf ab, dass künftig eine Version des AID 2 nach dem Prinzip des tailored testing zur Umsetzung gelangt. Je nachdem ob eine Testperson das erste, mittelschwere Item gelöst oder nicht gelöst hat, wird nachfolgend das schwierigste oder leichteste Item vorgegeben. Sobald die Testperson zumindest ein Item gelöst und eines nicht gelöst hat, werden Items vorgegeben, deren Schwierigkeit das Leistungsniveau der Testperson in etwa abbildet. Das heißt, dass der Personenparameter immer mehr verfeinert wird. Die Testung wird dann beendet, wenn ein im Vorhinein festgelegtes Abbruchkriterium erreicht wurde.

Diese Vorgehensweise setzt voraus, dass ein möglichst großer Pool an Items bereits an Hand einer Kalibrierungsstichprobe analysiert wurde und die Item- beziehungsweise Schwierigkeitsparameter dieser Testaufgaben bekannt sind (Kubinger, 2003 und 2006).

Im Wesentlichen lässt sich die Zielsetzung dieser Arbeit dahingehend konkretisieren, dass für vier Subtests des überarbeiteten AID 2 eine derartige Kalibrierung durchgeführt wurde.

## **4. Gruppenvorgabe**

Kubinger (2006) führt bezüglich der Vorteile der Gruppenverfahren vor allem die wirtschaftliche Effizienz, beziehungsweise Ökonomie der Verfahren, sowie die weitgehende Testleiterunabhängigkeit an. Amelang (2002) merkte dazu aber auch einige Nachteile an, wie beispielsweise das Ausfallen, beziehungsweise Einschränken, der Möglichkeit einer Verhaltensbeobachtung, sowie die etwaige Benachteiligung von Personen mit geringeren Lese- oder Rechtsschreibkenntnissen. Auch die Gefahr des „Abschreibens“ besteht bei der Vorgabe in



der Gruppe immer, wobei man sich durch Parallelförmigkeiten helfen kann. Auch die Gruppensituation als solche ist als zusätzlicher Stressor nicht zu vernachlässigen.

Für Kinder und Jugendliche ist davon auszugehen, dass sie das (konzentrierte) Arbeiten im Gruppensetting aus ihrem Schulalltag gewohnt sind. Dennoch ist zu bemerken, dass im Gruppensetting gewisse ansprechende Materialien, die im Einzelsetting ihren Einsatz finden (beispielsweise Puzzles, Mosaikwürfel etc.) im Gruppensetting nicht verwendet werden können. Zudem sind die Beziehung zum Testleiter beziehungsweise Testleiterin und die damit einhergehende Motivation für Kinder und Jugendliche der Leistung gewissermaßen zuträglich. Ob also die Vorgabe im Einzel- oder im Gruppensetting vorzuziehen ist, kann mit Sicherheit nicht klar beantwortet werden, sondern wird für den Einzelfall zu entscheiden sein.

## **5. Computerverfahren**

Dass der Einsatz von technologischen Geräten im Wachsen begriffen ist, muss an dieser Stelle nicht diskutiert werden. Dennoch ist es interessant einige Zahlen und Fakten zu nennen. Laut Statistik Austria haben im Jahr 2011 78,1 Prozent der österreichischen Haushalte zumindest einen Computer und 75,4 Prozent verfügen über Internetzugang.

Auch in den schulischen Alltag und den Freizeitbereich von Jugendlichen finden Computer, Handys, Smartphones, Spielkonsolen etc. immer mehr ihren festen Platz. So verfolgte eine 2007 durchgeführte Studie von Lampert et al. das Ziel herauszufinden, wie verbreitet die Nutzung elektronischer Medien bei 11- bis 17-jährigen Jugendlichen ist. Die Ergebnisse demonstrieren die Rolle technischer Gerätschaften eindrucksvoll. 76 Prozent der Zielgruppe nutzen täglich Computer und Internet durchschnittlich 3,8 Stunden. Immerhin 34 Prozent der Burschen und 18 Prozent der Mädchen verbringen täglich rund 5 Stunden mit technischen Geräten.

Es erscheint daher nur allzu schlüssig, dass man auch in der Erstellung psychologisch-diagnostischer Verfahren bestrebt ist auch eine PC-Vorgabe zu erstellen.

Die Vorteile der Nutzung des Computers in der Psychologischen Diagnostik werden an anderer Stelle erschöpfend beschrieben, beispielsweise bei Kubinger (2009). Dennoch soll an dieser Stelle vor allem auf die Möglichkeit der ansprechenden Gestaltung des Testmaterials für Kinder und Jugendliche hingewiesen werden. Auch bezüglich der adaptiven Vorgabe im Sinne des „tailored testing“ bietet die PC Nutzung ganz klar Vorteile gegenüber Paper-Pencil Verfahren im Einzelsetting.

## 6. Das *Rasch*-Modell

Das dichotome logistische Testmodell nach Georg Rasch, das zur einfacheren Lesbarkeit meist nur *Rasch*-Modell genannt wird, kann für die Psychologische Diagnostik als eines der wichtigsten Modelle der Item-Response-Theory gesehen werden und zeichnet sich im Gegensatz zu anderen Modellen der IRT durch seine empirische Prüfbarkeit durch statistische Modelltests aus (Kubinger & Draxler, 2007; Kubinger et al., 2010)

Die Einsatzmöglichkeiten des *Rasch*-Modells gehen also über die schlichte Prüfung, ob ein Modell gewisse Daten beschreibt, hinaus. Das *Rasch*-Modell stellt die Forderung nach einer Generalisierbarkeit von Vergleichen und Unterschieden; sowohl zwischen Items, als auch zwischen Personen. „Nach *Georg Rasch* geht es also um die „spezifische Objektivität von Vergleichen“ und um Prüfmöglichkeiten, diese für bestimmte Messprozesse, konkret psychologische Tests, empirisch als gegeben oder nicht gegeben abzutesten.“ (Kubinger & Draxler, 2007, S. 134) Die Gültigkeit dieser spezifisch objektiven Vergleiche stellt eine wichtige Voraussetzung für die adaptive Vorgabe von psychologisch-diagnostischen Verfahren dar. Im Wesentlichen meint man mit spezifisch objektiven Vergleichen, dass die Testergebnisse zweier Testpersonen auch dann miteinander verglichen werden können, wenn unterschiedliche Items aus dem Itempool vorgegeben wurden (Kubinger, 1989). Abgesehen davon kann dem *Rasch*-Modell auch Stichprobenunabhängigkeit bescheinigt werden, da sich eben nicht nur die Testleistungen von Testpersonen unabhängig von der vorgegebenen Items vergleichen lassen, sondern auch die Schwierigkeiten von Items unabhängig davon, welche Stichprobe aus einer Population gezogen wurde (Kubinger, 1989, 2003).

Einer der wichtigsten Einsatzbereiche des *Rasch*-Modells findet sich vor allem dann für die Überprüfung psychologisch-diagnostischer Verfahren, bei denen als Testkennwert die Anzahl gelöster Aufgaben bestimmt wird. Hier muss das *Rasch*-Modell Gültigkeit haben um einen fairen Vergleich zwischen den Testpersonen zu garantieren. Nur so kann die gesamte relevante Information hinsichtlich der Fähigkeit der Testperson in das Ergebnis einbezogen werden, da die Gültigkeit des *Rasch*-Modells sicherstellt, dass die sogenannte Verrechnungsfairness gegeben ist. Man kann also mit Sicherheit sagen, dass eine Person mit einem höheren Testwert auch tatsächlich leistungsfähiger ist, als eine Person mit niedrigem Testwert (Kubinger, 2006).

Wie aus Formel 6.1 hervorgeht, definiert das *Rasch*-Modell die Wahrscheinlichkeit, „dass  $T_p$  v Item  $i$  löst („+“), in Abhängigkeit des Personenparameters  $\xi_p$ , das ist die wahre Fähigkeit von  $p$ , und des Itemparameters  $\sigma_i$ , das ist die (wahre) Schwierigkeit von  $i$ “ (Kubinger, 2006, S. 85).

$$P(+|\xi_v, \sigma_i) = \frac{e^{\xi_v - \sigma_i}}{1 + e^{\xi_v - \sigma_i}}$$

(6.1)

Die beiden Parameter,  $\xi$  und  $\sigma$ , werden im Rahmen des *Rasch*-Modells geschätzt, wodurch dann ein Rückschluss auf das Antwortverhalten der Testperson möglich wird. Genauer gesagt bedeutet das, dass man berechnen kann mit welcher Wahrscheinlichkeit eine Testperson ein Item löst, wenn man die Schwierigkeit und die Fähigkeit der Testperson kennt. Dieser Zusammenhang ist für die adaptive Vorgabe von Items wesentlich.

Das *Rasch*-Modell stützt sich im Wesentlichen auf drei Grundannahmen. Die Annahme der lokalen stochastischen Unabhängigkeit, stellt die erste dieser Voraussetzungen dar. Unter der lokalen stochastischen Unabhängigkeit versteht man laut Kubinger (2006, S. 85): „Ob eine Tp ein Item löst oder nicht löst, hängt, abgesehen vom Zufall nur von ihrer Fähigkeit und von der Schwierigkeit des Items ab, nicht aber davon, welche anderen Items sie schon gelöst hat oder noch lösen wird.“

Eine weitere wichtige Voraussetzung stellt die Grundannahme der erschöpfenden Statistik dar. Eine erschöpfende Statistik meint, dass die Anzahl der gelösten Items für eine Testperson, also der Rohwert, eine erschöpfende Statistik für deren Personenparameter darstellt. Also umfasst dieser Rohwert alle notwendigen Informationen über die Fähigkeit einer Person, allerdings unabhängig davon, welche Testaufgaben diese Person tatsächlich gelöst hat (Kubinger, 1989). Die dritte relevante Grundannahme besteht darin, dass alle Testaufgaben eine, beziehungsweise dieselbe, latente Eigenschaft oder Fähigkeit messen (Eindimensionalität) (Kubinger, 2003).

Die Gültigkeit des *Rasch*-Modells in Bezug auf einen gewissen Datensatz, belegt gleichzeitig die Gültigkeit dieser Grundannahmen.

Die Überprüfung des *Rasch*-Modells erfolgt, wie eingangs erwähnt, durch die Anwendung statistischer Modelltests. An dieser Stelle wird auf den Likelihood-Ratio-Test nach Andersen (1973) eingegangen, da dieses Verfahren auch in der vorliegenden Arbeit seine Anwendung findet. Der Likelihood-Ratio-Test nach Andersen macht sich die Stichprobenunabhängigkeit des *Rasch*-Modells zu Nutze und analysiert im Wesentlichen die Itemparameterdifferenzen an Hand von Stichprobensubgruppen. Diese Stichprobensubgruppen werden hinsichtlich des Scores, also die gelösten Aufgaben pro Testperson, oder anderen externen Kriterien, wie beispielsweise Geschlecht oder Alter, gebildet.

Wenn das *Rasch*-Modell gilt, sind, bis auf zufällige Schwankungen, keine Unterschiede der Itemparameterschätzungen pro Item zwischen den definierten Subgruppen festzustellen. (Glas & Verhelst, 1995)

Der Grafische Modelltest bietet dafür auch eine anschauliche grafische Darstellung, indem die Parameterschätzungen in einem rechtwinkligen Koordinatensystem an Hand einer 45° Geraden pro Subgruppe abgetragen werden. Je näher die Itemparameterschätzungen an ebendieser Geraden liegen, desto eher kann von der Gültigkeit des *Rasch*-Modells ausgegangen werden (Kubinger, 2006).

### **III. Empirischer Teil**

#### **7. Beschreibung der Subtests**

Wie schon erwähnt, bezieht sich die vorliegende Studie auf jene vier Subtest, die bereits in den Vorgängerstudien von Böck und Eiter herangezogen wurden: „Funktionen Abstrahieren“, „Angewandtes Rechnen“, „Wortschatztest“ und „Alltagswissen“.

Eine erschöpfende Beschreibung dieser Subtests und zu deren Entstehung findet sich bei Böck (2010) und Eiter (2011). An dieser Stelle wird auf eine ausführliche Erläuterung verzichtet, die vorgegebenen Aufgaben und Items werden lediglich umrissen.

##### **7.1 Funktionen Abstrahieren**

Der Subtest „Funktionen Abstrahieren“ wurde an erster Stelle im Testheft vorgegeben. Bereits in der ursprünglichen Version des AID 2 ist ein Subtest mit demselben Titel und weitgehend ähnlichen Inhalten enthalten. „Mit dem Untertest 9 Funktionen Abstrahieren soll die Fähigkeit erfasst werden, durch Abstraktion zu einer Begriffsbildung zu gelangen.“ (Kubinger, 2009, S. 11). Während im AID 2 eine gemeinsame Funktion zweier Objekte, die verbal vorgegeben werden, genannt werden soll, sollten die Testpersonen in dieser Version des AID zu einem Objekt zwei weitere Objekte aus fünf Distraktoren finden, die eine gemeinsame Funktion umschreiben. Zur Verdeutlichung findet sich nachfolgend ein Beispielitem:

##### **Einfamilienhaus**

Reihenhaus    Fuchsbau    Vogelnest    Mietshaus    Hundehütte

(Item übernommen aus Böck, 2010)

Den Schülern und Schülerinnen wurden 15 Testaufgaben dieser Art zur selbstständigen Bearbeitung vorgegeben.

Für die Verrechnung galt, dass ein Item nur dann als richtig gezählt wurde, wenn beide Objekte oder Begriffe, die mit dem dritten eine gemeinsame Funktion erfüllen, angekreuzt wurden.

Die Testaufgaben wurden vollständig aus der Vorgängerstudie (Böck, 2010) übernommen und für die Vorgabe lediglich in eine neue Reihenfolge gebracht um auch in den neu zusammengestellten Parallelversionen annähernd ansteigende Schwierigkeit zu erhalten.

Vor der eigentlichen Studie wurde eine kleine Vorerhebung mit einer 9-jährigen und einer 12-jährigen Testperson gemacht. Abgeleitet aus diesen Daten und den Ergebnissen der Vorgängerstudie (Böck, 2010) ergab sich, dass für diesen Subtest maximal 9 Minuten zur

Bearbeitung ausreichen. Spätestens nach Ablauf dieser Zeitspanne, wurden die Testpersonen von dem Testleiter oder der Testleiterin gebeten zum nächsten Subtest überzugehen.

## 7.2 Angewandtes Rechnen

Auch im AID 2 selbst gibt es einen Subtest, der „Angewandtes Rechnen“ heißt. Im Rahmen dieses Subtests werden der Testperson mathematische Testaufgaben vorgelesen; dabei besteht ab einem gewissen Alter der Testpersonen die Möglichkeit zum Mitlesen in einem Textheft. Diese Aufgaben sollen dann innerhalb einer Zeitbegrenzung gelöst werden (Kubinger, 2009). „Der Untertest 3 Angewandtes Rechnen soll weitgehend unabhängig von schulischen Rechenfertigkeiten prüfen, inwieweit die Testperson bei der Problemlösung alltäglicher Aufgabenstellungen durch entsprechende Schlussfolgerungen die passenden Rechenoperationen anzuwenden imstande ist.“ (Kubinger, 2009, S. 9.) Eine Beispielaufgabe der Gruppenversion wiederum zur Veranschaulichung:

Anne hat sechs Eier in einer Schachtel gekauft. Allerdings stolpert sie und fällt hin. Dabei gehen zwei Eier kaputt. Wie viele ganze Eier hat Anne jetzt noch in der Schachtel?



(Item übernommen aus Eiter, 2011)

Das Testheft umfasste jeweils 15 dieser Textaufgaben, alle in freiem Antwortformat, die die Schüler und Schülerinnen ohne Verwendung von Hilfsmitteln (Taschenrechner, Handy etc.) lösen sollten. Nebenrechnungen konnten direkt auf dem Testbogen gemacht werden; falls jemand zu wenig Platz hatte, wurden weitere Blätter ausgegeben.

Die Verrechnung erfolgte auch hier nach dem richtig-oder-falsch-Prinzip. Das heißt, dass eine Antwort nur dann als richtig gewertet wurde, wenn das Ergebnis korrekt war. Richtig gerechnete Zwischenschritte wurden nicht gezählt.

Für diesen Subtest wurden die Testaufgaben aus der Vorgängerstudie (Eiter, 2011) ident übernommen. Wie bereits beim vorhergehenden Subtest wurde lediglich die Reihenfolge verändert. Im Testheft waren diese Aufgaben an zweiter Stelle.

Aus der Vorerhebung mit zwei Testpersonen (9 und 12 Jahre alt) und in Anbetracht der Testdauer aus der Vorgängerstudie (Eiter, 2011), wurden die Testpersonen nach maximal 12 Minuten gebeten zum nächsten Subtest überzugehen.

### 7.3 Wortschatz

Die Konstruktion dieses Subtests erfolgte in Anlehnung des Subtests „Synonyme finden“ aus dem AID 2 (Böck, 2010). Der Untertest „Synonyme finden“ verfolgt die Messung des elementaren Sprachverständnisses. Der Testperson wird ein Begriff verbal vorgegeben und sie wird dazu angehalten einen alternativen Begriff mit derselben Bedeutung zu nennen. (Kubinger, 2009) „Der Untertest 6 Synonyme Finden soll das elementare Sprachverständnis prüfen, nämlich in wie weit die Testperson die Bedeutung sprachgebundener Begriffe erfasst bzw. über einen Wortschatz verfügt, der solche Begriffe alternativ ausdrücken lässt.“ (Kubinger, 2009, S. 10)

Nachdem die verbale Vorgabe in der Gruppenversion nicht zu verwirklichen war, wurden in der Vorgängerstudie (Böck, 2010) neue Items generiert. Das Hauptaugenmerk lag nun nicht mehr auf dem Finden von Synonymen, sondern viel mehr beim Definieren von Begriffen. Die Items bestanden nun jeweils aus einem Begriff, der durch einen nachstehenden Satz definiert wurde. An jeweils zwei Stellen im Satz sollten die Testpersonen aus je vier Distraktoren die richtigen auswählen um den Satz zu vervollständigen; erneut wird ein Beispielitem angeführt:

**Hören** bedeutet sich eines

- Sinnes
- Zeichens
- Geräusches
- Geruchs

mittels

- der Augen
- der Ohren
- der Nase
- des Verstandes

bewusst zu werden.

(Item übernommen aus Böck, 2010)

Zur Bearbeitung folgten 20 dieser Testaufgaben. Im Testheft war der Untertest „Wortschatz“ an dritter Stelle zu finden.

Die Items dieses Subtests wurden für die Verwendung in der vorliegenden Studie dahingehend überarbeitet, als dass versucht wurde die Satzstruktur so einfach wie möglich zu gestalten. So wurde vermieden, dass Verwirrung durch die Verwendung der unterschiedlichen Präpositionen oder Artikel entsteht.

Dementsprechend lautete beispielsweise ein Item in der Originalversion von Böck (2010):

**Triumph**

Triumph bedeutet eine/-n

- Verlust
- Nachteil
- Erfahrung
- Erfolg

im Kampf zu haben und eine/-n

- Gleichstand gegenüber
- Niederlage gegenüber
- Sieg über
- Friedenskuss mit

dem/den Gegner.

(Item übernommen aus Böck, 2010)

In der vorliegenden Version wurde dann versucht, Artikel oder Präpositionen direkt an richtiger Stelle einzubauen, um die Lesbarkeit des Satzes zu vereinfachen.

### Triumph

Triumph bedeutet

- einen Verlust
- einen Nachteil
- eine Erfahrung
- einen Erfolg

im Kampf zu haben und

- einen Gleichstand gegenüber dem
- eine Niederlage gegenüber dem
- einen Sieg über dem
- einen Friedensschluss mit dem

Gegner.

Inhaltlich wurden die Testaufgaben allerdings, bis auf die Reihenfolge, ident zu jenen der Vorgängerstudie (Böck, 2010) übernommen.

Hervorgehend aus der Vorerhebung, die zur Überprüfung der Verständlichkeit und der notwendigen Testdauer diente, wurde entschieden, dass die Testpersonen für die Bearbeitung maximal 15 Minuten Zeit haben sollten.

## 7.4 Alltagswissen

Im AID 2 findet sich ebenfalls ein Subtest mit dem Titel „Alltagswissen“. „Der Untertest 1 Alltagswissen soll die Fähigkeit prüfen, sich Sachkenntnisse über Inhalte anzueignen, die in der heutigen Gesellschaft alltäglich sind.“ (Kubinger, 2009, S. 9) Die Vorgabe erfolgt mündlich. Die Testpersonen sollen ebenfalls verbal eine alltägliche Frage zu Sachwissen beantworten.

Der vorliegende Untertest „Alltagswissen“ ist dem Subtest aus dem AID 2 sehr ähnlich. Die Testpersonen sollten 15 Alltagsfragen kurz und knapp, meistens reichte ein Name oder Wort aus, schriftlich in freiem Antwortformat beantworten. Nachfolgend eine Beispielaufgabe:

Wie viele Beine haben Spinnentiere? \_\_\_\_\_

(Item übernommen von Eiter, 2011)

In der Formulierung der Fragen der Vorgängerstudie (Eiter, 2011) tauchten immer wieder Ungereimtheiten und unklare Formulierungen auf. Innerhalb dieser Studie wurde allerdings darauf verzichtet diese Items zu berichtigen, da die Vergleichbarkeit der Daten aus den beiden Studien ansonsten nicht mehr gewährleistet wäre.

Lediglich ein Item wurde vorab berichtet. Im Anschluss wurde es allerdings aus der Analyse ausgeschlossen, da diese Korrektur zu deutlich abweichenden Antworten führte.



Aus der Vorerhebung mit zwei Testpersonen ergab sich, dass die Bearbeitungszeit von 10 Minuten für diesen Subtest auch für Testpersonen mit langsameren Lese- und Schreibtempo absolut ausreicht. Dieser Untertest stellte den Abschluss der Testung dar.

### **7.5 Überarbeitung der Instruktion**

Für die Testvorgabe innerhalb dieser Studie wurden die Instruktion und die Vorgabe geringfügig überarbeitet. Die Vorgabe erfolgte nicht, wie in den Vorgängerstudien, mit einzelnen Blättern, sondern es wurde für jede Testperson ein Testheft angefertigt.

Die Instruktion wurde mit Hilfe zweier Testpersonen (im Alter von 9 und 12 Jahren) auf Verständlichkeit überprüft und gegebenenfalls verändert. Vor allem die Instruktion für den Subtest „Funktionen Abstrahieren“ stellte sich als zu ausführlich und komplex heraus.

Außerdem wurde im Layout auf eine übersichtliche Darstellung großer Werte gelegt. Wenn möglich sollte ein Item immer abgeschlossen in einer Zeile dargestellt werden.

Anstelle von Buchstaben wurden die Distraktoren durch kleine Kästchen gekennzeichnet, die die Testpersonen dann ankreuzen sollten.

Für den Subtest „Wortschatz“ wurde ein zusätzliches Element zur Klarheit eingeführt. Die Testpersonen sollten in diesem Untertest an zwei Stellen in einem Satz einen der Distraktoren auswählen. Um klar zu stellen, dass an beiden Stellen im Satz eine Auswahl zu treffen ist, wurden die einen Distraktoren hellgrau und die anderen dunkelgrau unterlegt. In der Instruktion konnte so einfach klargestellt werden, dass sowohl aus dem hell- als auch aus den dunkelgrauen Kästchen jeweils eine Antwortmöglichkeit anzukreuzen sei.

Sämtliche Änderungen in der Vorgabe sollten zu keinen abweichenden Ergebnissen führen, sondern lediglich das Prozedere vereinfachen und zu weniger Zwischenfragen seitens der Schüler und Schülerinnen führen.

## 8. Testdesign

Wie bereits eingangs erwähnt, wurde für die vorliegende Studie ein verlinktes Design gewählt. Innerhalb der schon mehrfach genannten Vorgängerstudien von Böck (2010) und Eiter (2011) wurde bereits ermittelt, ob die Testaufgaben innerhalb der vorgegebenen Testformen die notwendige Voraussetzung für die weitere Verwendung, nämlich die der *Rasch*-Modell-Konformität, erfüllen.

Um nun ermitteln zu können, ob auch der komplette Untertest, beziehungsweise das vollständige Itempool eines Subtests, *Rasch*-Modell-konform ist, wurden die Testformen verknüpft, beziehungsweise die Testaufgaben verlinkt.

Verlinkte Testdesigns basieren darauf, dass eine Gruppe von Testpersonen Teil A der Items vorgegeben wird und einer anderen Gruppe von Testpersonen Teil B. Allerdings befinden sich sowohl in Teil A, als auch in Teil B, teilweise idente Items, die sogenannten „linking items“. Üblicherweise kann eine derartige Verlinkung so gestaltet werden, dass es nur einige wenige „linking items“ gibt, die die Testformen verknüpfen. Da es innerhalb dieser Studie aber von Vorteil war die gesamte Stichprobe zu maximieren, erschien es sinnvoll, möglichst große Überschneidungen zu erhalten.

In den vorhergehenden Studien von Bock (2010) und Eiter (2011) wurden für jeden Subtest jeweils vier Testformen, die je nach Untertest 15 bis 20 Items beinhalteten, vorgegeben. Diese vier Testformen wurden im Rahmen dieser Arbeit derart restrukturiert, dass drei Testformen entstanden sind. Diese Testformen wurden sich überschneidenden Gruppen vorgegeben. Abbildung 1 dient der Veranschaulichung.

Urspr. Testform 1 Items 1-15	Urspr. Testform 2 Items 16-30	Urspr. Testform 3 Items 31-45	Urspr. Testform 4 Items 46-60
Neue Testform 1 Items 9-23		Neue Testform 2 Items 24-38	Neue Testform 3 Items 39-53

Abbildung 1: Darstellung des Testdesigns

Mittels des Softwarepackages eRM in „R“ lässt sich der LR-Test nach Andersen mit derart verlinkten Designs berechnen.

## **9. Durchführung der Studie**

Dieses Kapitel widmet sich der Erhebung der Daten, also die Stichprobenakquise und -beschreibung, Beschreibung der Testung selbst und die deskriptive Analyse der Stichprobe.

### **9.1 Stichprobenakquirierung**

Nachdem zu dieser Studie bereits zwei Vorgängerstudien (Böck, 2010 und Eiter, 2011) bestehen, war es, um die Vergleichbarkeit der Daten zu gewährleisten, notwendig, weitgehend dieselben Schüler und Schülerinnen zu testen. Die Datenerhebungen von Böck und Eiter erfolgte 2010 in vier verschiedenen oberösterreichischen Schulen.

Mit Unterstützung von Johanna Böck, der Autorin der Vorgängerstudie, die persönliche Kontakte zu den Schulen hatte, wurde im Frühjahr 2011 mit der Kontaktaufnahme zu den betreffenden Schulen begonnen. Im April 2011 erfolgte das Ansuchen um Genehmigung der Testung beim Landesschulrat Oberösterreich (die Projektbeschreibung findet sich im Anhang).

Konkret handelte es sich um eine Volksschule, in der auf Grund der notwendigen Lesekompetenz nur die 3. und 4. Schulstufe getestet wurden, zwei Hauptschulen und ein Privatgymnasium. Die Datenerhebung erfolgte, komprimiert an einigen Tagen, im Juni 2011.

Für die Vorbereitung der Schuldirektoren und-direktorinnen und der Lehrerschaft sowie der Eltern, wurde ein Elternbrief, mit einer kurzen Projektbeschreibung und der Bitte um Einverständnis in die Testung, formuliert (auch der Elternbrief ist im Anhang nachzulesen). Die Reaktionen der Eltern auf den Elternbrief waren unterschiedlich. So haben manche Eltern von der Möglichkeit der persönlichen Kontaktaufnahme mit der Testleiterin Gebrauch gemacht und sich telefonisch genauer informiert. Die Rücklauf- und Einverständnisquote der Eltern unterlag klassenweise großen Schwankungen. Leider konnte nicht nachvollzogen werden aus welchen Gründen sich Eltern gegen die Teilnahme ihrer Kinder an der Testung entschieden. Sowohl Kinder, deren Eltern eindeutig abgelehnt hatten, als auch Kinder, die die Einverständniserklärung nicht abgegeben hatten (entweder, weil die Eltern kein Interesse hatten, oder weil die Einverständniserklärung verloren, bzw. vergessen wurde) nahmen nicht an der Testung teil.

Außerdem konnte nicht nachvollzogen werden, warum in manchen Klassen alle Eltern zustimmten und in anderen nur etwa zwei Drittel der Schüler und Schülerinnen teilnehmen durften. Denkbar ist allerdings die Hypothese, dass die Eltern vor allem am Land gut vernetzt sind. Wenn dann manche Eltern gegen die Teilnahme an der Testung sind, überzeugen sie möglicherweise in persönlichen Gesprächen auch andere Eltern davon ihre Kinder nicht mitmachen zu lassen, vor allem, weil die letzte Testung in diesen Schulen noch nicht lange zurück lag.

## 9.2 Testvorgabe

Für die Testung wurden die drei Parallelformen in Form von drei farblich unterschiedlichen Testheften vorbereitet. So konnte gewährleistet werden, dass auf einen Blick erkennbar ist, dass Sitznachbarn beziehungsweise Sitznachbarinnen nicht dieselben Testaufgaben zur Bearbeitung erhalten. Für die Testung wurde die Dauer von einer Schulstunde, also 50 Minuten, veranschlagt, was für die meisten Testpersonen auch ausreichend war. Die Testung erfolgte zumeist im Klassenzimmer, manchmal wurde aus logistischen Gründen auch in andere Räume, wie zum Beispiel den PC-Raum, ausgewichen. Pro Klasse war ein Testleiter beziehungsweise eine Testleiterin anwesend. Zumeist waren die Lehrer oder Lehrerinnen zu Beginn, für die Einführung, dabei und zogen sich dann zurück.

Anfangs erfolgte für die Schüler und Schülerinnen eine ausführliche Instruktion, in der sie über den Ablauf der Testung informiert wurden. Zudem wurde ihnen mitgeteilt, dass die Ergebnisse ausschließlich zu Forschungszwecken verwendet werden und nicht an die Lehrer und Lehrerinnen weitergeleitet oder gar in die Note miteinbezogen werden. Nachdem unterschiedliche Testleiter beziehungsweise Testleiterinnen zum Einsatz kamen, wurde eine schriftliche Instruktion vorbereitet, sodass gewährleistet werden konnte, dass alle Schüler und Schülerinnen dieselben Informationen erhalten.

Für jeden Subtest wurden die entsprechenden Übungsaufgaben mündlich gelöst und es wurde sichergestellt, dass alle Testpersonen die Instruktion verstanden haben, indem auch abweichende (falsche) Lösungen besprochen wurden. Die Subtests wurden dann eigenständig bearbeitet und am Ende jedes Subtests wurden die Testpersonen dazu angehalten zusammenzuwarten, sodass die nächste Übungsaufgabe gemeinsam besprochen werden konnte. Falls jemand sehr viel schneller war, wurde ein anderes Testheft zur Beschäftigung vorgegeben. Schüler und Schülerinnen, die keine Einverständniserklärung hatten und auch keine andere Beschäftigung für die Dauer der Testung hatten, konnten an der Testung teilnehmen, allerdings wurden diese Testbögen anschließend vernichtet.

Falls am Ende noch Zeit blieb, verbrachte der/die Testleiter/-in die restliche Dauer der Schulstunde in der Klasse. Oft wurden, nach Absammeln der Testbögen, offene Fragen beantwortet oder die Schüler/-innen hatten Fragen zum Studium oder zur Diplomarbeit. Als kleines Dankeschön erhielten die Testpersonen nach der Testung etwas Süßes, worüber sie sich sehr gefreut haben.

### **9.2.1 Beobachtungen bei der Datenerhebung**

Die Datenerhebung verlief weitgehend reibungslos. Die Lehrer und Lehrerinnen zeigten sich wohlwollend und interessiert und auch die Schüler und Schülerinnen gingen größtenteils motiviert an die Bearbeitung der Aufgaben heran. Die Tatsache, dass in der vorletzten Schulwoche getestet wurde, sorgte für eine unbeschwerte, lockere Atmosphäre ohne übertriebenen Leistungsdruck. Die sehr einfachen Aufgaben sorgten bei den älteren Testpersonen für großes Amusement, sodass deren Ernsthaftigkeit bei der Bearbeitung dieser Items teilweise anzuzweifeln ist.

Für einzelne Volksschüler gestaltete sich das Verstehen der Instruktion teilweise etwas schwierig und mündliche Erklärungen waren notwendig. Immer wieder waren auch gewisse Begriffe, die für die Lösung des Items notwendig waren, unklar. Diese Begriffe wurden dann kurz erklärt.

In einer Klasse musste kurzfristig die Reihenfolge der Testaufgaben geändert werden, da die Testleiterin sich um ein krankes Kind kümmern musste und für keine Fragen zur Verfügung stand. Daher wurde mit dem Subtest „Angewandtes Rechnen“ begonnen, da es bei diesem Subtest erfahrungsgemäß zu weniger Zwischenfragen kam.

Insgesamt wurden sechs Testbögen aus der Analyse ausgeschlossen. Eine Person konnte nicht bis zum Ende mitmachen, weil sie auf Grund einer Krankheit bereits nach 15 Minuten den Unterricht verließ. Die fünf anderen Schüler und Schülerinnen waren Integrationskinder in Integrationsklassen. Diese Kinder haben zumeist selbstständig oder mit Unterstützung einer Hilfslehrerin an der Testung teilgenommen, konnten aber nur einen Bruchteil der Aufgaben lösen und haben auch die Instruktion nicht selbstständig verstanden.

## **10. Deskriptive Statistiken**

Die interessierenden deskriptiven Statistiken bezüglich der vorliegenden Daten beschränken sich auf die Variablen Geschlecht, Schulstufe und Schultyp.

In einem ersten Schritt wird auf die Datenerhebung, die im Rahmen dieser Studie durchgeführt wurde, eingegangen.

Nachdem die weitere Berechnung allerdings in Kombination mit den im Vorfeld von Böck (2010) und Eiter (2011) erhobenen Daten erfolgt, wird in einem zweiten Schritt auch die Gesamtgruppe deskriptiv betrachtet.

## 10.1 Beschreibung der aktuellen Daten

Im Juni 2011 wurden 667 Kinder und Jugendliche mit der vorliegenden Testbatterie getestet.

Die Geschlechterverteilung ist annähernd homogen. An der Studie nahmen 356 Burschen und 311 Mädchen teil. In Prozent ausgedrückt waren es 53,4 Prozent Burschen und 46,6 Prozent Mädchen.

Nahezu alle Testpersonen gaben als Muttersprache Deutsch an. Lediglich 1,5 Prozent, also 10 Personen, hatten eine andere Muttersprache (Serbisch, Albanisch, Türkisch, Tschechisch, Arabisch, Niederländisch).

Lediglich 4,7 Prozent der Gruppe waren Volksschüler, 44,2 Prozent der Testpersonen waren Hauptschüler und 51,1 Prozent besuchten das Linzer Privatgymnasium.

Jeweils rund 20 Prozent der Gruppe waren in der 5., 6., 7. beziehungsweise 8. Schulstufe. Der Rest der Gruppe verteilte sich auf die 3., 4. und 9. Schulstufe. Dementsprechend liegt auch der Altersmedian in diesem Bereich, nämlich bei 156 Monaten, beziehungsweise 13 Jahren.

## 10.2 Beschreibung der Gesamtgruppe

Nachdem in weiterer Folge mit den Daten aus der vorliegenden aber auch aus den vorhergehenden Studien gearbeitet wird, sei auch diese Gesamtgruppe an dieser Stelle deskriptiv beschrieben.

Die Gesamtgruppe umfasste zwischen 1342 und 1346 Personen. Diese Anzahl schwankt je nach Subtest, da Böck (2010) und Eiter (2011) die Subtests nicht in einem Testheft, sondern einzeln ausgegeben und abgesammelt hatten. So kam es dazu, dass einzelne Blätter verschwunden sind, wodurch für jeden Subtest eine geringfügig abweichende Anzahl an Testpersonen entstand.

### 10.2.1 Geschlecht

Tabelle 1 fasst die Geschlechterverteilung für die vier Subtests zusammen. Generell kann aber angemerkt werden, dass die Geschlechterverteilung als homogen zu betrachten ist.

Tabelle 1: Geschlechterverteilung in der Stichprobe

Subtest	Funktionen Abstr.		Angew. Rechnen		Synonyme Finden		Alltagswissen	
	männlich	weiblich	männlich	weiblich	männlich	Weiblich	männlich	weiblich
<b>Geschlecht</b>								
<b>Anzahl</b>	710	632	710	633	710	636	709	636
<b>Prozent</b>	52,9	47,1	52,7	47,0	52,7	47,3	52,6	47,3
<b>Fehlend</b>	4		3		0		1	
<b>Gesamt</b>	1342		1343		1346		1345	

### 10.2.2 Alter

Analog zur oben ausgeführten Altersverteilung, fällt die Altersverteilung der Gesamtgruppe aus. Nachdem in den Vorgängerstudien (Böck, 2010; Eiter, 2011) keine Angaben zum Alter gemacht wurden, musste in Anlehnung an das Alter mit der Schulstufe gearbeitet werden. Auch wenn klar ist, dass diese Variable durch Repetenten und übliche Schwankungen des Altersbereichs innerhalb einer Schulklasse einer gewissen Verzerrung unterliegt. Dennoch erscheint die Verwendung der Schulstufe schlüssig, da viele Testaufgaben sich an der Schulstufe, beziehungsweise am vorgesehenen Lehrplan, orientieren.

Tabelle 2: Altersverteilung der Gesamtgruppe

Schulstufe	Funktionen Abstr.	Angew. Rechnen	Synonyme Finden	Alltagswissen
3.	33	32	33	33
4.	33	33	33	33
5.	296	295	296	296
6.	286	288	288	290
7.	308	308	308	304
8.	286	286	287	287
9.	101	101	101	102
Fehlend	3	3	0	1
<b>Gesamt</b>	1343	1343	1346	1345

Leider konnten keine weiteren deskriptiven Daten, wie beispielsweise die Muttersprache der Gesamtgruppe, demonstriert werden, da in den entsprechenden Vorgängerstudien keine Angaben dazu gemacht wurden. In der nachfolgenden Analyse hinsichtlich der *Rasch*-Modell-Konformität der Items wurden, neben dem Teilungskriterium Score, die dargestellten Kriterien Geschlecht und Alter zur Bildung der Subgruppen herangezogen.

## 11. Analyse der Rasch-Modell-Konformität der Items

Die vorliegenden Itempools der vier Subtests sollten mittels des *Likelihood-Ratio-Tests* (LRT) nach Anderson (1973) auf Rasch-Modell-Konformität überprüft werden. Zur Analyse und Veranschaulichung der einzelnen Items wurde der Grafische Modelltest herangezogen. Im Falle eines signifikanten LRT lieferten der z-Test nach Fischer und Scheiblechner, sowie die Item-Fit-Statistik nach Hatzinger und Mair, zusätzliche Informationen über die Güte der Items. Ausgeschlossen wurden die Items allerdings nur dann, wenn plausible inhaltliche Gesichtspunkte formuliert werden konnten.

Folgende Teilungskriterien wurden für die Analyse verwendet:

### Internes Teilungskriterium:

- **Score:**

Hier wurden Rohwerte, die unter dem Median lagen oder dem entsprachen jenen Rohwerten, die über dem Median lagen, gegenübergestellt.

### Externe Teilungskriterien:

- **Geschlecht:**

Die Stichprobe wurde in männliche und weibliche Testpersonen geteilt.

- **Alter:**

Auf Grund der fehlenden Altersangaben in den Vorstudien wurde dieses Teilungskriterium im Sinne der Schulstufe interpretiert, ungeachtet dessen, dass Repetenten und Schüler, die Klassen überspringen, die Altersangaben verzerren. Dementsprechend wurde die Gruppe am Median der Schulstufen geteilt.

Konkret bedeutet das 3. bis 6. Schulstufe (ca. 8 bis 12 Jahre) versus 7. bis 9. Schulstufe (ca. 13 bis 16 Jahre)

Die Items sämtlicher Subtests wurden bezüglich folgender forschungsleitender Nullhypothese begutachtet: „Das dichotom logistische Modell von Rasch gilt“. Die entsprechende Alternativhypothese lautet demnach: „Das dichotom logistische Modell von Rasch gilt nicht“. Die Überprüfung der Daten erfolgte mit einer Irrtumswahrscheinlichkeit von  $\alpha=0.05$ , das gültige Signifikanzniveau wurde allerdings auf  $\alpha=0.01$  adjustiert um den Fehler 1. Art nicht allzu groß werden zu lassen, da jeder der Datensätze drei abhängigen LRT (entsprechend der drei Teilungskriterien) unterzogen wurde.



Bei Auftreten von signifikanten Modellabweichungen, die außerdem inhaltlich plausibel erschienen, wurden sukzessive Items ausgeschieden um letztlich a posteriori *Rasch*-Modell-Konformität zu erreichen. Gleichzeitig wurde aber im Vorhinein festgelegt, dass nicht mehr als 20 Prozent der vorliegenden Items ausgeschieden werden sollten, da andernfalls die Güte des Subtests anzuzweifeln ist.

## 11.1 Funktionen Abstrahieren

In einer ersten Parameterschätzung für den ersten vorgegebenen Subtest ergaben sich, wie Tabelle 3 zu entnehmen ist, signifikante Modellabweichungen.

Von den insgesamt 60 Items im Subtest „Funktionen Abstrahieren“, konnte in zwei von drei Teilungskriterien zumindest eines der Items nicht geschätzt werden, da diese Items auf Grund ungünstiger Verteilungen in den Subgruppen von der Berechnung ausgeschlossen wurden. Konkret bedeutet das, dass diese nicht geschätzten Items von allen Testpersonen oder keiner Testperson innerhalb einer Subgruppe gelöst wurden.

Wie viele Items pro Teilungskriterium geschätzt wurden, geht aus Tabelle 1 hervor und ist durch  $df+1$  definiert.

Tabelle 3: Erste Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Funktionen Abstrahieren“

Teilungskriterium	$\chi^2$ (LRT)	Df	$\chi^2_{krit} (\alpha=0,01)$	P
Score	201.271	58	85.95	0.000
Geschlecht	353.894	55	82.29	0.000
Alter	151.093	59	87.17	0.000

Auf Grund der Abweichungen im Grafischen Modelltest, dem z-Test und der Item-Fit-Statistik zeigte sich welche Items sich nicht modellkonform verhalten, da diese Verfahren im Gegensatz zum Likelihood-Ratio-Test eine itemweise Überprüfung der Testaufgaben ermöglichen (Kubinger 1989).

Nach inhaltlichen Analysen zur Güte der Items, wurden die auffälligen Testaufgaben sukzessive ausgeschlossen; dementsprechend wurde nach Ausschluss eines Items der LRT für alle Teilungskriterien erneut berechnet.

Die Grafischen Modelltests nach der ersten Parameterschätzung können den Abbildungen 2 bis 4 entnommen werden. Die Ergebnisse der z-Test und der Item-Fit-Statistik kann der interessierte Leser im Anhang nachlesen. An dieser Stelle werden nur die auffälligen Items besprochen.

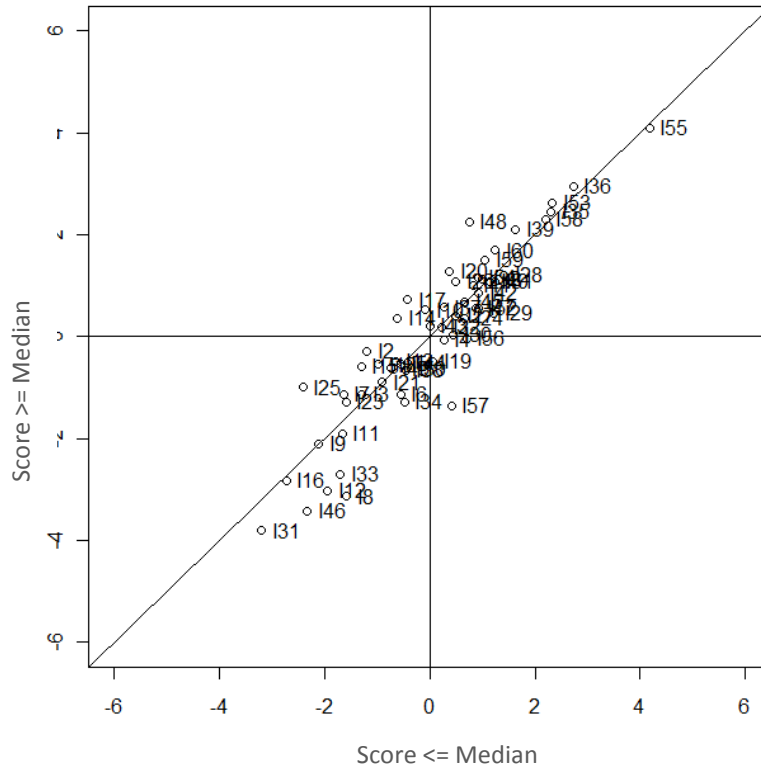


Abbildung 2: Grafischer Modelltest für das Teilungskriterium Score vor Ausschluss eines Items, Subtest „Funktionen Abstrahieren“

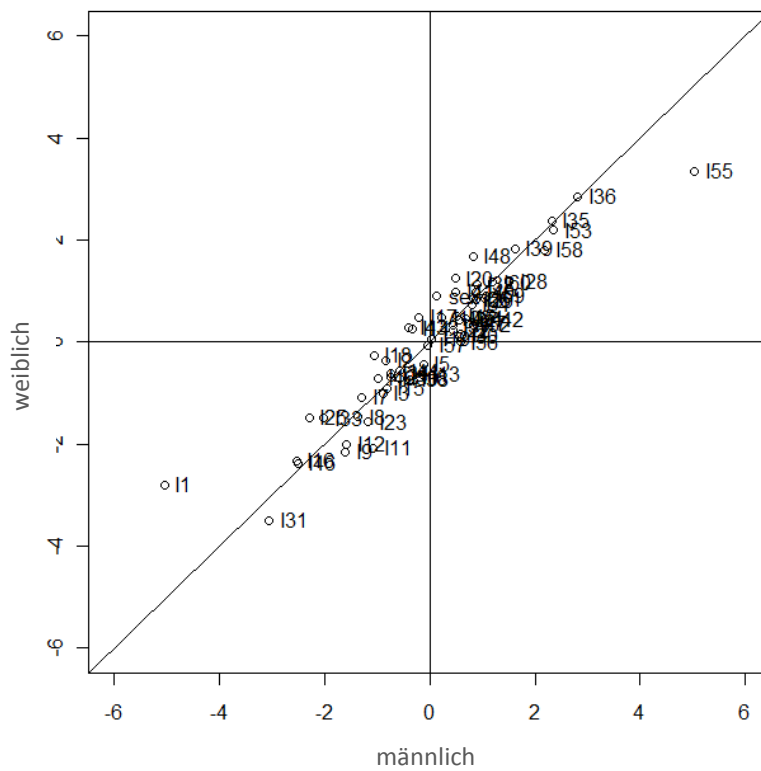


Abbildung 3: Grafischer Modelltest für das Teilungskriterium Geschlecht vor Ausschluss eines Items, Subtest „Funktionen Abstrahieren“

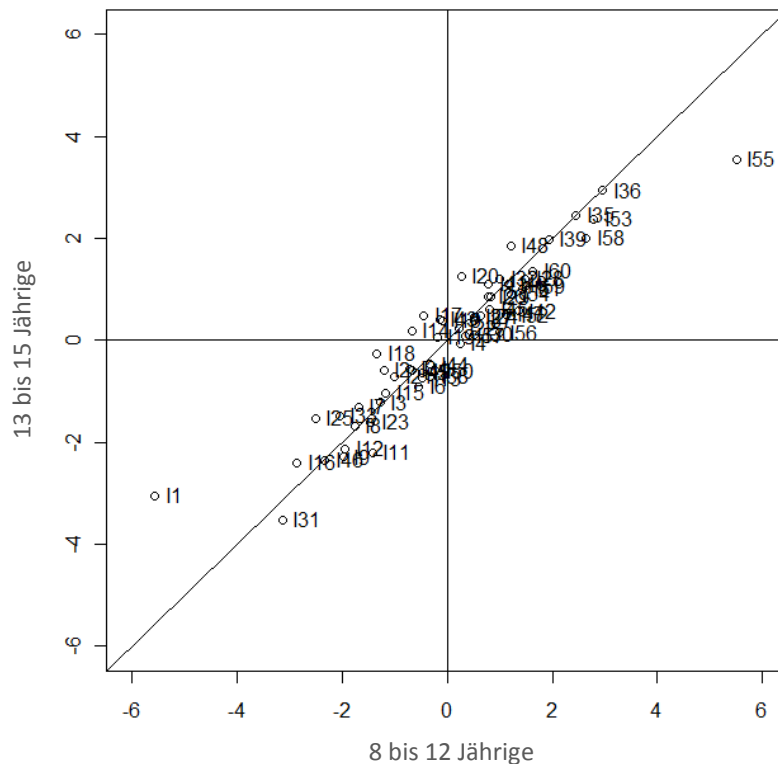


Abbildung 4: Grafischer Modelltest für das Teilungskriterium Alter vor Ausschluss eines Items, Subtest „Funktionen Abstrahieren“

### 11.1.2 Itemselektion

Aus den vorliegenden Verfahren zur Überprüfung wurden nun die auffälligen Items ersichtlich: Der z-Test wurde mit einem strengen Signifikanzniveau von  $\alpha=0.01$  berechnet, da die meisten vorgegebenen Items sogenannte „linking items“ waren und daher ohnehin eher signifikant ausfallen.

Im z-Test wurden beim internen Splitkriterium die Items 14, 15, 18, 20, 25, 29 und 48 signifikant. Items 1 wurde a priori auf Grund ungünstiger Verteilungen in den Subgruppen ausgeschlossen (siehe oben). Beim Splitkriterium Geschlecht wurde das Item 2 signifikant. Item 1 wurde ebenfalls a priori nicht in die Berechnung miteinbezogen. Für das dritte Splitkriterium (Alter) wurden im z-Test die Items 14, 17, 18, 20, 25 und 42 signifikant.

In der Itemfit-Statistik fielen die Items 2, 4, 14, 17, 20, 25, 32, 36, 39 und 48 mit signifikanten  $p$ -Werten auf.

Diese angegebenen Items zeigten zu einem überwiegenden Teil auch in den Grafischen Modelltests eine abweichende Position vom Modell (siehe Abbildungen 2, 3 und 4).

In der Vorgängerstudie von Böck (2010) wurden die folgenden Items nach der ersten Modellschätzung ausgeschlossen: 2, 17, 25, 48 und 57.

Nachdem die genannten Items bis auf das Item 57 (hier liegt aber beim internen Splitkriterium eine deutliche Abweichung im Grafischen Modelltest vor) auch im Rahmen dieser Studie signifikante Werte aufwiesen, erschien es nur allzu schlüssig, die Testaufgaben auch hier auszuschließen. Böck (2010) hat zu diesen Items bereits ausführlich auf mögliche Fehlerquellen analysiert, daher soll hier nur noch kurz und ergänzend darauf eingegangen werden.

Des Weiteren wurden nach Analyse von Item-Fit-Statistik, z-Test und Grafischem-Modelltest die Items 1, 14, 18, 20, 29, 55 als auffällig identifiziert und nach inhaltlicher Analyse sukzessive ausgeschlossen.

### 11.1.3 Itemanalyse

Inhaltliche Analyse Item 1:

#### Apfel

Schokolade    Birne    Pommes    Zwetschke    Eis

(Item übernommen aus Böck, 2010)

Das Item 1 wurde im internen Teilungskriterium und im Teilungskriterium Geschlecht von vornherein ausgeschlossen, da die Verteilung innerhalb der Subgruppen ungünstig war. Zusätzlich lag es auch im Grafischen Modelltest für das Teilungskriterium Alter weit von der Geraden entfernt.

Die Itemparameter, (Tabelle c) zeigen zudem für das Teilungskriterium Alter, dass das Item für die älteren Testpersonen vergleichsweise schwerer zu lösen war, als für die jüngeren. Im Rahmen dieser Studie werden durchwegs die Leichtigkeitsparameter berechnet, da in *eRm* ausschließlich diese angegeben werden. Dementsprechend gilt, je höher der Leichtigkeitsparameter, desto leichter ist das Item.

Grundsätzlich ist zu bedenken, dass in diesem Subtest nach einer gemeinsamen Funktion oder Verwendung gefragt wird. Auch wenn man die Testaufgabe richtig beantwortet Apfel/Birne/Zwetschke wird hier keine gemeinsame *Funktion*, sondern lediglich eine Gemeinsamkeit, also alle drei Begriffe beschreibe eine Obstsorte, bezeichnet. Die gemeinsame Verwendung wäre, dass alle drei Objekte essbar sind, allerdings bezeichnen auch die anderen Begriffe Nahrungsmittel.

Daher ist bei diesem Item von einem Konstruktionsfehler auszugehen, da die korrekte Beantwortung im engeren Sinne nicht der Instruktion entspricht.

#### Inhaltliche Analyse Item 2:

##### **Lockenstab**

Kamm    Schere    Rundbürste    Lockenwickler    Pinzette

(Item übernommen aus Böck, 2010)

Dieses Item wurde bereits bei Böck (2010) aus dem Itempool entfernt. Im Teilungskriterium Geschlecht zeigte sich ein signifikantes Ergebnis ( $p=0.001$ ) und die Itemparameter weisen darauf hin, dass das Item den männlichen Testpersonen schwerer gefallen ist. Böck (2010) merkt dazu an, dass Mädchen eher wissen würden wozu die genannten Objekte verwendet werden und daher einen Vorteil in der Beantwortung haben könnten. Außerdem sei es aufgefallen, dass die Burschen vermehrt nachgefragt hätten was ein Lockenstab sei, was auch in der aktuellen Studie auffallend war.

Demnach ist hier von einem beträchtlichen Einfluss an geschlechtsspezifischem Fachwissen auszugehen.

#### Inhaltliche Analyse Item 14:

##### **Eishockey**

Tennis    Golf    Basketball    Polo    Volleyball

(Item übernommen aus Böck, 2010)

Item 14 wurde in zwei Splitkriterien (internes Teilungskriterium,  $p=0.001$ ; Alter,  $p=0.004$ ) und der Item-Fit-Statistik ( $p=0.000$ ) signifikant. Betrachtet man die dazugehörigen Leichtigkeitssparameter (Tabelle c), wird deutlich, dass die Personen mit einem höheren Score (als der Median) und einem höheren Alter sich vergleichsweise schwerer mit dem Item getan haben.

In der Testvorgabe war auffallend, dass sehr viele Testpersonen nicht wussten welche Sportart Polo ist, beziehungsweise wie Polo gespielt wird. Dementsprechend ist anzunehmen, dass der Einfluss von Fachwissen hier eine wichtige Rolle spielt. Somit könnte auch erklärt werden, warum generell fähigere Personen gerade in diesem Item schlechter abschneiden, da nicht nur Abstraktionsvermögen, sondern auch (sportliches) Allgemeinwissen zur korrekten Beantwortung nötig ist und damit die Eindimensionalität nicht gegeben ist.

#### Inhaltliche Analyse Item 17:

##### **Mikrofon**

Videorekorder    Lausprecher    Hörgerät    Antenne    Scheinwerfer

(Item übernommen aus Böck, 2010)

Auch dieses Item wurde bereits in der Vorgängerstudie (Böck 2010) ausgeschieden. In der vorliegenden Studie wurde das Item in der Item-Fit-Statistik ( $p=0.000$ ) und im Split-Kriterium Alter ( $p=0.001$ ) signifikant; zudem wich es in allen drei Kriterien im Grafischen Modelltest deutlich von der Geraden ab. Der Itemparameter zeigt, dass die Testpersonen mit höheren Leistungen und höherem Alter bei diesem Item eher auf keine richtige Lösung kamen.

Als Erklärung dienen laut Böck (2010) erneut der Einfluss von Fachwissen und die Uneindeutigkeit des Items. Dementsprechend spielt technisches Fachwissen zur Funktionsweise der Geräte eine nicht unwesentliche Rolle. Bei Testpersonen, die komplexer dachten (vorwiegend ältere und jene mit höheren Leistungen in dem Subtest) interpretierten beispielsweise auch die Funktion des Aufnehmens von akustischen Signalen (Mikrofon, Hörgerät, Videorekorder). (Böck 2010).

Die Uneindeutigkeit dieses Items lässt sich auch noch an anderen denkbaren Lösungen demonstrieren; So haben zum Beispiel auch Mikrofon, Lautsprecher und Scheinwerfer die Funktion ein Geschehen auf einer Bühne sichtbar und hörbar zu machen.

Inhaltliche Analyse Item 18:

**Regenschirm**

- Badehaube    Schal    Gummistiefel    Stirnband    Stöckelschuhe

(Item übernommen aus Böck, 2010)

Item 18 wurde im z-Test beim internen Splitkriterium ( $p=0.000$ ) und beim Teilungskriterium Alter ( $p=0.000$ ) signifikant. Die dazugehörigen Leichtigkeitsparameter finden sich in Tabelle c.

Nachdem in diesem Subtest immer die zwei Begriffe gefunden werden sollte, die eine gemeinsame Funktion mit dem fett gedruckten Begriff erfüllen, liegt hier ein Konstruktionsfehler nahe. Je nach Interpretation des Items sind zwei Antwortmöglichkeiten denkbar.

Dementsprechend haben Regenschirm, Badehaube und Gummistiefel gemeinsam, dass sie vor Wasser schützen. Regenschirm, Schal und Stirnband erfüllen allerdings die gemeinsame Funktion vor Witterung zu schützen.

Inhaltliche Analyse Item 20:

**Förster**

- Schriftsteller    Polizist    Reinigungskraft    Maler    Maurer

(Item übernommen aus Böck, 2010)

Dieses Item zeigte bei den Teilungskriterien Score ( $p=0.001$ ) und Alter ( $p=0.000$ ) und in der Item-Fit-Statistik ( $p=0.009$ ) signifikante Ergebnisse. Den Leichtigkeitsparametern (Tabelle c) entsprechend war die Testaufgabe für die Testpersonen mit höheren Leistungen und höherem Alter schwerer.

Erneut ist hier auf einen Konstruktionsfehler und die damit einhergehende Uneindeutigkeit hinzuweisen, da die Berufsgruppen Förster, Polizist und Reinigungskraft eine ordnende und kontrollierende Funktion ausüben, gleichzeitig Förster, Maler und Mauerer aber handwerkliche Berufe darstellen.

Inhaltliche Analyse Item 25:

#### **Springschnur**

Murmeln    Trampolin    Ball    Sprungbrett    Sandschaufel

(Item übernommen aus Böck, 2010)

Item 25 wies beim internen Teilungskriterium ( $p=0.000$ ) sowie beim Alter ( $p=0.006$ ) und in der Item-Fit-Statistik ( $p=0.000$ ) signifikante Werte auf. Aus den Itemparametern lässt sich ableiten, dass Testpersonen mit besseren Leistungen in diesem Subtest deutliche Schwierigkeiten hatten das Item zu lösen.

Auf den ersten Blick fällt die Beantwortung leicht; sowohl die Springschnur, als auch das Trampolin und das Sprungbrett sind Spiel- bzw. Sportgeräte, die springend genutzt werden. Allerdings haben das Trampolin und das Sprungbrett, im Gegensatz zu der Springschnur, die zusätzliche Funktion rückfedernd zu sein. Dementsprechend ist es durchaus denkbar, dass Testpersonen, die komplexer dachten andere Lösungen fanden.

Böck (2010) weist demnach beispielsweise darauf hin, dass vor allem Personen mit höheren Rohwerten auch noch die Bewegungsart des Spielgeräts bedachten. Sowohl die Murmeln, der Ball, als auch die Springschnur werden zur Verwendung in „Drehung“ versetzt.

Murmeln, Ball und Springschnur ist außerdem gemeinsam, dass sie als Spielgeräte in der Gruppe, bzw. für Mannschaftssport, verwendet werden können.

Inhaltliche Analyse Item 29:

#### **Ampel**

Lampe    Bodenmarkierung    Auto    Straßenschild    Verkehr



(Item übernommen aus Böck, 2010)

Aus den Itemparametern (Tabelle c) kann man schließen, dass Item 29 generell ein schwer zu lösendes Item ist. Besondere Schwierigkeiten hatten allerdings Personen mit höheren Testwerten. Beim internen Splitkriterium ergab sich ein signifikanter Wert ( $p=0.008$ ).

So fiel in der Sichtung der Testbögen ein Antwortmuster auf, dementsprechend es sein könnte, dass Kinder, die noch etwas abstrakter in der Itemlösung vorgehen, an die direkte (optische) Funktion der Ampel, also das Leuchten, dachten und somit Lampe ankreuzten. Die dritte Lösung war abgeleitet daraus dann Auto, da auch dieses an gewissen Stellen leuchtet.

Inhaltliche Analyse Item 48:

**Socken**

- Handschuhe    Schuhe    Strumpfhose    Bademantel    Mütze

(Item übernommen aus Böck, 2010)

Diese Testaufgabe wurde im z-Test beim internen Splitkriterium ( $p=0.000$ ) und in der Item-Fit-Statistik ( $p=0.000$ ) signifikant. Außerdem wurde es bereits in der Vorgängerstudie ausgeschieden.

Böck (2010) führt hier einen Konstruktionsfehler an. Als gemeinsame Funktion wurde das Warmhalten (der Füße) verstanden, wonach Schuhe und Strumpfhose die richtige Antwort darstellen. Schuhe hätten aber nicht immer eine wärmende Funktion und gleichzeitig wärmen Handschuhe und Mütze den Körper auch.

Es erscheint also nur allzu schlüssig, dass wieder einmal vor allem Testpersonen mit höheren Scores eine Antwort auf einer höheren Abstraktionsebene gefunden und das Wärmen als generelle Funktion, dem „Schutz“ eines spezifischen Körperteils (nämlich der Füße) vorgezogen haben.

Inhaltliche Analyse Item 55:

**Hass**

- Liebe    Gleichgültigkeit    Vernunft    Freude    Gefühl

(Item übernommen aus Böck, 2010)

Item 55 wurde ausgeschlossen, da es im Grafischen Modelltest in den Splitkriterien Geschlecht und Alter sehr deutlich von der Geraden abweicht. Auf Grund der Itemparameter (Tabelle c) lässt sich ableiten, dass die Testaufgabe vor allem für die Mädchen und die jüngeren Testpersonen schwerer war.

Bei korrekter Beantwortung handelt es sich bei der gemeinsamen Funktion um die Benennung von Gefühlsqualitäten (Hass/Liebe/Gleichgültigkeit). Für jüngere Kinder ist es möglicherweise recht schwierig Begriffe wie Gleichgültigkeit oder Vernunft entsprechend zu interpretieren. Demnach kommt zur Lösung des Items zusätzlich zur Abstraktionsfähigkeit auch noch Wissen über die Wortbedeutung hinzu.

Für die Geschlechtsunterschiede konnten keine eindeutigen Erklärungen gefunden werden, außer dass es plausibel erscheint, dass emotionale Begriffe bei Mädchen und Burschen im (prä)pubertären Alter unterschiedliche Assoziationen auslösen, die die Beantwortung der Frage beeinflussen.

Inhaltliche Analyse Item 57:

**Internat**

Hort    Schwimmbad    Schlafzimmer    Kindergarten    Park

(Item übernommen aus Böck, 2010)

Bereits in der Vorgängerstudie (Böck 2010) wurde dieses Item ausgeschlossen und in der vorliegenden Studie weicht es im grafischen Modelltest im internen Splitkriterium deutlich von der Geraden ab.

Böck weist darauf hin, dass die zentrale Funktion in diesem Fall nicht eindeutig ist. Dementsprechend kann (wie ursprünglich gemeint) als Funktion, die Betreuung von Kindern angenommen werden (Internat, Hort, Kindergarten). Eine andere gemeinsame Funktion wäre aber das „Schlafen“. Sowohl in einem Internat, als auch in einem Schlafzimmer übernachtet man und im Kindergarten wird auch Mittagsschlaf gehalten.

#### **11.1.4 Modellschätzung nach Itemselektion**

Nach Ausschluss der oben genannten Items, wurde der *LR*-Test erneut durchgeführt. Die Nullhypothese („Das dichotom logistische Modell von Rasch gilt“) konnte beibehalten werden und somit kann von *Rasch*-Modell-Konformität für die verbleibenden Items ausgegangen werden. Insgesamt wurden 11 Items selektiert, was 18 Prozent der Gesamtanzahl der Items entspricht.

Im internen Splitkriterium wurde nun zusätzlich Item 8 auf Grund ungünstiger Verteilung innerhalb der Subgruppen ausgeschlossen.

Tabelle 4: Parameterschätzung des Likelihood Ratio Tests (LRT) nach Itemselektion, Subtest Funktionen Abstrahieren

<b>Teilungskriterium</b>	<b><math>\chi^2</math> (LRT)</b>	<b>Df</b>	<b><math>\chi^2_{\text{krit}} (\alpha=0,01)</math></b>	<b>P</b>
<b>Score</b>	59.403	47	72.44	0.106
<b>Geschlecht</b>	72.198	48	73.68	0.014
<b>Alter</b>	62.263	47	72.44	0.067

Nachfolgend finden sich die entsprechenden Diagramme der Grafischen Modelltests der *Rasch*-Modell-konformen Version des Subtests „Funktionen Abstrahieren“.





Im internen Teilungskriterium mussten a priori bereits eine beträchtliche Anzahl von Items ausgeschlossen werden: Auf Grund ungünstiger Verteilungen in den Subgruppen wurden im internen Teilungskriterium zusätzlich zu Item 51 die Aufgaben 1, 2, 3, 4, 5, 6, 7, 8, 16, 17, 19, 21, 23, 34, 37, 47, 54 und 55 eliminiert.

Für das Teilungskriterium Geschlecht wurde zusätzlich zu Item 51 das Item 16 ausgeschlossen und für das dritte Teilungskriterium Alter die Testaufgaben 3, 21 und 32.

Bedenkt man also, dass beim internen Teilungskriterium bereits 19 Items nicht in die Berechnung einbezogen werden konnten, ist eine weitere Überprüfung schwierig umzusetzen, da festgelegt wurde, dass nicht mehr als 20 Prozent der Items aus dem Pool entfernt werden sollten.

Dennoch wird auch in diesem Fall nach Analyse und Elimination weiterer auffälliger Items überprüft, ob das dichotom logistische Modell nach Rasch Gültigkeit hat.

Auch anhand der Grafischen Modelltests nach der ersten Parameterschätzung (Abbildungen 8-10) sind starke Modellabweichungen erkennbar.

Die Ergebnisse des z-Test und der Item-Fit-Statistik, sowie die Leichtigkeitsparameter sind im Anhang zu finden (Tabellen d, e, f).

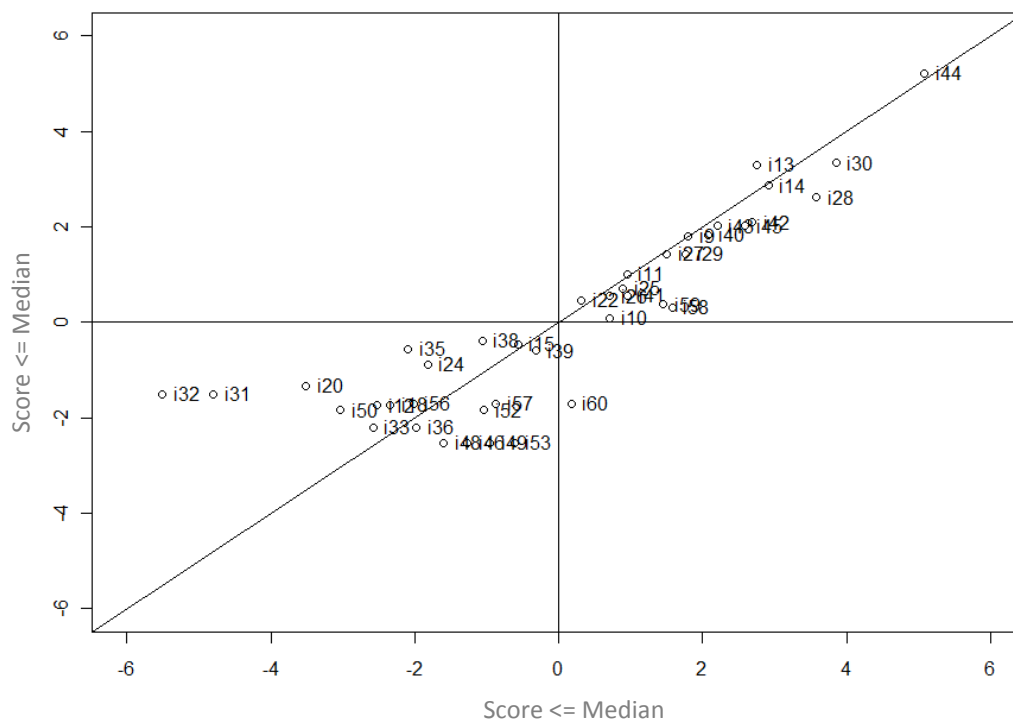


Abbildung 8: Grafischer Modelltest für das interne Teilungskriterium vor Ausschluss eines Items, Subtest „Angewandtes Rechnen“

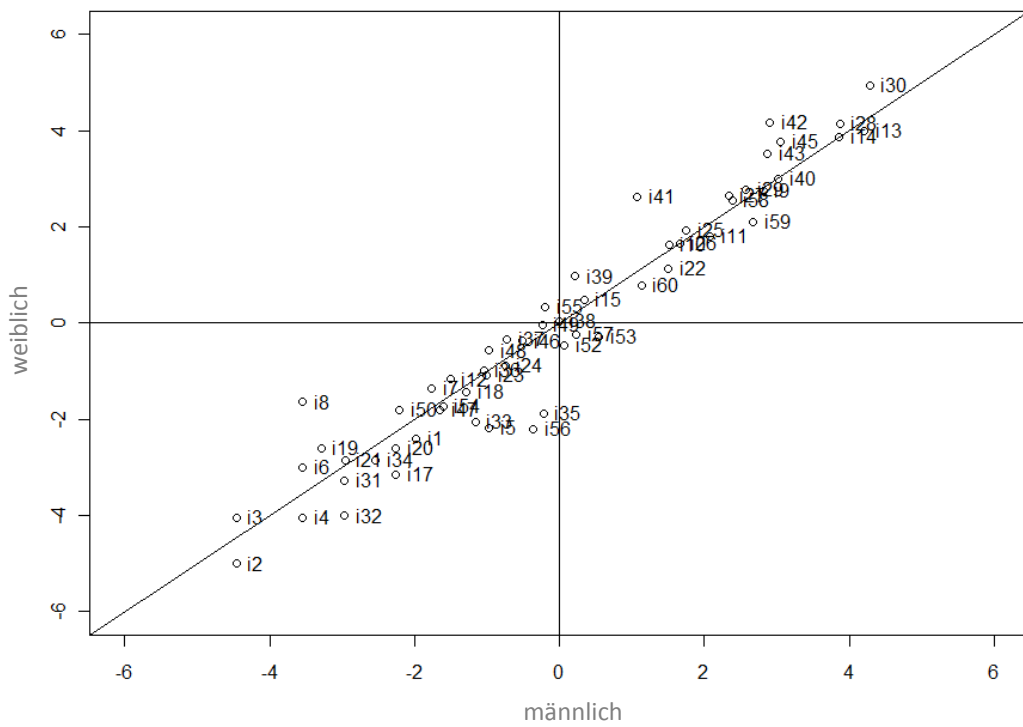


Abbildung 9: Grafischer Modelltest für das Teilkriterium Geschlecht vor Ausschluss eines Items, Subtest „Angewandtes Rechnen“

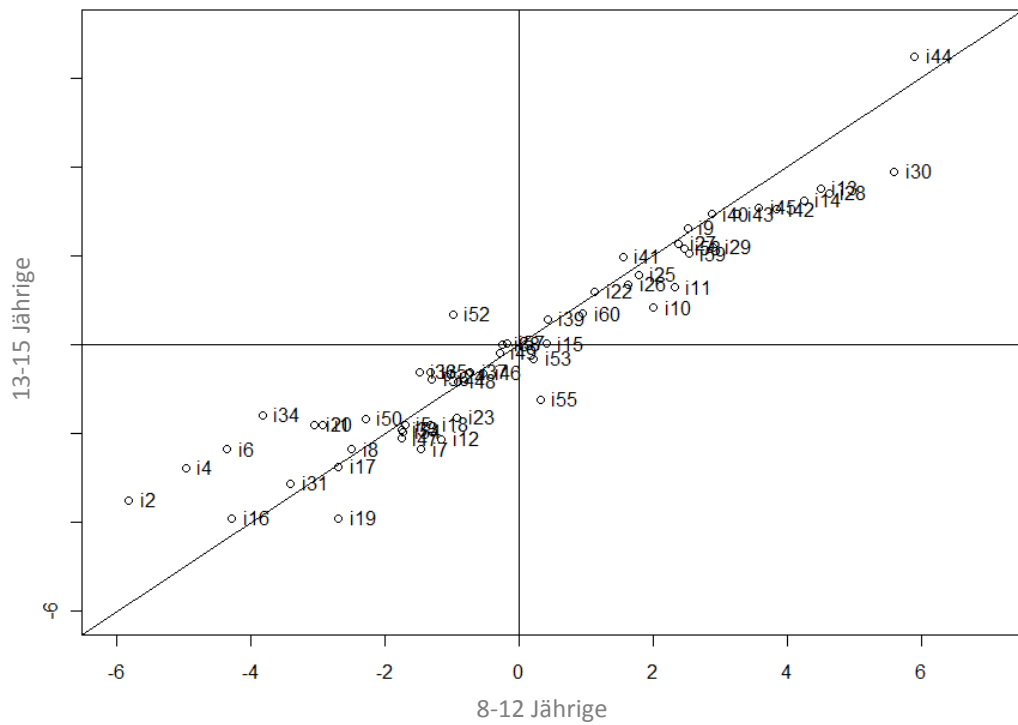


Abbildung 10: Grafischer Modelltest für das Teilkriterium Alter vor Ausschluss eines Items, Subtest „Angewandtes Rechnen“

### 11.2.1 Itemselektion

Anhand der vorliegenden Ergebnisse wurden nun die auffälligen Items bestimmt und sukzessive ausgeschlossen.

Beim internen Splitkriterium wurden im z-Test folgende Items signifikant: 20, 31, 32. Zusätzlich wurden 19 Testaufgaben von vornherein nicht in die Berechnung miteinbezogen (siehe oben).

Beim Teilungskriterium Geschlecht wurden die Items 35, 41 und 42 im z-Test signifikant und unter Berücksichtigung des dritten Teilungskriteriums die Testaufgaben 10, 28, 30 und 52.

Des Weiteren wiesen in der Item-fit-Statistik die Items 2, 4, 14, 17, 20, 25, 32, 36, 39 und 48 signifikante Werte auf.

Die signifikanten Testaufgaben wichen durchwegs auch in den jeweiligen Grafischen Modelltests deutlich von der Geraden ab (siehe Abbildungen 8-10).

Abgesehen von all jenen Items, die a priori auf Grund von ungünstigen Verteilungen in den Subgruppen ausgeschlossen wurden, wurden in der vorliegenden Vorgängerstudie (Eiter, 2011) nur die Testaufgaben 27 (auf Grund eines Konstruktionsfehlers) und 51 (weil es auf Grund des Antwortverhaltens der Testpersonen für keines der Teilungskriterien in die Analyse einbezogen werden konnte) aus dem Itempool entfernt.

Nach Begutachtung der auffälligen Items wurden in dieser Studie zusätzlich die Testaufgaben 2, 4, 16, 32 und 35 aus dem Itempool sukzessive ausgeschlossen.

### 11.2.2 Itemanalyse

Inhaltliche Analyse Item 2:

Eine Hündin bekommt 7 Hundebabys. 4 davon werden an Familien verschenkt. Wie viele Hundebabys bleiben bei der Mutter?
--

(Item übernommen aus Eiter, 2011)

Item 2 wurde im Splitkriterium Geschlecht ( $p=0.002$ ) und in der Item-Fit-Statistik ( $p=0.001$ ) signifikant. Entsprechend der Leichtigkeitsparameter ist davon auszugehen, dass die Testaufgabe den männlichen Testpersonen schwerer gefallen ist (siehe Tabelle f im Anhang).

Grundsätzlich ist in diesem Fall die Verwendung des Begriffes „Hundebabys“ zu kritisieren, da optimalerweise der korrekte Begriff „Welpen“ verwendet werden sollte.

Generell liegt die Hypothese nahe, dass Mädchen dieses Thema emotional eher betrifft, als Burschen und sie eher Erfahrungen mit und Interesse an Welpen haben und daher in der Beantwortung gewissenhafter vorgehen.



Des Weiteren ist zu bemerken, dass das Item nur von 1,7 Prozent der Testpersonen nicht gelöst wurde und seine Aussagekraft daher eingeschränkt ist.

#### Inhaltliche Analyse Item 4:

Peter bekommt 23 Gummibärchen geschenkt. 3 davon isst er gleich auf. 10 verschenkt er an seine Schwester. Wie viele Gummibärchen hat Peter dann noch?

(Item übernommen aus Eiter, 2011)

Dieses Item wurde in der Item-Fit-Statistik ( $p=0.005$ ) signifikant und zeigte in den Grafischen Modelltests Abweichungen von der Geraden, vor allem im Splitkriterium Alter. Dementsprechend war das Item für ältere Testpersonen schwieriger zu lösen.

Vor allem die älteren Testpersonen nahmen die leichten Aufgaben nicht allzu ernst und amüsierten sich über die Einfachheit der Items. Mit zunehmender Schwierigkeit wurden sie immer ehrgeiziger und ernsthafter. Dementsprechend liegen hier Flüchtighkeitsfehler sehr nahe und lassen sich auch aus den Antworten der Testpersonen ableiten. Nachdem das Item aber nur von 2,8 Prozent der Testpersonen nicht gelöst wurde, ist der Informationsgewinn ohnehin anzuzweifeln.

#### Inhaltliche Analyse Item 32:

Susi hat 3 Birnen. Sie isst eine Birne.  
Wie viele hat sie noch?

(Item übernommen aus Eiter, 2011)

Item 32 wurde im z-Test ( $p=0.001$ ), im Grafischen Modelltest für das interne Teilungskriterium und in der Item-Fit-Statistik ( $p=0.000$ ) auffällig. Außerdem wurde es bei der Berechnung für das Teilungskriterium Alter, auf Grund ungünstiger Verteilung innerhalb der Subgruppen, aus dem Pool entfernt.

Entsprechend der Leichtigkeitsparameter, scheinen Testpersonen mit höheren Leistungen in diesem Subtest, eher Schwierigkeiten mit der Lösung der Testaufgabe gehabt zu haben. Generell ist das Item sehr einfach zu lösen. Vor allem Personen, denen Rechnen eher leicht fällt, sind daher eventuell versucht ungenau zu lesen und Flüchtighkeitsfehler zu begehen.

Eventuell ist die Testaufgabe auch etwas unglücklich formuliert; Susi hat auch nach dem Verzehr der Birne immer noch drei Birnen, auch wenn sich eine davon bereits in ihrem Magen befindet. Die Testaufgabe wurde in der vorhergehenden Datenerhebung (Eiter, 2011) vor allem von älteren (8. Schulstufe) und leistungsfähigeren Testpersonen nicht gelöst. Gerade in diesem Alter kann es sein, dass die Antwort eher als „Scherz“ zu verstehen war, da die Ernsthaftigkeit der älteren Teilnehmer bei den leichteren Aufgaben teilweise angezweifelt werden könnte.

Abgesehen davon ob diese Hypothese korrekt ist, zeigt sich, dass es vor allem dann, wenn es für die Testpersonen um nichts geht und die Testung keine Konsequenzen hat, besonders wichtig ist, konkret und eindeutig zu formulieren um zu verhindern, dass „Scherzantworten“ gegeben werden.

Inhaltliche Analyse Item 35:

Ein Mann hat einen neuen Teich in seinen Garten angelegt. Er hat acht Goldfische gekauft und gibt diese in den Teich. Von seiner Nachbarin erhält er drei weitere geschenkt und gibt diese ebenfalls in den Teich.

Wie viele Goldfische sind jetzt im Teich?

(Item übernommen aus Eiter, 2011)

Dieses Item wurde im z-Test ( $p=0.003$ ) und in der Item-Fit-Statistik ( $p=0.000$ ) signifikant und weicht in den Grafischen Modelltests deutlich von der Geraden ab.

Die Testaufgabe fiel den Testpersonen mit höheren Leistungen in dem Subtest und den Burschen schwerer.

Auffallend ist, dass die Testaufgabe im Vergleich zu anderen Aufgaben dieser Leichtigkeitsstufe länger ausformuliert ist und daher ein höheres Augenmerk beim sinnerfassenden Lesen, als beim Rechnen liegt.

So zeichneten sich die typischen (falschen) Antworten der Burschen als klassische Konzentrationsfehler aus; beispielsweise lautete die Antwort 4 (also wurde überlesen, dass schon 8 Fische im Teich waren), oder 12 (hier wurde ein Flüchtigkeitsrechenfehler begangen).

Inhaltliche Analyse Item 51:

An einem kleinen Apfelbaum hängen neun Äpfel. Bei einem Sturm fallen zwei hinunter.  
Wie viele Äpfel hängen nun noch daran?

(Item übernommen aus Eiter, 2011)

Dieses Item wurde in allen drei Splitkriterien auf Grund des Antwortverhaltens der Testpersonen a priori ausgeschieden. In der anschließenden Analyse stellte sich heraus, dass die Testaufgabe insgesamt nur von einer Person nicht gelöst werden konnte und daher kaum Information liefert. Dementsprechend kann auf eine weitere Verwendung verzichtet werden.

Anmerkung zu Item 27:

Ein Haus wird gebaut und soll mehrere Stöcke haben. Die Zimmerhöhe muss 2 Meter betragen, und die Decke muss 75cm dick sein. Wie viele Stöcke kann man maximal bauen, wenn die Höhe des Hauses ohne Dach maximal 8 Meter betragen darf?

(Item übernommen aus Eiter, 2011)

Wie schon bei Eiter (2011) angemerkt sollte dieses Item mit Vorsicht behandelt werden, beziehungsweise vor der weiteren Verwendung in der Formulierung konkretisiert werden. Je nachdem ob die Decke, bzw. das Dach (75cm) in die Gesamthöhe (8m) miteinberechnet werden sollte lautet die Antwort 2 oder 3 Geschoße.

In der vorliegenden Studie wurde den Schüler/-innen gesagt, dass sie die mögliche Anzahl der Stockwerke exklusive Dach berechnen sollen, also lautete die richtige Antwort 3 Stockwerke. Das Item zeigte sich in keinem der angewendeten Verfahren auffällig, also wurde es vorerst im Itempool belassen.

### 11.2.3 Modellschätzung nach Itemselektion

Für den Subtest „Angewandtes Rechnen“ wurde versucht eine möglichst geringe Anzahl an Items aus dem Itempool zu entfernen, da von vornhinein eine beträchtliche Menge auf Grund ungünstiger Antwortverteilungen in den Subgruppen ausgeschieden werden musste.

Insgesamt wurden also nur 5 Items entfernt, was bei einem Pool von 60 Testaufgaben rund 8 Prozent der Testaufgaben entspricht. Nichts desto trotz ist die Anzahl der Testaufgaben, die im internen Teilungskriterium in die Analyse eingehen, bereits von 60 auf 37 verringert.

Tabelle 6 gibt einen Überblick über die Ergebnisse des Likelihood-Quotienten-Tests nach der Selektion der auffälligen Items und veranschaulicht, dass die entsprechende Nullhypothese („Das dichotom logistische Modell von Rasch gilt.“) nicht für alle drei Splitkriterien beibehalten werden kann.

Anhand der Grafischen Modelltests (Abb. 11-13) werden jene Items erkennbar, die sich nach wie vor nicht modellkonform verhalten.

Tabelle 6: Parameterschätzung des Likelihood Ratio Tests (LRT) nach Itemselektion, Subtest „Angewandtes Rechnen“

Teilungskriterium	$\chi^2$ (LRT)	Df	$\chi^2_{krit} (\alpha=0,01)$	p
Score	55.643	38	61.16	0.032
Geschlecht	75.231	52	78.62	0.019
Alter	132.398	55	82.29	0.000



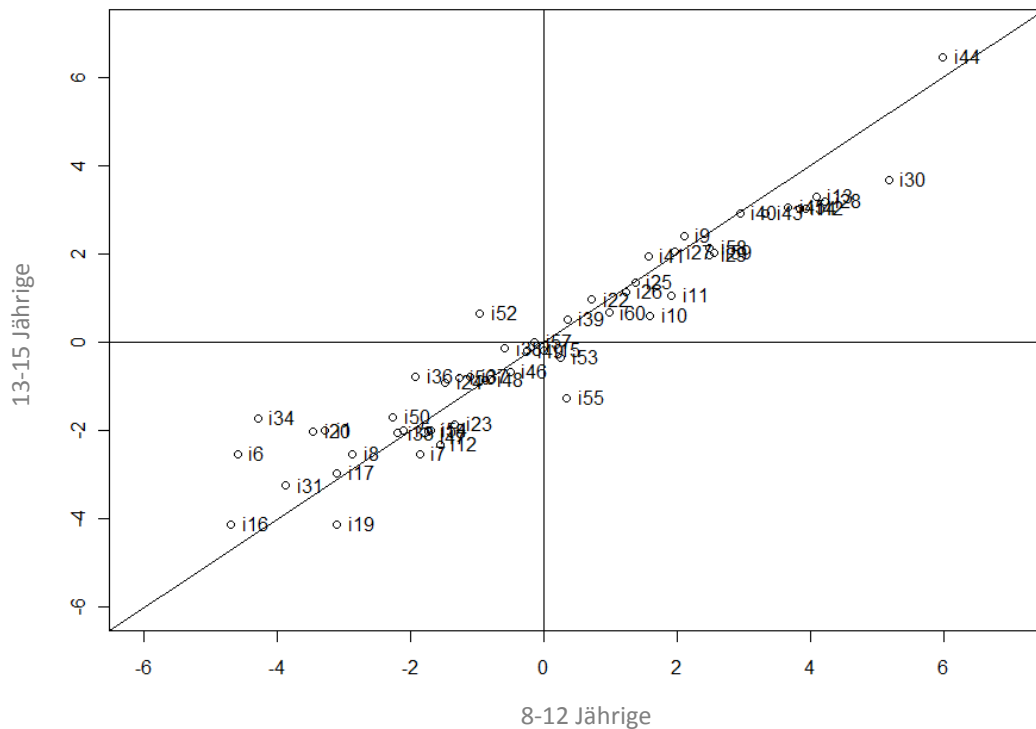


Abbildung 13: Grafischer Modelltest für das Teilungskriterium Alter nach Itemselektion, Subtest „Angewandtes Rechnen“

### 11.3 Wortschatz

Der dritte Subtest der auf *Rasch*-Modell-Konformität überprüft wurde, ist der Wortschatztest. Der Itempool für diesen Subtest umfasste 78 Testaufgaben. In der Vorgängerstudie (Böck, 2010) wurden diese in vier Testformen vorgegeben und in der vorliegenden Erhebung wurden drei überschneidende Parallelförmungen gebildet. Zwei der ursprünglichen Testformen enthielten 20 Items und die beiden anderen jeweils nur 19.

Item 70 war die einzige Testaufgabe, die vor der Modellschätzung auf Grund ungünstiger Verteilungen innerhalb der Subgruppen ausgeschieden werden musste.

Tabelle 7: Erste Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Wortschatz“

Teilungskriterium	$\chi^2$ (LRT)	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	P
Score	191.972	76	107.58	0.000
Geschlecht	127.257	77	108.77	0.000
Alter	195.294	77	108.77	0.000

Die Grafischen Modelltests nach der ersten Parameterschätzung (Abbildungen 14 bis 16) zeigen die Schätzung der einzelnen Items. Die Ergebnisse der Item-Fit-Statistik und des z-Test, sowie die Leichtigkeitsparameter finden sich im Anhang (Tabellen g, h und i).

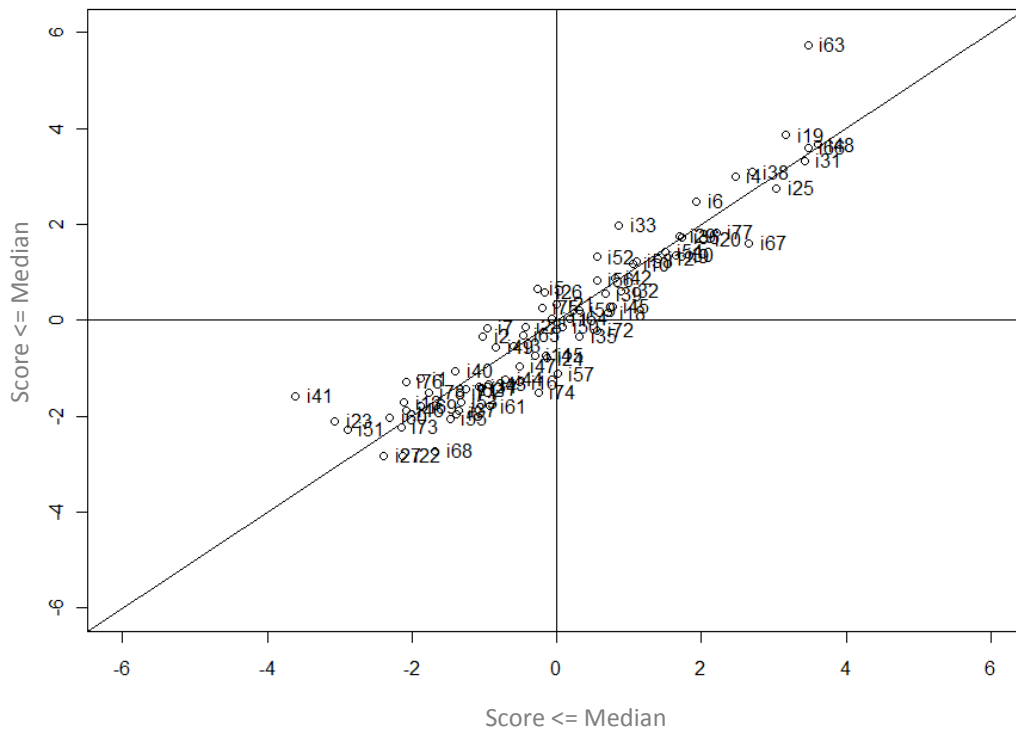


Abbildung 14: Grafischer Modelltest für das interne Teilungskriterium vor Ausschluss eines Items, Subtest „Wortschatz“

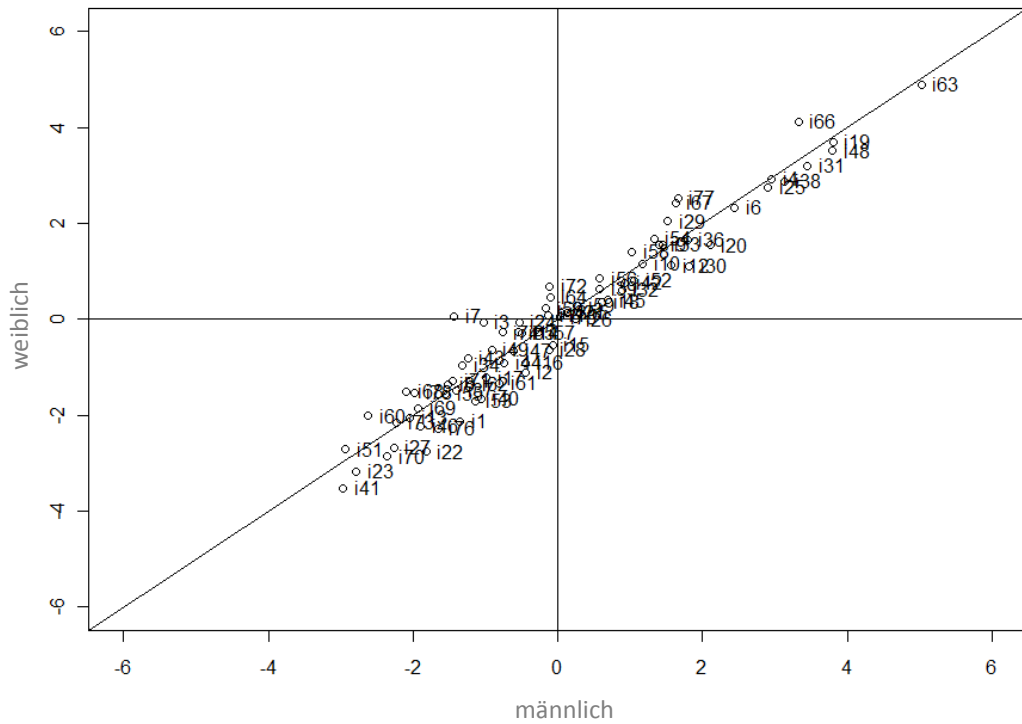


Abbildung 15: Grafischer Modelltest für das Teilungskriterium Geschlecht vor Ausschluss eines Items, Subtest „Wortschatz“

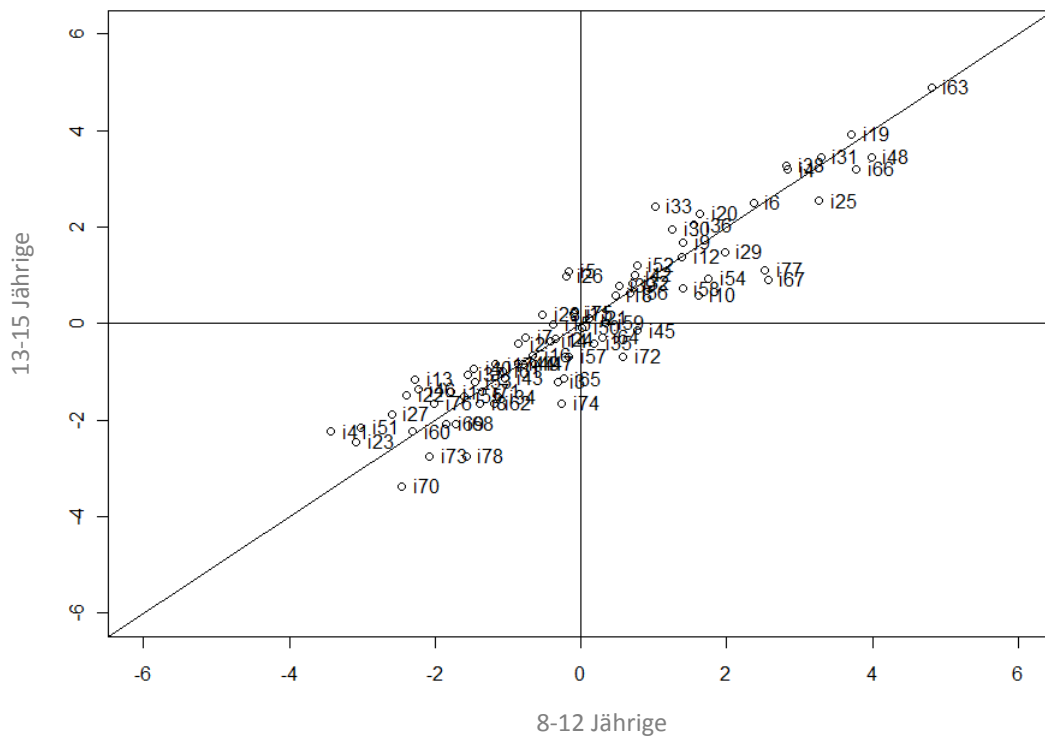


Abbildung 16: Grafischer Modelltest für das Teilungskriterium Alter vor Ausschluss eines Items, Subtest „Wortschatz“

### 11.3.1 Itemselektion

Die auf Grund der Ergebnisse im z-Test, in der Item-Fit-Statistik und im Grafischen Modelltest auffälligen Items wurden inhaltlich analysiert und anschließend sukzessive aus dem Itempool ausgeschlossen.

Beim internen Splitkriterium wurden im z-Test bei einem Signifikanzniveau von 1 Prozent folgende Testaufgaben signifikant: 26, 33, 41, 57, 53 und 67. Im Grafischen Modelltest zeigte zusätzlich Item 68 eine auffällige Abweichung von der Geraden.

Beim Splitkriterium Geschlecht wurde das Item 7 im z-Test signifikant und die Testaufgaben 22 und 67 wichen im Grafischen Modelltest deutlich von der Geraden ab. Für das dritte Teilungskriterium (Alter) ergaben sich im z-Test für die Items 5, 26, 33, 45 und 67 signifikante Werte und die Testaufgaben 70, 72, 74 und 77 wurden im Grafischen Modelltest zusätzlich auffällig.

In der Item-Fit-Statistik wurden die Items 2, 5, 13, 21, 26, 28, 33, 40, 41, 48, 52, 55, 59, 63, 66 und 77 signifikant.

In der Vorgängerstudie (Böck, 2010) wurden die Testaufgaben 5, 7, 41 und 59 aus dem Itempool entfernt.

Nach Analyse der auffälligen Items wurden innerhalb dieser Studie die folgenden Items sukzessive ausgeschlossen: 2, 5, 7, 26, 30, 33, 35, 41, 52, 63, 67, 72, 77.

### 11.3.2 Itemanalyse

Inhaltliche Analyse Item 2:

#### Fragen

Jemanden um (über) eine	<input type="checkbox"/> Ermahnung <input type="checkbox"/> Antwort <input type="checkbox"/> Verzeihung <input type="checkbox"/> Durchsicht	über/für/auf etwas	<input type="checkbox"/> berichten. <input type="checkbox"/> zwingen. <input type="checkbox"/> bitten. <input type="checkbox"/> belügen.
-------------------------	--	--------------------	---

(Item gestaltet nach Vorlage von Böck, 2010)

Item 2 wurde im Grafischen Modelltest und in der Item-Fit-Statistik ( $p=0.000$ ) signifikant. Anhand der Leichtigkeitparameter (Tabelle i) ist erkennbar, dass die Testaufgabe den Testpersonen mit höherer Leistung in diesem Subtest und den Burschen schwerer gefallen ist.



Grundsätzlich fällt auf, dass diese Testaufgabe durch die Verwendung der verschiedenen Präpositionen recht missverständlich und anspruchsvoll formuliert ist. Es ist fraglich ob der ganze Satz und dadurch auch die Distraktoren nicht weniger komplex formuliert werden könnten. Eine mögliche Erklärung wäre, dass die Testaufgabe schwieriger zu lösen ist, wenn man den ganzen Satz liest (wie es Testpersonen mit besseren Leistungen möglicherweise getan haben), als wenn man nur die passenden Begriffe, assoziativ zum Begriff Fragen, aus den grauen Kästchen wählt.

Inhaltliche Analyse Item 5:

### Artist

Ein Artist ist ein	<input type="checkbox"/> Adelige <input type="checkbox"/> Philosoph <input type="checkbox"/> Schausteller <input type="checkbox"/> Schauspieler	der zumeist	<input type="checkbox"/> in Schulen <input type="checkbox"/> in Universitäten <input type="checkbox"/> im Zirkus <input type="checkbox"/> in Restaurants	auftritt.
--------------------	--	-------------	---	-----------

(Item gestaltet nach Vorlage von Böck, 2010)

Die Testaufgabe wies im z-Test im Splitkriterium Alter ( $p=0.008$ ) und in der Item-Fit-Statistik ( $p=0.000$ ) signifikante Werte auf. Das Item war für ältere Testpersonen verhältnismäßig schwieriger.

In der Vorgängerstudie (Böck, 2010) wurde diese Testaufgabe ausgeschlossen, da der Begriff Schausteller veraltet sei und von den Testpersonen nicht mehr verstanden würde.

Die Differenzierung zwischen Schauspieler und Schausteller fällt in jedem Fall schwer. Vor allem, wenn man bedenkt, dass modere (Zirkus-)Shows schauspielerische und artistische Elemente enthalten. Abgesehen davon ist das Item auch in der Hinsicht nicht eindeutig, als da Restaurants in denen Artisten auftreten immer populärer werden.

Inhaltliche Analyse Item 7:

### Helikopter

Ein Helikopter ist	<input type="checkbox"/> ein Ballon, der <input type="checkbox"/> ein Einsatzwagen, der <input type="checkbox"/> ein Unterwasserboot, das <input type="checkbox"/> ein Luftfahrzeug, das	durch eine/n motorisch betriebene/n	<input type="checkbox"/> Reifenachse <input type="checkbox"/> Rotor <input type="checkbox"/> Heliumspender <input type="checkbox"/> Schraube	angetrieben wird.
--------------------	---	-------------------------------------	---	-------------------

(Item gestaltet nach Vorlage von Böck, 2010)

Dieses Item wurde im z-Test im Splitkriterium Geschlecht ( $p=0.001$ ) signifikant und anhand der Leichtigkeitsparameter lässt sich sagen, dass die Testaufgabe für die Mädchen deutlich schwieriger zu lösen war.

Geschlechterstereotyp betrachtet passt auch der Ausschlussgrund der Vorgängerstudie (Böck, 2010) dazu. Dementsprechend wird mit diesem Item nicht nur der Wortschatz getestet, sondern auch das Fachwissen im technischen Bereich. So wussten manche Testpersonen, dass ein Helikopter ein Luftfahrzeug ist, aber nicht, durch welche technischen Mittel er betrieben wird.

Inhaltliche Analyse Item 26:

### Verlies

Ein Verlies ist ein/-e zumeist	<input type="checkbox"/> möblierte/-r	<input type="checkbox"/> Hütte.
	<input type="checkbox"/> gewärmte/-r	<input type="checkbox"/> Wohnraum.
	<input type="checkbox"/> unterirdische/-r	<input type="checkbox"/> Kerker.
	<input type="checkbox"/> vergessene/-r	<input type="checkbox"/> Gruft.

(Item gestaltet nach Vorlage von Böck, 2010)

Diese Testaufgabe wurde im z-Test in zwei der Splitkriterien (Score:  $p=0.010$ ; Alter:  $p=0.000$ ), sowie in der Item-Fit-Statistik ( $p=0.000$ ) signifikant. Die Testaufgabe fiel den Testpersonen mit besseren Leistungen und den Älteren etwas schwerer.

Betrachtet man die Antworten der Testpersonen, fallen keine Regelmäßigkeiten auf, warum die Testaufgabe für gewisse Testpersonen schwieriger sein könnte. Auffallend ist lediglich, dass die unterschiedlichsten Kombinationen gewählt wurden. Sämtliche Distraktoren außer möblierte/-r, gewärmte/-r und Hütte wurden häufig gewählt.

Es scheint also, als wäre den meisten Testperson ungefähr klar, was ein Verlies ist, aber sie können den Begriff, an Hand dieser Distraktoren, nicht genau definieren.

Inhaltliche Analyse Item 30:

### Emsig

Eine emsige Person ist	<input type="checkbox"/> glücklich	und	<input type="checkbox"/> lustig.
	<input type="checkbox"/> faul		<input type="checkbox"/> arbeitswillig.
	<input type="checkbox"/> fleißig		<input type="checkbox"/> arbeitsunwillig.
	<input type="checkbox"/> aggressiv		<input type="checkbox"/> ausgelassen.

(Item gestaltet nach Vorlage von Böck, 2010)

Item 30 wurde auf Grund der Abweichungen von der Geraden in den Grafischen Modelltests für die Teilungskriterien Median und Geschlecht ausgeschlossen.

Während der Testvorgabe war auffallend, dass viele Testpersonen nicht wussten, was das Wort emsig bedeutet. Das Wort emsig kommt außerdem im modernen Wortschatz kaum mehr vor. Die

Distraktoren arbeitswillig und arbeitsunwillig sind nicht optimal gewählt, da die Silbe „un“ bei ungenauem Lesen leicht übersehen werden kann.

Inhaltliche Analyse Item 33:

**Jahreszeit**

Die Jahreszeit unterteilt das Jahr in bestimmte	<input type="checkbox"/> Jahreskreise <input type="checkbox"/> mehrere Stunden <input type="checkbox"/> versch. Perioden <input type="checkbox"/> gewisse Sonnwenden	und zeichnet sich durch bestimmte	<input type="checkbox"/> wetterunabhängige <input type="checkbox"/> tagesabhängige <input type="checkbox"/> unscheinbare <input type="checkbox"/> klimatische	Eigenschaften aus.
---	---	-----------------------------------	--	--------------------

(Item gestaltet nach Vorlage von Böck, 2010)

Diese Testaufgabe wurde im z-Test für die Teilungskriterien Score ( $p=0.000$ ) und Alter ( $p=0.000$ ) und in der Item-Fit-Statistik ( $p=0.000$ ) signifikant. Das Item war für die Testpersonen mit höheren Leistungen in diesem Subtest und für die Älteren schwieriger.

Erklärbar wäre das dadurch, dass jene Personen mit besserer Leistung, bzw. ältere Testpersonen versuchen komplizierter zu denken, bzw. sich korrekter auszudrücken und daher verschiedene Distraktoren in Betracht ziehen. Sowohl Sonnwenden, als auch Jahreskreise und Perioden passen im weiteren Sinne zu Jahreszeiten.

Inhaltliche Analyse Item 35:

**Journalist**

Ein Journalist ist für	<input type="checkbox"/> die Verbreitung <input type="checkbox"/> das Erfinden <input type="checkbox"/> die Bewertung <input type="checkbox"/> das Unterdrücken	von Informationen durch	<input type="checkbox"/> Führung <input type="checkbox"/> Vorträge <input type="checkbox"/> Massenmedien <input type="checkbox"/> Tagung	zuständig.
------------------------	--	-------------------------	---	------------

(Item gestaltet nach Vorlage von Böck, 2010)

Item 35 wurde wegen der eher uneindeutigen Konstruktion ausgeschlossen. Auch wenn natürlich die ursprüngliche Funktion einen Journalisten in der Verbreitung von Informationen liegt, ist kaum mehr zu leugnen, dass auch das Erfinden, Bewerten und Unterdrücken von Informationen in den Tätigkeitsbereich von Journalisten fallen.

Inhaltliche Analyse Item 41:

**Leopard**

Der Leopard gehört zu der Familie der	<input type="checkbox"/> Edelsteine <input type="checkbox"/> Menschenaffen <input type="checkbox"/> Grünpflanzen <input type="checkbox"/> Katzen	und ist ein	<input type="checkbox"/> Beuteltier <input type="checkbox"/> Raubtier <input type="checkbox"/> Heilmittel <input type="checkbox"/> Schutzsymbol
---------------------------------------	---	-------------	--

(Item gestaltet nach Vorlage von Böck, 2010)

Testaufgabe 41 wurde im z-Test im Teilungskriterium Score ( $p=0.000$ ) und in der Item-Fit-Statistik ( $p=0.000$ ) signifikant. Die Testaufgabe ist daher für Testpersonen mit schlechteren Leistungen in diesem Subtest wesentlich leichter zu lösen. Eine Erklärung dafür nannte Böck (2010). Und zwar handelt es sich bei einem Leopard nicht nur um die bekannte Raubkatze, sondern auch um einen Edelstein aus der Familie der Jaspis. Dieser Stein soll über schützende und heilende Wirkung verfügen. Daher wäre eine andere Antwortkombination denkbar. Das Item ist also nicht eindeutig lösbar. Die Testpersonen, die von dem Edelstein wussten, versuchten also eventuell diese speziellere Erklärung zu finden.

Inhaltliche Analyse Item 52:

### Verleugnen

Verleugnen bedeutet nicht zu einer Person	<input type="checkbox"/> schreiten <input type="checkbox"/> stehen <input type="checkbox"/> kommen <input type="checkbox"/> laufen	obwohl man diese	<input type="checkbox"/> fürchtet. <input type="checkbox"/> kennt. <input type="checkbox"/> achtet. <input type="checkbox"/> liebt.
---	---	------------------	--

(Item gestaltet nach Vorlage von Böck, 2010)

Dieses Item zeigte im z-Test für das Splitkriterium Score ( $p=0.003$ ) und in der Item-Fit-Statistik ( $p=0.000$ ) signifikante Werte.

Im ersten Moment fällt die etwas holprige Formulierung auf, da der Satz eigentlich „Verleugnen bedeutet nicht zu einer Person ZU stehen, obwohl man diese kennt.“ Das Wort „zu“ fehlt allerdings und verursacht dadurch eine gewisse Irreführung und möglicherweise falsche Antworten. Die Distraktoren, die das Satzende bilden sind zudem wenig konkret. Man kann sowohl eine Person verleugnen, die man achtet, liebt, oder lediglich nur kennt.

Inhaltliche Analyse Item 63:

### Protektion

Bedeutet das	<input type="checkbox"/> Bewundern <input type="checkbox"/> Imitieren <input type="checkbox"/> Schützen <input type="checkbox"/> Beschatten	oder	<input type="checkbox"/> Beobachten <input type="checkbox"/> Isolieren <input type="checkbox"/> Begünstigen <input type="checkbox"/> Verehren	einer Person.
--------------	--	------	--	---------------

(Item gestaltet nach Vorlage von Böck, 2010)

Item 63 wurde im z-Test im internen Splitkriterium ( $p=0.003$ ), im grafischen Modelltest ( $p=0.000$ ) und in der Item-Fit-Statistik signifikant. Anhand der Itemparameter wird ersichtlich, dass das Item sehr schwer ist. Anhand der Rohdaten lässt sich nachvollziehen, dass lediglich 2 Prozent der Gesamtgruppe (also 8 von rund 390 Personen) in der Lage waren diese Testaufgabe korrekt zu lösen. Daher ist die Sinnhaftigkeit dieses Items generell anzuzweifeln.

Inhaltliche Analyse Item 67:

### Euphorie

Die Euphorie ist eine	<input type="checkbox"/> bedrückte <input type="checkbox"/> überschwängliche <input type="checkbox"/> nervöse <input type="checkbox"/> ruhige	Gemütsverfassung mit	<input type="checkbox"/> allgemeinen Gedanken. <input type="checkbox"/> allgemeiner Traurigkeit. <input type="checkbox"/> allgemeiner Hektik. <input type="checkbox"/> allgemeinem Hochgefühl.
-----------------------	--	----------------------	---

(Item gestaltet nach Vorlage von Böck, 2010)

Diese Testaufgabe wurde im z-Test internen Splitkriterium ( $p=0.009$ ) und im Teilungskriterium Alter ( $p=0.000$ ) signifikant. Zudem zeigten sich auch im Grafischen Modelltest deutliche Abweichungen von der Geraden.

Allerdings konnte keine plausible Erklärung für diese Modellabweichung gefunden werden. Bei der Testung ist aber aufgefallen, dass für die Testpersonen nicht klar war, wieso am Satzende in jedem Distraktor „allgemein“ vorkommt und es wurde gefragt, wie beispielsweise „allgemeinem Hochgefühl“ gemeint ist. Vielleicht könnte man für die weitere Verwendung also die Distraktoren in ihrer Formulierung verfeinern.

Inhaltliche Analyse Item 72:

### Burn Out

Burn Out bezeichnet eine/n besonders ausgeprägte/n berufliche/n und/oder familiäre/n	<input type="checkbox"/> Verletzung <input type="checkbox"/> Streit <input type="checkbox"/> Veränderung <input type="checkbox"/> Erschöpfung	aufgrund von	<input type="checkbox"/> Unachtsamkeit. <input type="checkbox"/> Gerüchten. <input type="checkbox"/> Überbelastung. <input type="checkbox"/> Missverständnissen.
--	--	--------------	---

(Item gestaltet nach Vorlage von Böck, 2010)

Diese Testaufgabe zeigte in allen drei Splikriterien im Grafischen Modelltest deutliche Abweichungen von der Geraden. Anhand der Leichtigkeitsparameter ist erkennbar, dass sich zwischen den Subgruppen Unterschiede finden lassen (Tabelle i).

Im Zuge der Itemanalyse fiel auf, dass der erste Teil des Satzes durch die Verwendung der verschiedenen Endungen sehr komplex formuliert ist. Auch wenn ein beträchtlicher Anteil der Testpersonen dieses Item lösen konnte, unterscheidet sich der Begriff „Burn Out“ deutlich von den anderen zu definierenden Begriffen. Die Lösung des Items erfordert nicht nur Wortverständnis, sondern auch (medizinisches) Fachwissen.

Inhaltliche Analyse Item 77:

### Neutral

Neutral bedeutet	<input type="checkbox"/> formlos <input type="checkbox"/> unparteiisch <input type="checkbox"/> unökonomisch <input type="checkbox"/> patriotisch	beziehungsweise	<input type="checkbox"/> neu <input type="checkbox"/> ungenau <input type="checkbox"/> ungeladen <input type="checkbox"/> positiv	zu sein.
------------------	--	-----------------	--	----------

(Item gestaltet nach Vorlage von Böck, 2010)

Diese Testaufgabe wurde in der Item-Fit-Statistik ( $p=0.000$ ) signifikant und zeigte im Grafischen Modelltest in allen drei Teilungskriterien Abweichungen von der Geraden.

Hier fällt auf, dass die Distraktoren unparteiisch, unökonomisch und patriotisch an und für sich schon Begriffe sind, die Kinder und Jugendliche herausfordern. Dementsprechend war die Testaufgabe auch für jüngere Kinder und für Kinder mit schlechteren Ergebnissen in diesem Subtest wesentlich schwieriger, da sie bereits mit dem Verständnis dieser Begriffe Schwierigkeiten hatten, wohingegen sie das Wort „neutral“ eventuell erklären hätten können.

### 11.3.3 Modellschätzung nach Itemselektion

Nachdem für diesen Subtest im Vorfeld nur wenige Items ausgeschieden werden mussten und der Untertest insgesamt mehr Items enthielt, als die vorherigen, konnten 13 Testaufgaben eliminiert werden ohne die 20 Prozentrate zu überschreiten.

Bei einem Itempool von 78 Aufgaben wurden demnach, inklusive des a priori ausgeschlossenen Items, 18 Prozent der Testaufgaben eliminiert.

Tabelle 8 veranschaulicht die Ergebnisse der zweiten Parameterschätzung. Nach Ausschluss der Items kann für den Subtest keine *Rasch*-Modell-Konformität angenommen werden. Allerdings zeigen zwei der drei Subtests bei einem Signifikanzniveau von  $\alpha=0.01$  *Rasch*-Modell-Konformität; daher kann für die Teilungskriterien Score und Geschlecht die Nullhypothese beibehalten werden.

Tabelle 8: Parameterschätzung des Likelihood Ratio Tests (LRT) nach Itemselektion, Subtest „Wortschatz“

Teilungskriterium	$\chi^2$ (LRT)	Df	$\chi^2_{\text{krit}} (\alpha=0,01)$	p
Score	91.292	63	92.01	0.011
Geschlecht	90.347	64	93.22	0.017
Alter	110.515	64	93.22	0.000

Anhand der Grafischen Modelltests (Abb. 14 bis 16) wird die Modellkonformität der einzelnen Items ersichtlich

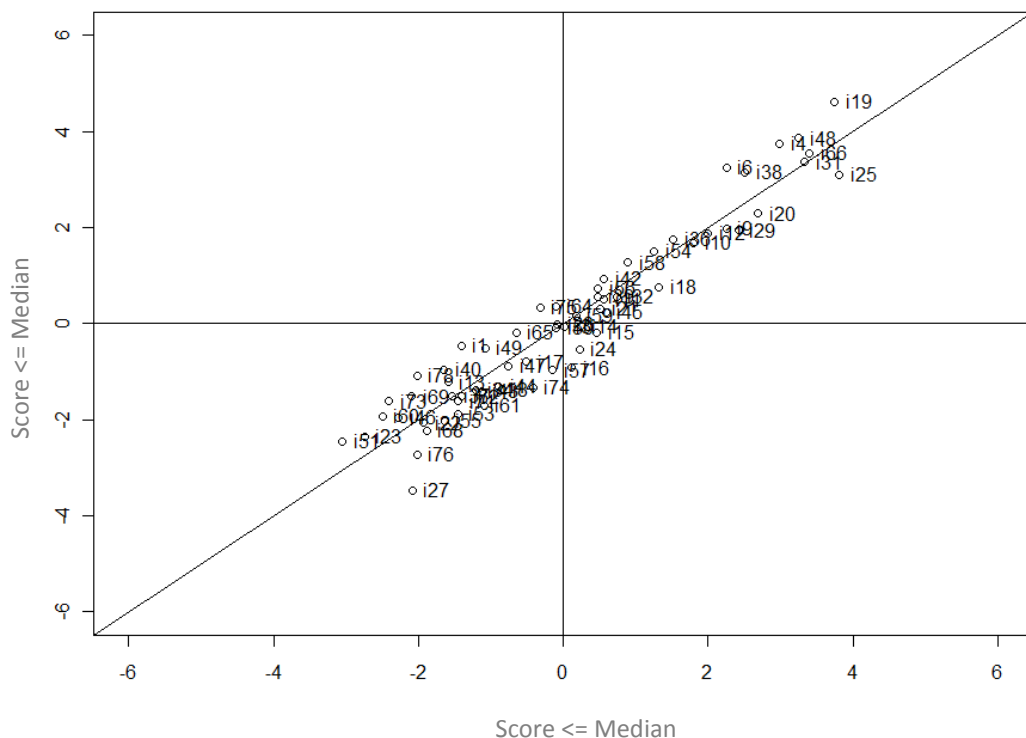


Abbildung 17: Grafischer Modelltest für das Teilungskriterium Score nach Itemselektion, Subtest „Wortschatz“





## 11.4 Alltagswissen

Die vorgegebene Testbatterie wurde durch den Subtest „Alltagswissen“ komplettiert. Das Itempool umfasste pro Testform 15 Testaufgaben, wobei in ein Item schon vor der Analyse ausgeschieden wurde. In der Vorgängerstudie (Eiter, 2011) befand sich ein Item mit Konstruktionsfehler. Das Item wurde in dieser Version zwar berichtigt, allerdings wurde nicht bedacht, dass dadurch die Vergleichbarkeit nicht mehr gegeben ist.

Die erste Parameterschätzung ergab ein Ergebnis, das dazu veranlasst die forschungsleitende Nullhypothese zu verwerfen. In allen drei Teilungskriterien gilt das dichotom logistische Modell nach Rasch nicht.

Tabelle 9: Erste Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Alltagswissen“

Teilungskriterium	$\chi^2$ (LRT)	Df	$\chi^2_{\text{krit}} (\alpha=0,01)$	p
Score	294.122	52	78,62	0.000
Geschlecht	353.894	55	82,29	0.000
Alter	470.438	56	83,51	0.000

Nach der ersten Parameterschätzung wurde Item 57 ausgeschlossen, da es von keiner Testperson korrekt beantwortet wurde. Außerdem wurden im internen Teilungskriterium die Items 2, 17, 20, 28 und 59 auf Grund ungünstiger Antwortverteilungen innerhalb der Subgruppen ausgeschlossen. Beim Teilungskriterium Geschlecht wurden die Testaufgaben 42 und 59 a priori eliminiert und beim dritten Teilungskriterium (Alter) das Item 59. Dementsprechend ging Item 59 in keinem Teilungskriterium in die Analyse ein.

Die Grafischen Modelltests nach der ersten Parameterschätzung (Abbildungen 19 bis 21) bilden die einzelnen Items an Hand der drei Splitkriterien ab.

Die Tabellen für den z-Test, die Item-Fit-Statistik und die Leichtigkeitsparameter befinden sich im Anhang (Tabellen j, k und l).

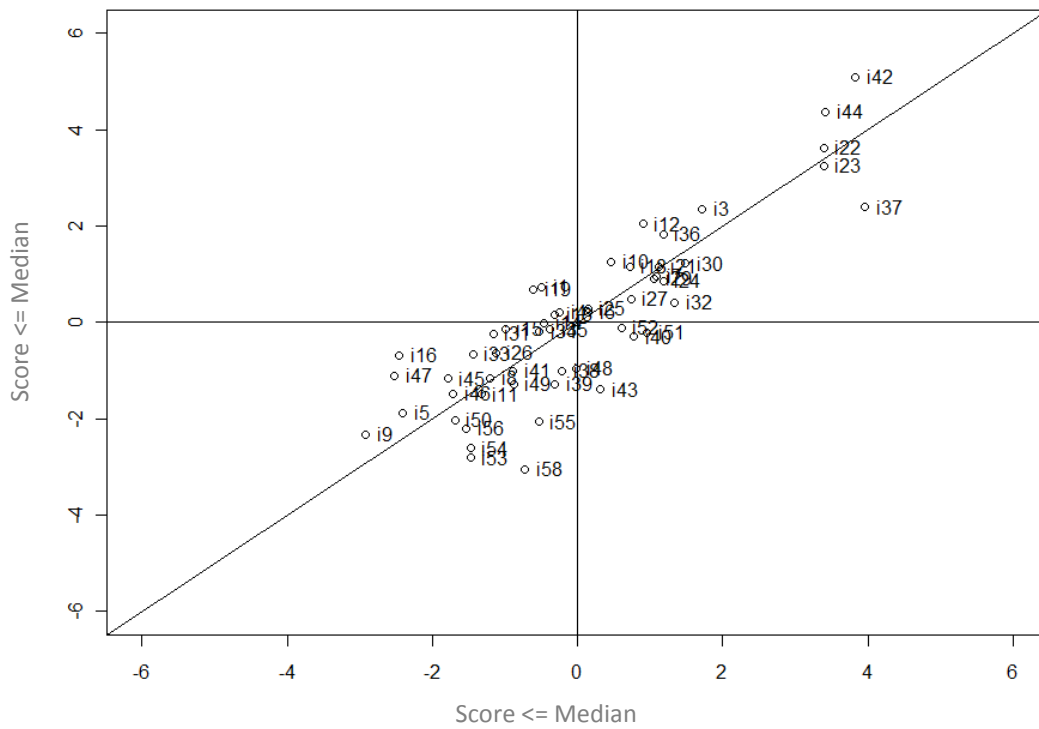


Abbildung 20: Grafischer Modelltest für das Teilungskriterium Score vor Itemselektion, Subtest „Alltagswissen“

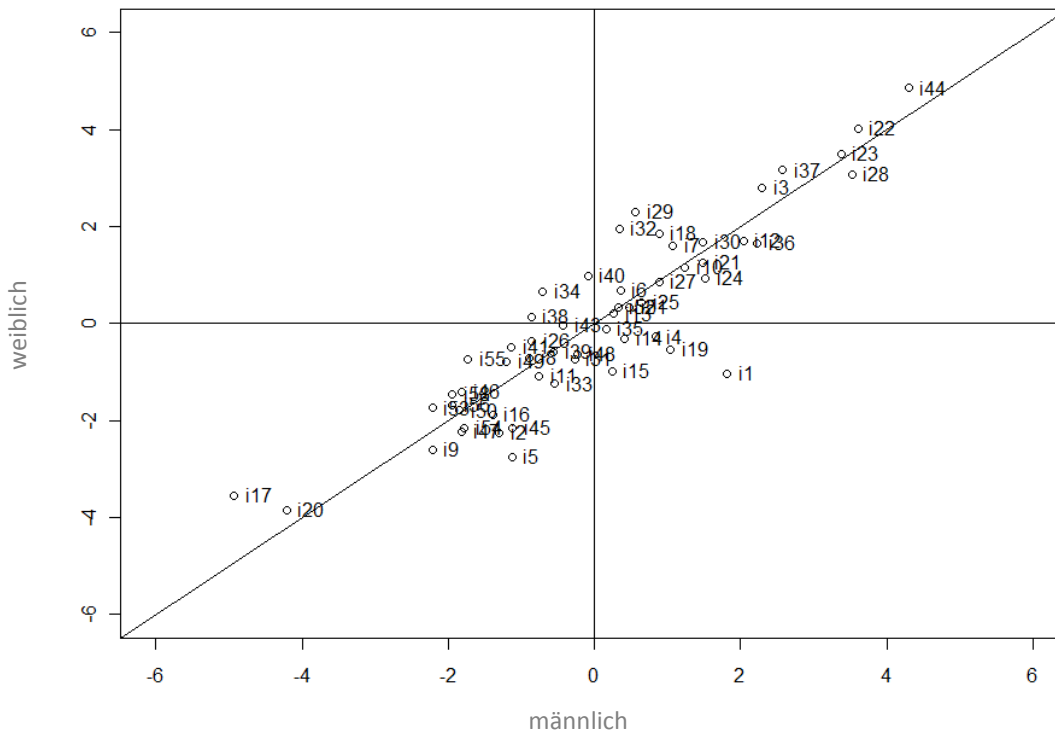


Abbildung 21: Grafischer Modelltest für das Teilungskriterium Geschlecht vor Itemselektion, Subtest „Alltagswissen“

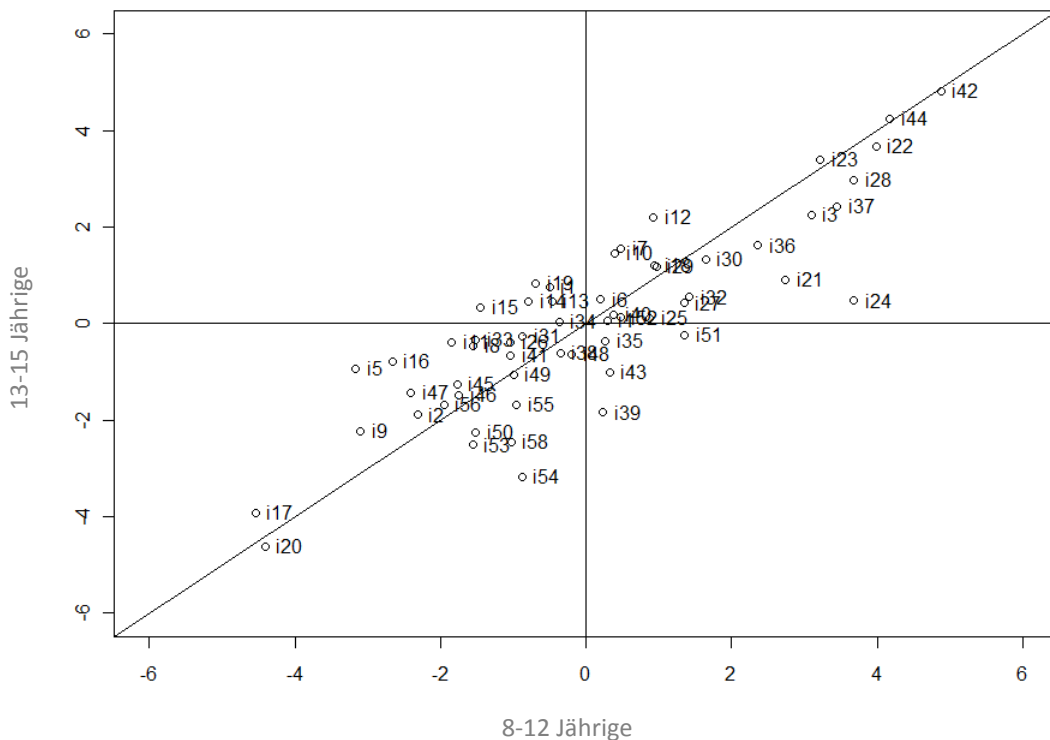


Abbildung 22: Grafischer Modelltest für das Teilungskriterium Alter vor Itemselektion, Subtest „Alltagswissen“

### 11.4.1 Itemselektion

Auf Grund der ersten Parameterschätzung liegt die Vermutung nahe, dass auch durch die Elimination auffälliger Items keine Konformität mit dem dichotom logistischen Modell nach Rasch zu erreichen sein wird. Dennoch wird an dieser Stelle besprochen, welche Testaufgaben sich in den angewendeten Verfahren als auffällig herausstellten.

Im z-Test wurden beim internen Splikriterium die Items 1, 12, 16, 19, 31, 33, 39, 40, 43, 47, 48, 51, 53, 55 und 58 signifikant; beim Teilungskriterium Geschlecht die Testaufgaben 1, 5, 15, 19, 29, 32, 34, 38, 40 und 45. Im letzten Splitkriterium Alter waren die Items 1, 5, 10, 11, 12, 13, 14, 15, 16, 19, 21, 24, 25, 27, 32, 33, 39, 43, 47, 51, 54 und 58 auffällig.

In der Item-Fit-Statistik wurden die Testaufgaben 3, 9, 15, 16, 19, 42, 48 und 49 signifikant.

Die Punkteschwärme in den drei Grafischen Modelltests sind sehr weit gestreut und dementsprechend ist davon auszugehen, dass auch in diesem Untertest viele Items Auffälligkeiten zeigten.

In der Vorgängerstudie (Eiter, 2011) wurden die Testaufgaben 1, 3, 13, 18, 22, 23, 33, 41, 43 und 45 ausgeschlossen.

Auf Grund dieser Ergebnisse und der Daten aus den vorliegenden Verfahren (z-Test, Grafischer Modelltest und Item-Fit-Statistik) wurden die Testaufgaben 1, 15, 16, 19, 32 und 43 ausgeschlossen.

### 11.4.2 Itemanalyse

Inhaltliche Analyse Item 1:

Wie nennt man die Nase bei Pferden?

(Item übernommen aus Eiter, 2011)

Item 1 wurde im z-Test im internen Teilungskriterium ( $p=0.005$ ), im Teilungskriterium Geschlecht ( $p=0.000$ ) und auch im Teilungskriterium Alter ( $p=0.005$ ) signifikant. Zudem zeigten sich in den Grafischen Modelltests (Geschlecht und Alter) deutliche Abweichungen von der Geraden.

An Hand der Leichtigkeitsparameter (Tabelle I) wird deutlich, dass vor allem im Teilungskriterium Geschlecht beträchtliche Unterschiede in der Schwierigkeit des Items vorliegen. Dieser Unterschied lässt sich schlüssig erklären: Geschlechterstereotyp interessieren sich Mädchen mehr für Tiere und vor allem kommen Mädchen eher in Kontakt mit Pferden, da sie öfter als Burschen Reitunterricht nehmen. Dementsprechend kann man davon ausgehen, dass dieses Item nicht unbedingt nur Alltagswissen erfragt, sondern eine zusätzliche geschlechterspezifische Komponente enthält, sowie spezialisiertes Wissen die Lösung deutlich erleichtert. So kann auch erklärt werden warum sich Personen mit schlechteren Leistungen in diesem Untertest leichter mit dem Item getan haben. Personen, die beispielsweise reiten, oder mit einem Pferdehof oder Reitstall zu tun haben, wissen die Antwort eher unabhängig vom Alltagswissen.

Inhaltliche Analyse Item 15:

In welchem Monat beginnt (astronomisch) der Frühling?

(Item übernommen aus Eiter, 2011)

Diese Testaufgabe wurde im z-Test für die Teilungskriterien Geschlecht ( $p=0.000$ ) und Alter ( $P=0.000$ ), sowie in der Item-Fit-Statistik ( $p=0.001$ ) signifikant. Außerdem wurden in allen drei Grafischen Modelltests deutliche Abweichungen von der Geraden ersichtlich.

Auch wenn klar ist, dass der Begriff „astronomisch“ verdeutlichen soll, dass nicht der klimatische, sondern eben der astronomische Frühlingsbeginn gemeint ist, ist an Hand der Reaktionen der Schüler und Schülerinnen davon auszugehen, dass die Verwendung des Begriffs eher verwirrend als verdeutlichend wirkte. So haben manche Testpersonen gefragt, was nun genau gemeint sei. Die Personen, die also wussten, dass es sich beim astronomischen Frühlingsbeginn gleichzeitig um den weitläufig bekannten Frühlingsbeginn (21. März) handelt, hatten in der Beantwortung einen Vorteil.

Inhaltliche Analyse Item 16:

Wie viele Beine haben Spinnentiere?

(Item übernommen aus Eiter, 2011)

Item 16 wurde im internen Teilungskriterium im z-Test ( $p=0.000$ ) und im Grafischen Modelltest, sowie im Teilungskriterium Alter ebenfalls im z-Test ( $p=0.000$ ) und im Grafischen Modelltest auffällig. Außerdem ergab sich in der Item-Fit-Statistik ein signifikanter Wert ( $p=0.000$ ). Auf Grund der Leichtigkeitsparameter zeigt sich, dass die Testaufgabe für die Testpersonen mit besseren Leistungen in diesem Subtest und die älteren Testpersonen schwieriger war.

Im Zuge der Auswertung wurde klar, dass hier viele Testpersonen ganz einfach geraten haben. Die Antworten streuen zwischen 4 und 20 Beinen. Für die deutlichen Modellabweichungen im internen Teilungskriterium konnte keine direkte Erklärung gefunden werden. Für das Splitkriterium Alter kann allerdings nach Begutachtung des Lehrplans für Biologie und Umweltkunde der Hauptschulen und AHS-Unterstufe gesagt werden, dass Verwandtschaftsverhältnisse von Tiergruppen und Waldtiere, beziehungsweise auch Wirbellose im Konkreten, Stoff der 5. und 6. Schulstufe sind. Dadurch würde sich erklären, dass jüngere Kinder sich noch eher an die Antwort erinnern, als ältere.

Inhaltliche Analyse Item 19

Welcher Ton wird gespielt, wenn alle Löcher der (Sopran)Blockflöte zugehalten werden?

(Item übernommen aus Eiter, 2011)

Item 19 würde in allen verfügbaren Verfahren auffällig. Der z-Test wurde in allen drei Splitkriterien signifikant ( $p=0.000$  für alle drei Splitkriterien) und auch in der Item-Fit-Statistik ergab sich ein signifikanter Wert ( $p=0.000$ ). In den Grafischen Modelltests für alle drei Teilungskriterien zeigten sich ebenso auffällige Abweichungen von der Geraden.

Die Leichtigkeitsparameter weisen darauf hin, dass die Testaufgabe für die Personen mit besseren Leistungen in dem Subtest, für Burschen und für ältere Testpersonen schwieriger war.

Wie schon bei Item 1 ausgeführt, handelt es sich bei dieser Frage zwar um Alltagswissen, dennoch werden Personengruppen, die an Musik sehr interessiert, beziehungsweise vorgebildet sind, bevorzugt. Nachdem davon ausgegangen werden kann, dass Mädchen eher Blockflöte spielen, fällt ihnen die Beantwortung der Frage leichter.

#### Inhaltliche Analyse Item 32

Welche Firma hat Bill Gates gegründet?

(Item übernommen aus Eiter, 2011)

Diese Testaufgabe zeigte im Grafischen Modelltest für alle drei Splitkriterien auffällige Abweichungen von der Geraden. Außerdem wurde die Testaufgabe im z-Test für die Teilungskriterien Geschlecht ( $p=0.000$ ) und Alter ( $p=0.007$ ) signifikant.

Auf Grund der Leichtigkeitsparameter lässt sich sagen, dass dieses Item für die Mädchen und die jüngeren Testpersonen deutlich schwerer zu lösen war. Geht man von geschlechtstypischen Interessen aus, ist anzunehmen, dass sich Burschen eher für technische Themen interessieren und daher auch eher wissen, wer Bill Gates ist. Auch, dass die jüngeren Testpersonen eher Schwierigkeiten in der Beantwortung des Items haben, ist nachvollziehbar, da sie sich noch nicht so sehr mit Computern auseinandersetzen und in der Schule zumeist auch noch keinen Informatikunterricht haben.

#### Inhaltliche Analyse Item 45

Ein stärker und länger anhaltendes Steigen der Kurse an der Börse wird auch wie genannt?

(Item übernommen aus Eiter, 2011)

Diese Testaufgabe wurde zwar in vergleichsweise weniger Verfahren signifikant, konnte aber generell nur von 0.67 Prozent der Gesamtgruppe gelöst werden. Daher ist die Sinnhaftigkeit der Testaufgabe zu überdenken; selbst die AHS Lehrerschaft hatte Schwierigkeiten die richtige Antwort zu nennen. Auch die korrekte Beantwortung der Testaufgabe ist nicht ganz eindeutig. So lautet die richtige Antwort im Wirtschaftsjargon „Hausse“, gleichzeitig wären aber auch Begriffe wie Boom, Hochkonjunktur etc. denkbar.

Das Item zeigte zudem im Grafischen Modelltest für die Teilungskriterien Score und Geschlecht auffällige Abweichungen von der Geraden und wurde auch im z-Test für das Splitkriterium Geschlecht signifikant ( $p=0.005$ ).

### 11.4.3 Zweite Modellschätzung

Nach Ausschluss dieser in den Verfahren zur Überprüfung besonders auffälligen und auch inhaltlich nicht einwandfreien Items wurde ein weiteres Mal für alle drei Splitkriterien ein *Likelihood*-Ratio-Test durchgeführt.

Tabelle 10: Zweite Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Alltagswissen“

Teilungskriterium	$\chi^2$ (LRT)	df	$\chi^2_{\text{krit}} (\alpha=0,01)$	p
Score	228.155	46	71.20	0.000
Geschlecht	197.63	48	73.68	0.000
Alter	363.349	49	74.92	0.000

Tabelle 10 zeigt, dass die empirischen  $\chi^2$  Werte nach wie vor sehr hoch sind und *Rasch*-Modell-Konformität weiterhin verneint werden muss.

### 11.4.4 Analyse weiterer inhaltlich auffälliger Items

Nachdem in der Itemanalyse einige weitere Testaufgaben auffielen, die sich eher als inhaltlich, denn als empirisch auffällig zeigten, sollen diese hier auch noch besprochen werden.

Inhaltliche Analyse Item 13

Außer Wachs verbraucht die Flamme einer Kerze noch?

(Item übernommen aus Eiter, 2011)

Testaufgabe 13 wurde bereits in der Vorgängerstudie (Eiter, 2011) aus dem Itempool ausgeschieden. In der vorliegenden Studie zeigte die Testaufgabe im Teilungskriterium Alter Auffälligkeiten. Im z-Test wurde es signifikant ( $p=0.005$ ) und auch im Grafischen Modelltest wich es von der Gerade ab.

Zu kritisieren ist, wie auch schon Eiter (2011) angemerkt hat, dass die Beantwortung der Frage nicht eindeutig ist. Dementsprechend verbraucht eine Kerze drei Komponenten: Wachs, den Docht und Sauerstoff. Entweder man lässt nun alle drei Antworten gelten (was dazu führen würde, dass Antworten auf unterschiedlichen Abstraktionsniveaus gleichsam behandelt werden würden), oder man müsste die Fragestellung konkretisieren.

Inhaltlich Analyse Item 18:

Wo wurde die erste Atombombe abgeworfen?

(Item übernommen aus Eiter, 2011)

Auch diese Aufgabe wurde auf Grund der unkonkreten Formulierung bereits in der Vorgängerstudie (Eiter, 2011) ausgeschieden. In der aktuellen Studie fiel ein signifikanter Wert z-Test des Splitkriteriums Geschlecht ( $p=0.002$ ) und eine Unregelmäßigkeit im Grafischen Modelltest auf. Die Testaufgabe war für die Mädchen schwieriger, was sich dadurch erklären würde, dass (bei geschlechtsstereotyper Betrachtung) Burschen eher an dem Thema interessiert sind und die Frage daher eher beantworten konnten.

Es waren auch zumeist Burschen, die Lösungen wie Wüste oder USA nannten (gemeint waren Atomwaffentests). Weiters waren die Testpersonen unschlüssig, ob sie das Land, oder die Stadt nennen sollten. Bei weiterer Verwendung sollte diese Testaufgabe dahingehend überarbeitet werden, dass nach der Stadt gefragt wird, in der die erste Atombombe im Rahmen eines kriegerischen Akts abgeworfen wurde.

Inhaltliche Analyse Item 31

Wie viele Planeten hat unser Sonnensystem?

(Item übernommen aus Eiter, 2011)

So einfach und klar diese Testaufgabe für ältere Personen scheint, so unklar ist sie für Kinder und Jugendliche. Die astronomische Forschung macht im Eiltempo Fortschritte und die 9-Planeten-Theorie ist längst nicht mehr „State of the Art“. Daher kam es bei dieser Frage zu vielen Zwischenfragen seitens der Testpersonen – mit oder ohne Pluto? Soll Charon (Mond des Pluto) als eigener Planet betrachtet werden?

Das Item zeigte im z-Test für das interne Teilungskriterium einen signifikanten Wert ( $p=0.001$ ) und wich in den Grafischen Modelltests für die Splitkriterien Score und Alter deutlich von der Geraden ab. Es erscheint schlüssig, dass die Testpersonen mit höherer Allgemeinbildung und auch ältere



Testpersonen die Frage differenzierter betrachteten und nicht klassischerweise mit 9 Planeten antworteten.

### 11.4.5 Dritte Modellschätzung

Der Ausschluss der inhaltlich auffälligen Testaufgaben führte zu keiner wesentlichen Verbesserung des Modells. Dennoch sollte man davon Abstand nehmen die besprochenen Items in der Form für weitere Untersuchungen zu verwenden. Daher wird abschließend die dritte Parameterschätzung besprochen.

Tabelle 11: Dritte Parameterschätzung des Likelihood Ratio Tests (LRT), Subtest „Alltagswissen“

Teilungskriterium	$\chi^2$ (LRT)	df	$\chi^2_{krit} (\alpha=0,01)$	P
Score	228.302	43	67.46	0.000
Geschlecht	170.398	45	69.96	0.000
Alter	349.63	46	71.20	0.000

Auf Grund der vorliegenden Ergebnisse kann davon ausgegangen werden, dass auch das weitere Ausschließen von Testaufgaben zu keiner Konformität mit dem dichotomen logistischen Modell nach Rasch führt. Daher muss für diesen Subtest für alle drei Teilungskriterien die forschungsleitende Alternativhypothese, „Das dichotom logistische Modell von Rasch gilt nicht“, angenommen werden.

Die nachfolgenden Grafischen Modelltests nach der dritten Parameterschätzung ermöglichen die itemweise Betrachtung.

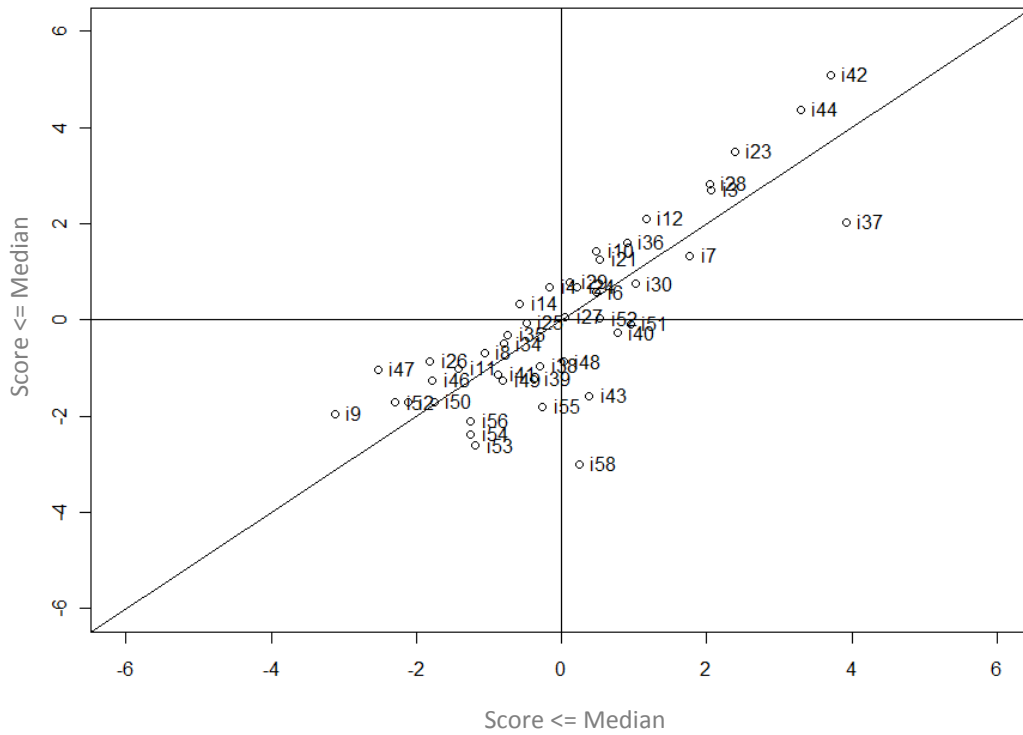


Abbildung 23: Grafischer Modelltest für das interne Teilungskriterium nach Itemselektion, Subtest „Alltagswissen“

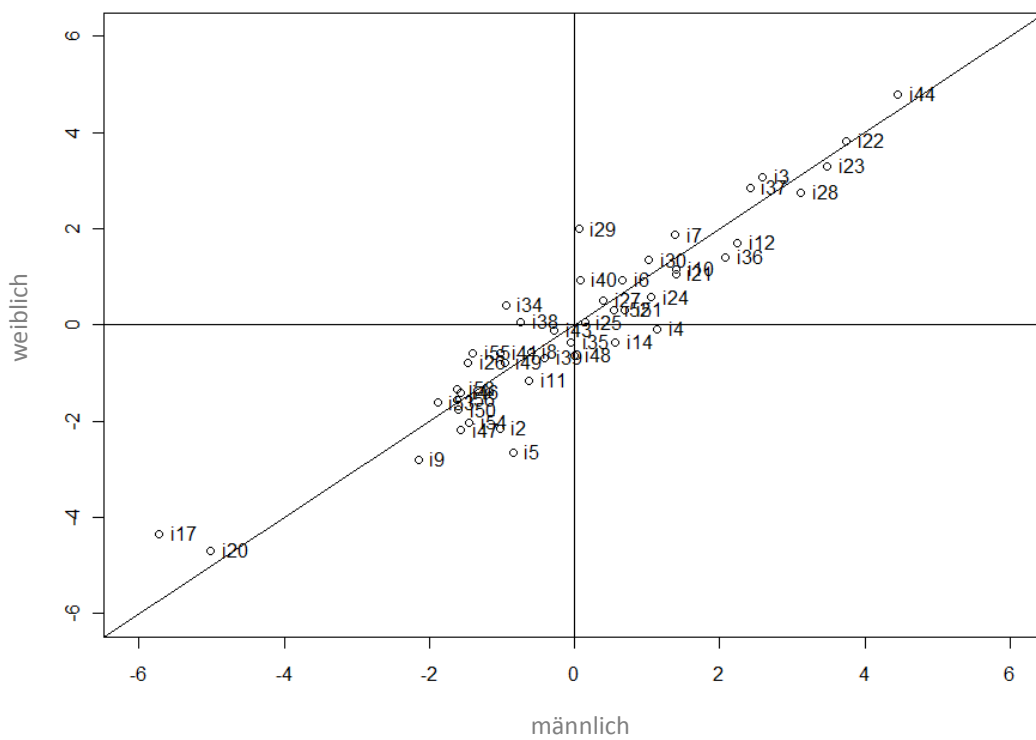


Abbildung 24: Grafischer Modelltest für das Teilungskriterium Geschlecht nach Itemselektion, Subtest „Alltagswissen“

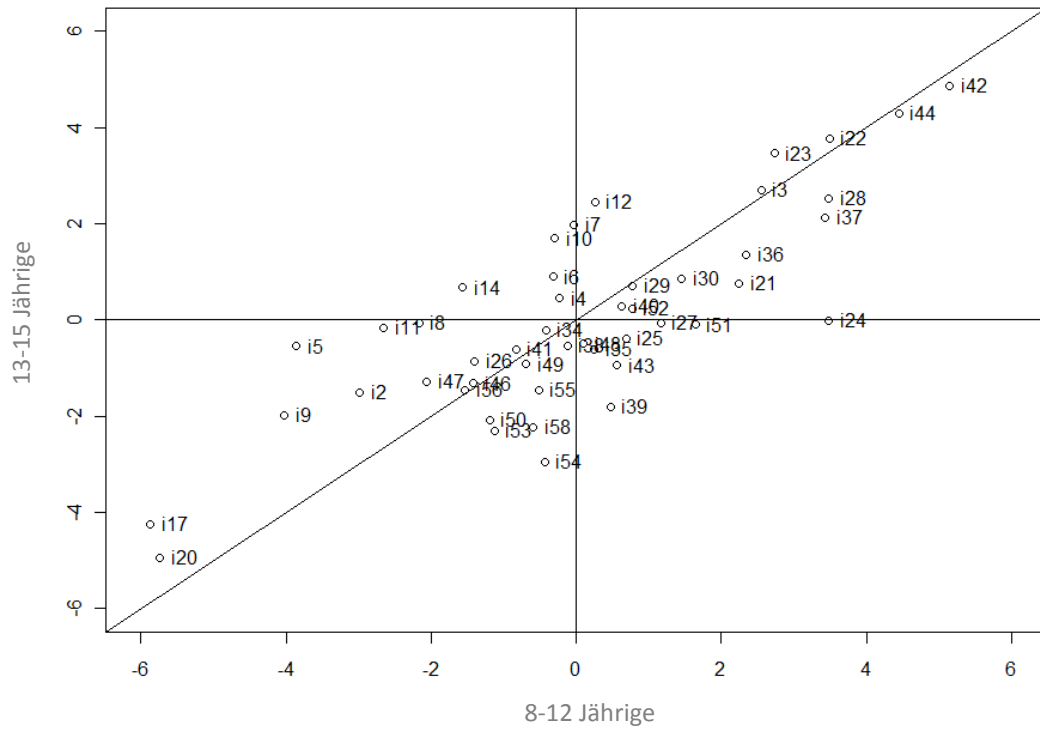


Abbildung 25: Grafischer Modelltest für das Teilungskriterium Alter nach Itemselektion, Subtest „Alltagswissen“

## 12. Diskussion und Ausblick

Generell zeigte sich während der Testung, dass die Vorgabe der neuen AID Aufgaben in der Gruppe problemlos funktioniert. Bei den Testpersonen aller Schulstufe kam es kaum zu Verständnisproblemen und die Testungen verliefen im Allgemeinen reibungslos. Vor allem für die Schüler und Schülerinnen der 3. und 4. Schulstufe, war das überraschend, da durch die Testung bereits hohe Lesekompetenz vorausgesetzt wird.

Mit der geringfügig überarbeiteten Instruktion und der Vorgabe eines Testhefts wurden gute Erfahrungen gemacht. Durch die unterschiedlichen Farben des Testhefts war für die Schüler und Schülerinnen klar, dass die jeweiligen Sitznachbarn unterschiedliche Testformen bearbeiten und abschreiben sinnlos wäre. Andererseits provoziert diese Vorgehensweise nach Testpersonen mit der gleichen Version, also der gleichen Farbe, in der näheren Sitzumgebung zu Kooperationszwecken ausfindig zu machen.

Problematisch war lediglich, dass die Testung für die Vorgängerstudien von Böck (2010) und Eiter (2011) noch nicht allzu lange zurück lag. Da die Erhebung für die aktuelle Studie auf Grund der Vergleichbarkeit der Daten in denselben Schulen durchgeführt wurde, konnten sich manche Schüler/-innen an Testaufgaben erinnern, beziehungsweise nahmen die Testung etwas zu locker, da sie davon ausgingen bereits alle Antworten von den vorhergehenden Testungen zu kennen. An dieser Stelle muss auch angemerkt werden, dass für die Schüler und Schülerinnen eine ganz spezielle Situation entsteht. Sie werden zwar im Schulsetting getestet, gleichzeitig ist aber kein Lehrer/keine Lehrerin anwesend und die Leistung trägt nicht zur Note bei. Diese Tatsache führt dazu, dass die Schüler und Schülerinnen es mit so mancher Antwort nicht zu genau nehmen und sich über einfache Fragen amüsieren.

Zusätzlich ist natürlich zu bedenken, dass die gewählte Stichprobe nur bedingt als repräsentativ angesehen werden kann, da die Testung in einem Privat-Gymnasium in Linz, sowie an zwei Hauptschulen und einer Volksschule im Bezirk Urfahr-Umgebung durchgeführt wurde. Beispielsweise ist an den Daten zur Muttersprache bemerkbar, dass die Zielgruppe nicht repräsentativ für die österreichischen Pflichtschüler/-innen ist. Von rund 670 Kindern gaben lediglich 10 Personen eine andere Muttersprache als Deutsch an. Gerade was das Instruktionsverständnis der Testpersonen betrifft, muss also sicherlich eine weitere Studie in Pflichtschulen im städtischen Umfeld, beispielsweise in Wien, durchgeführt werden. Wie allgemein bekannt, macht der Anteil an Schüler/-innen mit Migrationshintergrund und anderen Muttersprachen als Deutsch, einen großen Anteil der Schülerschaft im städtischen Bereich aus. Um Fairness gegenüber dieser Zielgruppe zu gewährleisten, sind sicher noch Nachfolgestudien nötig. Abgesehen ist das Unterrichtsniveau in Hauptschulen im ländlichen Bereich

bekanntermaßen deutlich höher, als beispielsweise in Wiener Hauptschulen. Es wäre also möglich, dass in Wien erhobene Daten stark von den vorliegenden abweichen würden.

Außerdem ist zu überlegen, ob die beiden Subtests, die in einem offenen Antwortformat zu beantworten waren, nämlich „Alltagswissen“ und „Angewandtes Rechnen“ durch Entwicklung passender Distraktoren in Multiple Choice Format übertragen werden. Während das offene Antwortformat für „Angewandtes Rechnen“ weniger problematisch ist, da die Lösungen von Rechenbeispielen mehr oder weniger eindeutig sind, lässt eine geringfügig unkonkrete Formulierung der Frage im Subtest „Alltagswissen“ eine große Bandbreite an Antworten zu. So kann man auf die Frage „Wo wurde die erste Atombombe abgeworfen“ unterschiedlich antworten – sowohl Hiroshima, als auch Japan, oder ganz abweichende Antworten, wie beispielsweise Wüste oder auch Pazifik (wo erste Atombomben zu Testzwecken gezündet wurden) wären denkbar und genaugenommen korrekt. Daher sollte entweder eine sehr eindeutige Formulierung der Fragen und ein umfassender Katalog an zulässigen Antworten heraus gegeben, oder eben Multiple Choice Antwortformat in Betracht gezogen werden.

Das verlinkte Design machte die vorliegende Überprüfung der kompletten Testformen hinsichtlich des dichotomen logistischen Modells nach Rasch möglich. Im Zuge der Analyse der nicht modellkonformen Items, war auffällig, dass die Betrachtung der falsch gewählten Antworten auf Testpersonen-Ebene für die inhaltliche Analyse bezüglich Eindimensionalität und Konstruktionsfehlern sehr hilfreich war. Für jeweils ein Viertel der Items war das aber nicht möglich, da das Verlinkungsdesign derart gewählt wurde, dass sich aus vier Gruppen drei ergeben (siehe Abbildung XX). Es wäre aber, um die Datenqualität zu erhöhen denkbar gewesen, auch das letzte Viertel der jeweiligen Items zu einer neuen Testform zusammenzustellen.

Betreffend der forschungsleitenden Hypothesen, also der Überprüfung der Gültigkeit des dichotomen logistischen Modells nach Rasch, kann lediglich für einen der vier Untertests, nämlich „Funktionen Abstrahieren“ Modellkonformität angenommen werden. Zwei der Untertests, „Wortschatz“ und „Angewandtes Rechnen“, zeigten nach Ausschluss auffälliger Testaufgaben nur für das Teilungskriterium Alter signifikante Abweichungen. Der vierte Untertest „Alltagswissen“ weicht in allen drei Splitkriterien deutlich vom Modell ab.

Bezüglich des Teilungskriteriums Alter muss aber angemerkt werden, dass Alter in der vorliegenden Arbeit mit „Schulstufe“ gleichgesetzt wurde, da die Annahme naheliegt, dass Schüler und Schülerinnen der gleichen Schulstufe etwa gleich alt sind. Diese Annahme hält aber nicht, wenn man bedenkt, dass manche Schüler und Schülerinnen Klassen überspringen oder auch

wiederholen. Nachdem allerdings in den Vorgängerstudien von Böck (2010) und Eiter (2011) keine Angaben zum Alter enthalten wurden, konnte das Alter, in Jahren und Monaten ausgedrückt, nicht herangezogen werden. Für weitere Studien ist also anzuraten tatsächlich mit dem Alter der Testpersonen zu arbeiten.

Ebenso ist weiterer Forschungsbedarf im Bereich der itemgenerierenden Regeln zu sehen. Hier sollte in einem weiteren Schritt an Hand des LLTM (Linear Logistisches Testmodell, nachzulesen bei Fischer, 1995) die Überprüfung der itemgenerierenden Regeln erfolgen. Vor allem was den Subtests „Alltagswissen“ betrifft sind die itemgenerierenden Regeln zu überdenken, da der Likelihood-Ratio-Test nach Andersen hier deutliche Abweichungen vom *Rasch*-Modell zeigte.

Ziel des laufenden Prozesses könnte folglich die teilautomatische Generierung von Items sein.

Weiterer Forschungsbedarf tut sich im Bereich der Computertestung auf. Wie bereits in Abschnitt 5 ausgeführt, nimmt die Computernutzung in der jugendlichen Zielgruppe stets zu, beziehungsweise ist zu einer Selbstverständlichkeit geworden. Gerade was die adaptive Vorgabe und die Verwendung itemgenerierender Regeln betrifft, birgt die Computernutzung eindeutige Vorteile.

### 13. Zusammenfassung

Diese Diplomarbeit steht in unmittelbarem Zusammenhang mit zwei Vorgängerstudien von Johanna Böck (2010) und Angelika Eiter (2011). Im Rahmen dieser Studien wurde für jeweils zwei Subtests des Adaptiven Intelligenz Diagnostikum 2 (Version 2.2) von Kubinger (2009) nach itemgenerierenden Regeln ein Itempool für eine Gruppenversion des Verfahrens entwickelt. Die Items der neuen Subtests wurden im Wesentlichen den Originalitems des AID 2 nachempfunden, aber eben entsprechend der Gruppenvorgabe verändert.

Böck (2010) entwickelte die Testaufgaben für die Untertests „Funktionen Abstrahieren“ und „Wortschatz“; Eiter (2011) bearbeitete ein Itempool für die Subtests „Angewandtes Rechnen“ und „Alltagswissen“. Die entwickelten Items wurden einer Gruppe von rund 680 Schülern und Schülerinnen in oberösterreichischen Volks- und Hauptschulen, sowie einem Gymnasium vorgegeben. Auf Grund der für die Gruppentestung notwendigen Lesekompetenz der Testpersonen, wurde der Altersbereich auf 8;0 bis 15;11 Jahre angepasst.

Die einzelnen Testformen wurden anschließend auf Konformität mit dem dichotomen logistischen Modell nach Rasch unter Verwendung Andersens Likelihood-Quotienten-Test (1973) geprüft.

Im Zuge der vorliegenden Arbeit wurde eine Reanalyse der Daten von Böck (2010) und Eiter (2011) durchgeführt mit einer Verdopplung der Stichprobe und unter Zuhilfenahme eines verlinkten Testdesigns.

Zu diesem Zweck wurden im Juni 2011 in denselben oberösterreichischen Schulen 667 Schüler und Schülerinnen mit einer leicht überarbeiteten Testversion getestet.

Anschließend wurde durch das verlinkte Design ermöglicht alle erhobenen Daten hinsichtlich der *Rasch*-Modell-Konformität zu überprüfen. So konnte das Itempool jedes Untertests analysiert werden. Als Teilungskriterien für den Likelihood-Quotienten-Test nach Andersen wurde das interne Splikriterium Score und die beiden externen Kriterien Alter (operationalisiert nach der Schulstufe) und Geschlecht herangezogen.

Die Ergebnisse bescheinigen dem Untertest „Funktionen Abstrahieren“ nach Ausschluss von 11 Items auf Grund von Konstruktionsfehlern oder zweifelhafter Eindimensionalität Konformität mit dem *Rasch*-Modell. Für den Untertest „Angewandtes Rechnen“ wurde bereits a priori eine beträchtliche Anzahl von Items auf Grund ungünstiger Verteilungen in den Subgruppen ausgeschlossen. Nach der Analyse wurden lediglich 5 weitere Items aus dem Itempool entfernt. Dennoch konnte keine *Rasch*-Modell-Konformität für alle drei Teilungskriterien angenommen werden, da der LR-Test im Splikriterium Alter nach wie vor signifikant ausfällt.

Der dritte Subtest „Wortschatz“ wurde ebenfalls, auch nach Elimination von 13 Items aus dem Itempool, im LR-Test im Teilungskriterium Alter signifikant. Daher kann nicht von Konformität mit dem dichotomen logistischen Modell nach Rasch ausgegangen werden.

Der vierte und letzte Subtest „Alltagswissen“ zeigte im LR-Test gravierende Abweichungen vom *Rasch*-Modell. Auch nach Ausschluss der inhaltlich auffälligen Testaufgaben, konnte für keines der drei Teilungskriterien *Rasch*-Modell-Konformität bescheinigt werden.

Für Folgestudien liegt eine Überarbeitung der Testaufgaben und eine Reanalyse der itemgenerierenden Regeln, vor allem den Untertests „Alltagswissen“ betreffen, da in diesem Fall die Überprüfung auf *Rasch*-Modell-Konformität negativ ausgefallen ist.

Der Weiteren ist, wie bereits im Ausblick erwähnt, eine Überprüfung der itemgenerierenden Regeln an Hand des LLTM in Betracht zu ziehen.





## 14. Literatur

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

Böck, J. (2010). *AID 2 als Gruppentestung? Eine Machbarkeitsstudie*. Dipl. Arbeit, Universität, Wien.

Bundesgesetzblatt für die Republik Österreich (2000). 1119; 140 II 159. P.b.b. Verlagspostamt

Davies, A.A. & Davies, M. (2007). A Unified Approach to IRT Scale Linking and Scale Transformations. *Methodology*, 3(3), 115–124.

Eiter, A. (2011). *Das AID 2 als Gruppen- oder Computerverfahren. Eine Machbarkeitsstudie*. Dipl. Arbeit, Universität, Wien.

Drosdowski, G. (1989). *Duden Etymologie: Herkunftswörterbuch der deutschen Sprache*. Mannheim: Dudenverlag

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

Fischer, G.H. & Molenaar, I.W. (1995). *Rasch Models. Foundation, Recent Developments, and Applications*. New York: Springer Verlag.

Fischer, G.H. (1995). The Linear Logistic Testmodel. In Fischer, G.H. & Molenaar, I.W. (1995). *Rasch Models. Foundation, Recent Developments, and Applications*. (1. Aufl., S. 131-155). New York: Springer Verlag.

Glas, C.A.W. & Verhelst, N.D. (1995). Testing the Rasch Model. In Fischer, G.H. & Molenaar, I.W. (1995). *Rasch Models. Foundation, Recent Developments, and Applications*. (1. Aufl., S. 69-95). New York: Springer Verlag.

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. & 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.

Kubinger, K. D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum - Version 2.1 (AID 2)*. Göttingen: Beltz.

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

Kubinger, K.D. (2006). *Psychologische Diagnostik –Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.

Kubinger, K.D. & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell. *Diagnostica*, 53(3), 131-143.

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

Kubinger, K.D., Rasch, D. & Yanagida, T. (2011). *Statistik in der Psychologie. Vom Einführungskurs bis zur Dissertation*. Göttingen: Hogrefe

Lampert, T., Sygusch, R.& Schlack, R. (2007). *Nutzung elektronischer Medien im Jugendalter Ergebnisse des Kinder- und Jugendgesundheits surveys (KiGGS)*. Bundesgesundheitsbl-Gesundheitsforsch -Gesundheitsschutz 2007. 50, 643–652. Springer Medizin Verlag.

Poinstingl, H., Mair, P. & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modeling). Anwendung des Rasch-Modells (1-PL Modell)*. Lengerich: Pabst Science Publishers

Wild, B. (1989). Neue Simulationsstudien zur Effizienz verschiedener Parameterschätz-und Itemauswahl-Strategien beim „tailored-testing“. In K.D. Kubinger (Hrsg.), *Moderne Testtheorie - Ein Abriß samt neuesten Beiträgen*(2. Aufl., S. 19-83). Weinheim: Beltz



## 15. Anhang

### 15.1 Ansuchen an den Landesschulrat Oberösterreich

Institut für Entwicklungspsychologie  
und Psychologische Diagnostik  
Universität Wien  
Liebiggasse 5  
1010 Wien

An AR Ulrike Wagner  
**Landesschulrat Oberösterreich**  
Sonnensteinstraße 20  
4040 Linz

Wien, 13. April 2011

#### ***Umsetzung des AID als PC- bzw. Gruppenversion***

Das Adaptive Intelligenz Diagnostikum (kurz AID) ist eine Testbatterie für Kinder und Jugendliche zwischen 6 und 15;11 Jahren, die 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“. Der AID liegt bisher nur in einer Einzeltestversion, aktuell in der überarbeiteten zweiten Version (AID 2.2), vor.

Im Rahmen des aktuellen Forschungsprojekts soll nun ein Schritt in Richtung einer Gruppen- oder PC Version des AID gegangen werden. Durch den adaptiven Ansatz ist der AID 2 ein höchst ökonomisches und zumutbares Verfahren für die Testpersonen und die Testleiter. Für Fragestellungen, die allerdings die Testung größerer Gruppen, wie zum Beispiel ganzer Schulklassen, verlangen, bedeutet die Einzelvorgabe einen sehr hohen Zeit- und Organisationsaufwand. Die Testung in der Gruppe ließe sich auch einfacher in den Schulalltag einbetten, da sie den Schulbetrieb weniger behindert und effizient durchgeführt werden kann. Innerhalb kurzer Zeit wäre es möglich Daten, für eine Vielzahl an Kindern zu erheben und gegebenenfalls frühzeitig Fördermaßnahmen zu setzen.

Durch die Umsetzung in Form einer Gruppenversion ist auch der Schritt in Richtung PC nicht mehr weit. Die Testvorgabe am Computer ist inzwischen „State of the Art“ und bedeutet

zusätzlich zur Ökonomie der Gruppenvorgabe auch eine vereinfachte Vorgabe, raschere und unkompliziertere Auswertung, sowie völlig neue Optionen der Aufgabengestaltung und individuellen Anpassung.

Die ersten notwendigen Schritte um die Frage der Machbarkeit einer oben beschriebenen Version des AID zu beantworten, wurden bereits 2010 von zwei Diplomandinnen der psychologischen Fakultät Universität Wien gesetzt. Nun wird der notwendige nächste Schritt vorbereitet.

Für die Datenerhebung für dieses Forschungsprojekt ersuche ich daher Schulen in Oberösterreich und insbesondere den oberösterreichischen Landesschulrat um Unterstützung bei meinem Projekt.

Die Koordination des Forschungsprojekts erfolgt über die Universität Wien. In den Schulen vor Ort werden während des Testzeitraumes Testleiter/-innen und eine Koordinatorin zur Verfügung stehen. Der Aufwand für die jeweilige Schule soll so gering wie möglich gehalten werden.

Nachdem die Datenerhebungen der zu Grunde liegenden Diplomarbeiten aus dem letzten Jahr an vier oberösterreichischen Schulen durchgeführt wurden, habe ich mich erneut an diese Schulen gewandt. Dieses Vorgehen ist notwendig, um die Vergleichbarkeit der Daten zu gewährleisten. Abweichungen in der Schulwahl beziehungsweise gar im Bundesland würden zu gravierenden regionalen Verzerrungen führen, die die methodisch einwandfreie Umsetzung des Projekts gefährden würden.

Mit der Bitte um Unterstützung und den besten Grüßen,

---

**Petra Hofmayer**

Diplomandin am Institut für Entwicklungspsychologie und Psychologische Diagnostik

**Anlagen:**

Ansprechpartnerin  
Informationen und Organisatorisches zum Forschungsprojekt  
Untersuchungsziel und -ablauf  
Stichprobenszusammensetzung  
Elternbrief

## Ansprechpartner

Das vorliegende Forschungsprojekt wird im Rahmen der Diplomarbeit von Frau Petra Hofmayer umgesetzt.

Petra Hofmayer

Wiedner Hauptstraße 135/16

1050 Wien

✉ a0305157@unet.univie.ac.at

☎ +43 699 128 53 283

Die Diplomarbeitsbetreuung hat Univ.-Prof. Mag. Dr. Klaus Kubinger, Vorstand des Instituts für Entwicklungspsychologie und Psychologische Diagnostik, inne.

Univ.-Prof. Mag. Dr. Klaus Kubinger

Liebiggasse 5

1010 Wien

✉ klaus.kubinger@univie.ac.at

☎ +43 1 4277 478 50

## Informationen und Organisatorisches zum Forschungsprojekt

Ziel der Untersuchung ist es herauszufinden, ob eine Gruppen- oder PC Version des Adaptiven Intelligenzdiagnostikums machbar ist, ohne wesentliche Qualitätseinbußen, beispielsweise bezüglich der Rasch-Modell-Konformität, hinnehmen zu müssen.

Die Vorteile einer Gruppen- bzw. PC Testung liegen vor allem in folgenden Bereichen:

- Effizienzsteigerung
- Erhöhte Flexibilität
- Standardisierung und mehr Objektivität
- Automatisierung

Wie oben bereits erwähnt, wurden im Schuljahr 2009/2010 Voruntersuchungen in Oberösterreich zu diesem Forschungsprojekt im Rahmen zweier Diplomarbeiten durchgeführt.

Das vorgesehene Testinstrument ist der ursprünglichen Version des AID 2 nachempfunden. Die Items sind allerdings nicht ident.

Folgende Untertests kommen zur Anwendung:

- UT 1 Alltagswissen
- UT 3 Angewandtes Rechnen
- UT 6 Synonyme finden
- UT 9 Funktionen abstrahieren

Die Testung wird in einer Papier-Bleistift-Version vorgegeben und ist für Kinder und Jugendliche der 3. bis zur 9. Schulstufe geeignet.

### Eingesetztes Verfahren:

Den Testpersonen werden vier verschiedene Aufgabenstellungen vorgegeben:

#### *1. Alltagswissen*

15 Fragen zum Alltagswissen werden in einem freien Antwortformat vorgegeben und sollen von den Testpersonen beantwortet werden.

Beispiel aus dem Fragebogen:

- Bei welcher Farbe der Fußgängerampel ist das Gehen erlaubt? \_\_\_\_\_
- Von welchem Tier stammt der Hauptanteil der Milch, die verkauft wird? \_\_\_\_\_
- Nenne eine Primärfarbe. \_\_\_\_\_



## 2. *Angewandtes Rechnen*

Hier werden 15 Testvorgaben im freien Antwortformat vorgegeben.

Beispiel aus dem Fragebogen:

- Anne hat sechs Eier in einer Schachtel gekauft. Allerdings stolpert sie und fällt hin. Dabei gehen zwei Eier kaputt. Wie viele ganze Eier hat Anne jetzt noch in der Schachtel?

- Johanna hat zu ihrem Geburtstag ihre drei besten Freundinnen eingeladen. Jede Freundin bringt zwei kleine Geschenke für Johanna mit. Wie viele Geschenke bekommt Johanna somit insgesamt von ihren Freundinnen?

## 3. *Synonyme finden*

Für ein vorgegebenes Wort soll eine passende Definition gefunden werden. Hier werden 20 Aufgaben vorgegeben.

Beispiel aus dem Fragebogen:

Hören bedeutet sich eines

Sinnes

Zeichens

Geräusches

Geruchs

mittels

der Augen

der Ohren

der Nase

des Verstandes

bewusst zu werden.

## 4. *Funktionen abstrahieren*

Vorgabe eines Wortes und fünf Antwortmöglichkeiten. Die Testperson soll angeben, welche zwei der fünf möglichen Antworten eine zentrale Gemeinsamkeit mit dem vorgegebenen Wort teilen. Hier werden 15 Aufgaben vorgegeben.

Beispiel aus dem Fragebogen:

**Ofen**

Rauchfang

Kamin

Holz

Lampe

Heizungen

Die Durchführung der Testung findet im Klassenverband im Rahmen einer Schulstunde statt. Zu Beginn der Testung erfolgt eine Instruktion durch die Testleiterin, die auch während der gesamten Testung anwesend ist. Anschließend sollen die Kinder selbstständig arbeiten. Falls sich Fragen ergeben, ist die Testleiterin natürlich zu Stelle und stellt offene Punkte klar.

Die Eltern werden vorab durch einen Elternbrief verständigt und um ihr Einverständnis gebeten. Zusätzlich werden die Kontaktdaten der Testleiterin weitergegeben um allfällige Fragen zu klären.

### **Untersuchungsablauf**

Bisher wurden mit den in Frage kommenden Schulen die entsprechenden Rahmenbedingungen geklärt.

Die Schulen, die sich bereit erklärt haben, die Untersuchung zu unterstützen, hatten sich dankenswerterweise bereits im letzten Schuljahr dazu bereit erklärt, wodurch das Ausschließen regionaler Verzerrungen möglich ist.

Es handelt sich um folgende Schulen:

- Volksschule Vorderweißenbach
- Hauptschule Vorderweißenbach
- Hauptschule Bad Leonfelden
- Petrinum Linz

Bei der Terminvereinbarung wurde darauf Wert gelegt den Schulablauf so wenig wie möglich zu stören. Daher wurden Termine gewählt, die einerseits in die letzte Schulwoche (nach Notenschluss), sowie andererseits auf den Schultag nach den Pfingstferien fallen.

Als Testleiterin kommen neben der hauptverantwortlichen Diplomandin weitere Diplomanden/-innen aus dem Arbeitsbereich Diagnostik der Fakultät für Psychologie zum Einsatz, die bereits Erfahrung in der Testung von Jugendlichen vorweisen können.

Selbstverständlich werden die gewonnenen Daten im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke verwendet, die TestleiterInnen sind angewiesen auf die Anonymität der Testbögen zu achten und auch die Kinder und Jugendlichen werden darauf hingewiesen keinen Namen zu nennen.

Auf Grund des wissenschaftlichen Charakters des Forschungsprojekts können leider keine persönlichen Rückmeldungen erfolgen, da bislang keine Vergleichswerte für dieses neue Instrument vorliegen.

Die beschriebenen Untertests liegen jeweils in drei unterschiedlichen Formen vor, die an die Schüler/-innen ausgegeben werden. Dies erhöht die Vielfalt an Items, die auf Rasch-Modell-Konformität untersucht werden können. Gleichzeitig wird Abschreiben verhindert.

Mit mehreren Testleiter/-innen werden jeweils einige Klassen parallel getestet um die Zeit in der in den üblichen Schulbetrieb eingegriffen wird möglichst gering zu halten (optimalerweise nur ein Tag). Die gilt vor allem für das Petrinum in Linz, da diese Schule über vergleichsweise viele Klassen verfügt.

**Stichprobenzusammensetzung:**

<i>Schule</i>	<i>Schulstufen</i>	<i>Termin</i>	<i>Anzahl</i>	<i>Bundesland</i>
VS Vorderweißenbach	3., 4.	27.Juni 2011	40	Oberösterreich
HS Vorderweißenbach	5.bis 8.	1.Juli 2011	90	Oberösterreich
HS Bad Leonfelden	5.bis 8.	29., 30.Juni 2011	230	Oberösterreich
AHS Petrinum	5.bis 9.	15.Juni 2011	360	Oberösterreich
<b>INSGESAMT</b>			<b>720</b>	

## 15.2 Elternbrief

### Sehr geehrte Damen und Herren; Liebe Eltern!

Wissenschaftliche Forschung ist immer abhängig von freiwilligen Unterstützer/-innen, die es durch ihren Einsatz ermöglichen, Datenmaterial zu sammeln. Ich wende mich daher mit der Bitte an Sie, mich bei der Umsetzung eines Forschungsprojekts im Rahmen meiner Diplomarbeit zu unterstützen. Schon im letzten Schuljahr haben zwei meiner Kolleginnen eine Diplomarbeitserhebung an dieser Schule durchgeführt. Mein Folgeprojekt ist nun in großem Ausmaß davon abhängig an denselben Schulen Daten erheben zu können um beispielsweise regionale Einflüsse möglichst gering zu halten. Ziel des Projekts ist die Umsetzung des Adaptiven Intelligenzdiagnostikum 2 (kurz AID 2) als Computerversion. Der AID 2 ist ein Intelligenztest für Schulkinder, der häufig in der Schulleistungsdiagnostik eingesetzt wird.

Eine Computerversion des AID 2 würde die Durchführung von Gruppenuntersuchungen in Klassen ermöglichen, sodass Entwicklungspotentiale leichter festgestellt werden können. Die Vorgabe mittels Computer eröffnet zudem neue Möglichkeiten der Aufgabengestaltung.

Um die Machbarkeit einer Computerversion des AID 2 zu überprüfen ist es notwendig durch Erhebungen in verschiedenen Schulen Datenmaterial für die Auswertung zu erhalten.

Die Testung wird pro Klasse (im Klassenverband) eine Schulstunde dauern und beinhaltet Aufgaben wie das Lösen von altersgerechten Rechnungen oder Alltagsfragen. Die Kinder werden während der Testphase von erfahrenen Testleitern/-innen begleitet.

Im Rahmen der Testung und der gesamten Untersuchung ist die Wahrung der Anonymität selbstverständlich!

Mit ihrem Einverständnis zur Teilnahme ihres Kindes an der Datenerhebung leisten Sie einen wichtigen Beitrag für dieses Forschungsprojekt.

Beste Grüße und vielen Dank für ihre Unterstützung!

Petra Hofmayer

(Für Fragen stehe ich gerne unter 0699/12853283 oder [petra.hofmayer@gmail.com](mailto:petra.hofmayer@gmail.com) zur Verfügung)

----- ✂ ----- ✂ ----- ✂ ----- ✂ ----- ✂ ----- ✂ -----

Abschnitt zum Unterzeichnen

Ich erkläre mich einverstanden, dass mein/e Sohn/Tochter \_\_\_\_\_, Klasse \_\_\_\_\_, im Rahmen einer Schuleinheit an der Diplomarbeitstestung bezüglich AID 2 und Computerversion teilnimmt.

\_\_\_\_\_  
Ort, Datum

\_\_\_\_\_  
Unterschrift der/des Erziehungsberechtigten

## 15.3 Tabellen „Funktionen Abstrahieren“

Tab. a: Erste Parameterschätzung, z-Test, Subtest „Funktionen Abstrahieren“, ( $\alpha=0,01$ )

Item	z-Wert	p-Wert	z-Wert	p-Wert	z-Wert	p-Wert
------	--------	--------	--------	--------	--------	--------

	Score	Score	Geschlecht	Geschlecht	Alter	Alter
Item1	--	--	--	--	2.132	0.033
Item2	2.062	0.039	-3.395	0.001	1.462	0.144
Item3	0.297	0.767	-2.090	0.037	0.126	0.900
Item4	-0.741	0.459	-0.245	0.806	-0.719	0.472
Item5	0.021	0.985	-0.527	0.580	-0.367	0.714
Item6	-1.132	0.258	-1.893	0.066	-0.841	0.400
Item7	0.966	0.334	1.332	0.183	0.824	0.410
Item8	-1.481	0.139	0.117	0.907	0.150	0.881
Item9	0.016	0.987	1.087	0.277	-0.769	0.442
Item10	2.027	0.043	1.919	0.055	1.605	0.108
Item11	-0.511	0.610	0.607	0.544	-2.006	0.045
Item12	-1.450	0.147	0.431	0.666	-0.467	0.640
Item13	0.470	0.638	0.503	0.615	-0.881	0.378
Item14	3.191	0.001	-2.410	0.016	2.859	0.004
Item15	2.072	0.001	-0.758	0.448	0.440	0.660
Item16	2.027	0.038	-1.420	0.156	1.064	0.287
Item17	-0.120	0.904	1.748	0.081	3.432	0.001
Item18	4.185	0.000	-0.507	0.569	3.672	0.000
Item19	-1.699	0.089	2.063	0.039	0.906	0.365
Item20	3.326	0.001	1.246	0.213	3.643	0.000
Item21	0.043	0.966	2.527	0.012	0.958	0.338
Item22	-0.166	0.868	1.578	0.115	0.060	0.952
Item23	0.767	0.443	0.866	0.386	-0.446	0.656
Item24	-1.136	0.256	-0.684	0.494	-0.583	0.560
Item25	3.912	0.000	-1.263	0.207	2.734	0.006
Item26	2.317	0.021	-0.016	0.987	0.361	0.718
Item27	-0.289	0.773	-0.800	0.424	-0.245	0.807
Item28	-0.702	0.483	-0.988	0.323	-0.998	0.318
Item29	-2.669	0.008	0.677	0.499	0.050	0.960
Item30	-1.445	0.146	0.527	0.598	-1.540	0.123
Item31	-0.585	0.559	-1.409	0.159	-0.617	0.537
Item32	0.510	0.610	-0.896	0.370	0.801	0.423
Item33	-1.639	0.101	-0.320	0.749	1.658	0.097
Item34	-2.246	0.025	1.311	0.190	0.527	0.598
Item35	0.413	0.680	-1.219	0.223	0.004	0.997
Item36	0.536	0.592	-0.012	0.991	-0.028	0.978
Item37	1.207	0.227	-0.481	0.630	-0.511	0.609
Item38	-0.695	0.487	-0.606	0.544	-1.335	0.182
Item39	1.485	0.137	-1.397	0.162	0.189	0.850
Item40	-0.229	0.819	1.059	0.289	-0.186	0.852
Item41	0.708	0.479	-0.450	0.652	1.212	0.226
Item42	-0.274	0.784	1.280	0.200	-2.705	0.007
Item43	0.781	0.435	0.041	0.967	1.957	0.050
Item44	-0.276	0.783	0.255	0.799	-0.499	0.618
Item45	0.090	0.928	-1.061	0.289	-0.658	0.510
Item46	-1.054	0.292	-0.104	0.917	-0.079	0.937
Item47	-1.021	0.307	-0.047	0.963	-2.026	0.043
Item48	4.980	0.000	-1.834	0.067	2.088	0.037
Item49	0.299	0.765	0.211	0.833	0.223	0.823

Item50	-0.566	0.571	0.372	0.710	-1.148	0.251
Item51	-0.481	0.631	0.501	0.616	-1.400	0.162
Item52	-1.168	0.243	-1.033	0.302	-2.367	0.018
Item53	0.805	0.421	0.189	0.850	-1.115	0.265
Item54	0.596	0.551	-0.463	0.643	-0.754	0.451
Item55	-0.102	0.981	1.697	0.090	-1.776	0.076
Item56	-1.431	0.152	0.233	0.816	-1.973	0.049
Item57	-2.271	0.023	-0.021	0.983	-0.610	0.542
Item58	0.250	0.803	2.842	0.004	-1.389	0.165
Item59	1.092	0.275	1.192	0.233	-1.056	0.291
Item60	1.083	0.279	-0.869	0.385	-0.681	0.496

Tab. b: Itemfit-Statistik, vor Itemselektion, Subtests „Funktionen Abstrahieren“, ( $\alpha=0,01$ )

Item	$\chi^2$ (LRT)	Df	p-Wert	Item	$\chi^2$ (LRT)	Df	p-Wert
Item1	154.576	162	0.649	Item31	173.954	384	1.000
Item2	222.123	162	0.001	Item32	478.960	384	0.001
Item3	137.930	162	0.915	Item33	243.467	384	1.000
Item4	211.595	162	0.005	Item34	282.935	384	1.000
Item5	156.025	162	0.618	Item35	392.530	384	0.371
Item6	132.499	162	0.957	Item36	910.616	384	0.000
Item7	178.892	162	0.172	Item37	401.476	384	0.259
Item8	112.901	162	0.999	Item38	368.859	384	0.702
Item9	337.600	389	0.954	Item39	460.397	379	0.003
Item10	361.746	389	0.776	Item40	332.957	379	0.957
Item11	312.401	389	0.997	Item41	373.548	379	0.569
Item12	247.886	389	1.000	Item42	325.821	379	0.978
Item13	351.257	389	0.876	Item43	436.270	379	0.022
Item14	485.742	389	0.000	Item44	337.112	379	0.940
Item15	403.843	389	0.222	Item45	375.464	379	0.542
Item16	271.559	389	1.000	Item46	235.094	375	1.000
Item17	523.122	389	0.000	Item47	356.440	375	0.747
Item18	412.048	389	0.202	Item48	517.972	375	0.000
Item19	360.752	389	0.845	Item49	345.058	375	0.864
Item20	457.410	389	0.009	Item50	317.427	375	0.986
Item21	342.064	389	0.958	Item51	354.418	375	0.771
Item22	339.203	389	0.967	Item52	336.580	375	0.924
Item23	334.031	389	0.980	Item53	325.372	375	0.970
Item24	362.875	387	0.806	Item54	153.919	161	0.642
Item25	611.879	387	0.000	Item55	99.538	161	1.000
Item26	397.956	387	0.339	Item56	130.022	161	0.965
Item27	386.821	387	0.493	Item57	124.296	161	0.986
Item28	326.389	387	0.989	Item58	160.303	161	0.501
Item29	314.028	387	0.997	Item59	181.426	161	0.129
Item30	292.973	387	1.000	Item60	173.774	161	0.232

Tab. c: Itemparameter (Leichtigkeitsparameter), vor Itemselektion, Subtest „Funktionen abstrahieren“

Item Nr.	Itemparameter		Item-Nr.	Itemparameter		Item-Nr.	Itemparameter	
	Score<= Median	Score>= Median		Männlich	weiblich		8-12 Jahre	13-16 Jahre
1	--	--	1	--	--	1	5.55658745	3.05749359
2	1.1942714	0.2832282	2	0.27440954	1.84499211	2	1.21104964	0.57957736
3	1.2891938	1.1357489	3	0.86759770	1.84492995	3	1.27549809	1.21795239
4	-0.267933	0.0622406	4	-0.06817394	0.03488791	4	-0.24274512	0.06237917
5	0.5518160	0.5422118	5	0.50607704	0.73126718	5	0.48553844	0.64256981
6	0.5518161	1.1357488	6	0.38965544	1.19317959	6	0.54442705	0.91152422
7	1.6403010	1.1357488	7	1.90067465	1.26545184	7	1.68966413	1.30272891
8	1.5878271	3.1399739	8	1.80954759	1.75199122	8	1.76477283	1.69061498
9	2.1071142	2.0986708	9	2.31977406	1.90927230	9	1.98239476	2.29352731
10	0.0910288	-0.523600	10	0.18652754	-0.39019571	10	0.08999342	-0.38642639
11	1.6636883	1.9121772	11	1.81271664	1.60076860	11	1.42839525	2.20883359
12	1.9404043	3.0277439	12	2.11756076	1.95755153	12	1.94610769	2.12907208
13	0.6475692	0.4891358	13	0.72421102	0.56915064	13	0.47053241	0.74654435
14	0.6087928	-0.363045	14	-0.02226969	0.71406036	14	0.67301753	-0.17860389
15	1.3060477	0.5901059	15	1.04891063	1.29494464	15	1.17235137	1.02931633
16	2.7301289	2.8200418	16	2.46160951	3.08946856	16	2.86638096	2.39769703
17	0.4162401	-0.722795	17	0.27143567	-0.21228352	17	0.44905501	-0.48911861
18	0.9710812	0.5456774	18	0.82426587	0.99347907	18	1.34036190	0.26952553
19	-0.0641035	0.4864465	19	0.41509491	-0.15878719	19	0.19628277	-0.05505522
20	-0.3773513	-1.281454	20	-0.56596063	-0.91186181	20	-0.26543793	-1.26495196
21	0.9118698	0.8964130	21	1.29653549	0.54318021	21	0.99828930	0.71300569
22	-0.227366	-0.179243	22	0.06240973	-0.37208579	22	-0.21297200	-0.22957113
23	1.5953987	1.2837276	23	1.68729721	1.40469445	23	1.46211513	1.61497993
24	-0.661682	-0.355012	24	-0.54977383	-0.37238162	24	-0.62799715	-0.47487686
25	2.4171912	0.9937999	25	1.92373453	2.37913505	25	2.50791621	1.53052490
26	-0.488621	-1.079331	26	-0.74601128	-0.74185577	26	-0.77238671	-0.86727557
27	-0.488621	-0.412037	27	-0.50086668	-0.29341439	27	-0.51525455	-0.45118922
28	-1.413304	-1.214208	28	-1.36245835	-1.09624232	28	-1.48661443	-1.20891198
29	-1.2255079	-0.467520	29	-0.67223094	-0.84906211	29	-0.83143406	-0.84453048
30	-0.4412985	-0.039738	30	-0.15643970	-0.29341442	30	-0.51525456	-0.10689318
31	3.2046545	3.8086479	31	2.99019589	3.80941479	31	3.13683430	3.51616285
32	-0.997129	-1.132888	32	1.14768354	-0.91333719	32	-0.99679655	-1.20755122
33	1.699775	2.7014696	33	1.83379363	1.94435080	33	2.04775687	1.47754748
34	0.480521	1.2727801	34	0.90418158	0.53840537	34	0.71619090	0.56884820
35	-2.30075	-2.450994	35	-2.56944600	-2.17519554	35	-2.45345024	-2.45464921
36	-2.726882	-2.956253	36	-2.87210147	-2.86779383	36	-2.95206445	-2.94126712
37	-0.265013	-0.572262	37	-0.42638638	-0.30253734	37	-0.51653940	-0.38363074
38	0.459249	0.6680007	38	0.49928313	0.66547107	38	0.33206605	0.70623815
39	-1.625954	-2.096603	39	-2.08780608	-1.65497089	39	-1.92793946	-1.98724132
40	-1.152904	-1.086120	40	-0.94257352	-1.24532363	40	-1.17104697	-1.11775880
41	-0.818441	-1.017633	41	-0.94257350	-0.81596704	41	-0.77760013	-1.11775883
42	-0.932762	-0.854339	42	-0.69551776	-1.05691701	42	-1.35497262	-0.57152715
43	0.002755	-0.218206	43	-0.04857045	-0.06011824	43	0.14611308	-0.40143155
44	0.417502	0.5037493	44	0.50494425	0.43074123	44	0.32820320	0.47455686
45	-0.658602	-0.684027	45	-0.76236770	-0.46644939	45	-0.80449772	-0.61950818
46	2.324402	3.4147147	46	2.41535428	2.46720976	46	2.32353867	2.36340951
47	-0.887776	-0.568900	47	-0.72513059	-0.71097352	47	-1.12451172	-0.50384177
48	-0.752153	-2.259984	48	-1.74506922	-1.17606415	48	-1.20679939	-1.85675358
49	0.731768	0.6218384	49	0.76050713	0.69079965	49	0.67095823	0.59714074
50	0.41524	0.6218384	50	0.54142240	0.42215635	50	0.22430125	0.59714074



51	-1.237614	-1.088533	51	-1.05167512	-1.20425879	51	-1.43332488	-1.00164087
52	-0.934564	-0.568900	52	-0.88795485	-0.57650166	52	-1.20679940	-0.47948179
53	-2.3324664	-2.623802	53	-2.44174658	-2.50834322	53	-2.78060275	-2.37809861
54	-0.913766	-1.165460	54	-1.07023689	-0.87853864	54	-1.22040028	-0.90648256
55	-4.175939	-4.088255	55	-3.52681416	-5.38817176	55	-5.50609679	-3.55700167
56	-0.698657	0.0305455	56	-0.44038927	-0.53775531	56	-0.99040001	-0.15151557
57	-0.409740	1.3504973	57	-0.13338557	-0.12429281	57	-0.35026905	-0.08757929
58	-2.191957	-2.307208	58	-1.64495739	-3.05837203	58	-2.64282750	-1.99710824
59	-1.048981	-1.503008	59	-0.95765236	-1.45664246	59	-1.51420780	-1.07163626
60	-1.239797	-1.691550	60	-1.58550435	-1.2208451	60	-1.63545024	-1.34804285

## 15.4 Tabellen „Angewandtes Rechnen“

Tab. d: Erste Parameterschätzung, z-Test, Subtest „Angewandtes Rechnen“, ( $\alpha=0,01$ )

Item	z-Wert Score	p-Wert Score	z-Wert Geschlecht	p-Wert Geschlecht	z-Wert Alter	p-Wert Alter
Item1	--	--	-0.524	0.600	1.378	0.168
Item2	--	--	-0.319	0.750	1.463	0.143
Item3	--	--	0.273	0.785	--	--
Item4	--	--	-0.386	0.700	1.759	0.079
Item5	--	--	-1.685	0.092	-0.146	0.884
Item6	--	--	0.479	0.632	1.872	0.061
Item7	--	--	0.559	0.576	-1.124	0.261
Item8	--	--	1.849	0.064	0.160	0.873
Item9	0.008	0.994	-0.105	0.916	0.259	0.796
Item10	-0.902	0.367	0.280	0.779	-2.649	0.008
Item11	0.097	0.923	-0.618	0.536	-2.315	0.021
Item12	0.824	0.410	0.617	0.537	-1.645	0.100
Item13	0.781	0.435	-0.409	0.683	-1.851	0.064
Item14	-0.066	0.947	0.029	0.977	-1.940	0.052
Item15	0.133	0.894	0.318	0.751	-0.852	0.394
Item16	--	--	--	--	0.294	0.768
Item17	--	--	-1.131	0.258	-0.068	0.946
Item18	0.687	0.492	-0.275	0.784	-0.831	0.406
Item19	--	--	0.808	0.419	-1.072	0.284
Item20	2.629	0.009	-0.485	0.628	1.832	0.067
Item21	--	--	0.122	0.903	--	--
Item22	0.266	0.791	-0.997	0.319	0.148	0.882
Item23	--	--	-0.164	0.870	-1.312	0.190
Item24	1.253	0.210	-0.268	0.789	0.682	0.495
Item25	-0.335	0.738	0.485	0.628	-0.573	0.567
Item26	-0.297	0.767	-0.025	0.980	-0.744	0.457
Item27	-0.130	0.897	0.901	0.368	-0.324	0.746
Item28	-1.755	0.079	0.707	0.480	-2.770	0.006
Item29	-0.647	0.517	0.563	0.573	-1.901	0.057
Item30	-0.896	0.370	1.532	0.126	-3.360	0.001
Item31	3.144	0.002	-0.298	0.766	0.232	0.817
Item32	3.190	0.001	-0.823	0.410	--	--
Item33	0.358	0.720	-1.434	0.151	-0.222	0.824
Item34	--	--	-0.343	0.731	2.426	0.015
Item35	2.518	0.012	-2.992	0.003	1.382	0.167
Item36	-0.206	0.837	0.119	0.905	1.671	0.095
Item37	--	--	0.839	0.401	0.217	0.828
Item38	1.226	0.220	0.112	0.911	0.604	0.546
Item39	-0.423	0.672	1.701	0.089	0.310	0.757
Item40	-0.482	0.630	-0.037	0.971	0.136	0.892
Item41	-0.750	0.453	3.674	0.000	1.017	0.309
Item42	-1.210	0.226	2.889	0.004	-1.822	0.068
Item43	-0.375	0.708	1.518	0.129	-0.739	0.460
Item44	0.204	0.838	0.947	0.344	0.954	0.340
Item45	-1.152	0.250	1.690	0.091	-1.165	0.244
Item46	-1.007	0.314	0.239	0.811	-0.158	0.875

Item47	--	--	-0.181	0.856	-0.386	0.699
Item48	-0.745	0.456	0.631	0.528	0.111	0.912
Item49	-1.260	0.208	0.382	0.702	0.197	0.844
Item50	1.081	0.280	0.480	0.631	0.726	0.468
Item51	--	--	--	--	--	--
Item52	-0.765	0.444	-0.937	0.349	3.027	0.002
Item53	-1.532	0.125	-1.520	0.128	-0.921	0.357
Item54	--	--	-0.122	0.903	-0.190	0.849
Item55	--	--	0.740	0.459	-1.685	0.092
Item56	0.260	0.795	-1.885	0.059	0.587	0.557
Item57	-0.723	0.470	-0.663	0.507	0.286	0.775
Item58	-1.322	0.186	0.281	0.779	-0.470	0.638
Item59	-1.128	0.259	-0.903	0.367	-0.788	0.431
Item60	-1.637	0.102	-0.523	0.601	-0.385	0.700

Tab. e: Itemfit-Statistik, vor Itemselektion, Subtest „Angewandtes Rechnen“, ( $\alpha=0,01$ )

Item	$\chi^2$ (LRT)	df	p-Wert	Item	$\chi^2$ (LRT)	df	p-Wert
Item1	167.810	170	0.533	Item31	1432.626	376	0.000
Item2	118.619	170	0.999	Item32	2916.692	376	0.000
Item3	349.772	170	0.000	Item33	171.061	376	1.000
Item4	61.397	170	1.000	Item34	455.249	376	0.003
Item5	112.087	170	1.000	Item35	475.970	376	0.000
Item6	219.351	170	0.006	Item36	332.913	376	0.946
Item7	83.869	170	1.000	Item37	110.283	376	1.000
Item8	84.407	170	1.000	Item38	599.781	376	0.000
Item9	401.531	391	0.346	Item39	292.695	383	1.000
Item10	267.633	390	1.000	Item40	393.905	383	0.339
Item11	386.238	391	0.558	Item41	300.472	383	0.999
Item12	280.075	391	1.000	Item42	255.373	383	1.000
Item13	428.376	391	0.093	Item43	256.545	383	1.000
Item14	265.580	391	1.000	Item44	307.156	383	0.998
Item15	422.900	391	0.128	Item45	272.074	383	1.000
Item16	532.202	388	0.000	Item46	246.463	343	1.000
Item17	105.440	388	1.000	Item47	78.967	343	1.000
Item18	486.791	388	0.000	Item48	299.643	343	0.956
Item19	48.686	388	1.000	Item49	214.287	343	1.000
Item20	683.021	388	0.000	Item50	2490.414	343	0.000
Item21	153.209	388	1.000	Item51	--	--	--
Item22	403.676	388	0.281	Item52	319.217	343	0.817
Item23	192.424	388	1.000	Item53	284.220	343	0.991
Item24	724.500	378	0.000	Item54	37.066	125	1.000
Item25	365.627	378	0.667	Item55	63.929	125	1.000
Item26	303.426	378	0.998	Item56	212.379	125	0.000
Item27	401.657	378	0.193	Item57	121.043	125	0.583
Item28	443.656	378	0.011	Item58	102.178	125	0.933
Item29	282.181	378	1.000	Item59	107.396	125	0.870
Item30	347.488	378	0.868	Item60	88.781	125	0.994

Tab. f: Itemparameter (Leichtigkeitsparameter), vor Itemselektion, Subtest „Angewandtes Rechnen“

Item Nr.	Itemparameter		Item-Nr.	Itemparameter		Item-Nr.	Itemparameter	
	Score<= Median	Score>= Median		männlich	weiblich		8-12 Jahre	13-16 Jahre
1	--	--	1	1.9830511	2.4115344	1	2.9343577	1.8151008
2	--	--	2	4.4620869	5.0042469	2	5.8279719	3.4961791
3	--	--	3	4.4620870	4.0491596	3	--	--
4	--	--	4	3.5465581	4.0491596	4	4.9777866	2.7857687
5	--	--	5	0.9742857	2.1838384	5	1.7086553	1.8150417
6	--	--	6	3.5465581	3.0047588	6	4.3597494	2.3623859
7	--	--	7	1.7617153	1.3606128	7	1.4600668	2.3623859
8	--	--	8	3.5465580	1.6454367	8	2.4995503	2.3623859
9	-1.79816	-1.80342277	9	-2.809299	-2.761839	9	-2.509293	-2.626816
10	-0.70848	-0.08997529	10	-1.509946	-1.632558	10	-2.001199	-0.823044
11	-0.94886	-1.01202804	11	-2.077105	-1.806334	11	-2.315443	-1.284616
12	2.527488	1.74262510	12	1.5088661	1.1702057	12	1.1575144	2.1278163
13	-2.75200	-3.29103137	13	-4.200688	-3.992044	13	-4.490765	-3.506523
14	-2.91238	-2.86638555	14	-3.858095	-3.872657	14	-4.241495	-3.249229
15	0.566283	0.46995134	15	-0.339763	-0.482880	15	-0.402322	-0.015994
16	--	--	16	--	--	16	4.2891183	3.9109867
17	--	--	17	2.2570814	3.1557323	17	2.6991186	2.7530781
18	2.34117	1.73339635	18	1.2872463	1.4359743	18	1.3245850	1.8007282
19	--	--	19	3.2915247	2.6064682	19	2.6991186	3.9109867
20	3.514011	1.32428157	20	2.2570813	2.6064682	20	3.0580092	1.8007282
21	--	--	21	2.9577505	2.8567822	21	--	--
22	-0.31564	-0.46689944	22	-1.500345	-1.127753	22	-1.122728	-1.178131
23	--	--	23	1.0019319	1.0845751	23	0.9338102	1.6449846
24	1.820898	0.87763391	24	0.7492586	0.8831919	24	1.0689725	0.7257699
25	-0.89351	-0.71353868	25	-1.746663	-1.923326	25	-1.774818	-1.565399
26	-0.71227	-0.54974307	26	-1.657655	-1.648393	26	-1.626451	-1.352900
27	-1.50475	-1.43877608	27	-2.332067	-2.661796	27	-2.375435	-2.257003
28	-3.573553	-2.61375242	28	-3.864861	-4.148466	28	-4.617639	-3.406814
29	-1.769220	-1.43876874	29	-2.572670	-2.779526	29	-2.891766	-2.182417
30	-3.847080	-3.33737289	30	-4.279738	-4.946943	30	-5.586962	-3.888840
31	4.7948426	1.50125333	31	2.9708852	3.2845520	31	3.4083759	3.1285791
32	5.5026862	1.50125335	32	2.9708858	4.0105947	32	--	--
33	2.5836831	2.19933965	33	1.1657922	2.0563520	33	1.7601888	1.9225215
34	--	--	34	2.5423137	2.8462828	34	3.8299787	1.5890203
35	2.0970219	0.56959194	35	0.2205602	1.8712590	35	1.3311797	0.6225121
36	1.9809549	2.19933966	36	1.0448134	0.9830475	36	1.4912608	0.6225121
37	--	--	37	0.7264857	0.3350727	37	0.7291967	0.6225121
38	1.0631630	0.38211072	38	0.0105521	-0.035945	38	0.2501604	-0.005508
39	0.3176517	0.59999292	39	-0.214767	-0.973322	39	-0.417151	-0.559270
40	-2.069109	-1.83290959	40	-3.017059	-3.001480	40	-2.86847	-2.926023
41	-0.956005	-0.55659688	41	-1.069278	-2.631068	41	-1.543662	-1.966271
42	-2.690532	-2.09006594	42	-2.900274	-4.167164	42	-3.844487	-3.038348
43	-2.202969	-2.01949296	43	-2.871053	-3.521407	43	-3.243104	-2.926001
44	-5.076291	-5.21999000	44	-6.114262	-6.709033	44	-5.898551	-6.479284
45	-2.589575	-2.01948958	45	-3.046217	-3.777636	45	-3.574130	-3.066234
46	1.2700368	2.53401155	46	0.5138254	0.3729298	46	0.5390036	0.6383387
47	--	--	47	1.6532813	1.7959476	47	1.7627876	2.1190247
48	1.5953100	2.53401155	48	0.9714513	0.5744029	48	0.9064877	0.8330195
49	0.9565487	2.5340115	49	0.2421810	0.0281450	49	0.2913844	0.1773483
50	3.0387469	1.84078965	50	2.2119265	1.7959475	50	2.2963939	1.6641077

51	--	--	51	--	--	51	--	--
52	1.0404912	1.84078964	52	-0.061693	0.4707112	52	0.9907268	-0.681832
53	0.6199863	2.53401157	53	-0.545747	0.2803400	53	-0.218922	0.3148680
54	--	--	54	1.6044541	1.7262839	54	1.7371334	1.9686454
55	--	--	55	0.1938795	-0.331104	55	-0.310571	1.2333606
56	2.0168484	1.70722861	56	0.3599419	2.2045893	56	1.3032951	0.7853325
57	0.8697245	1.70722860	57	-0.228669	0.2450164	57	0.1767358	-0.031582
58	-1.5755732	-0.3076860	58	-2.382924	-2.557530	58	-2.458027	-2.168850
59	-1.4488881	-0.3721558	59	-2.668808	-2.108896	59	-2.526421	-2.039596
60	-0.1726617	1.70722858	60	-1.127210	-0.790507	60	-0.942845	-0.689495

## 15.5 Tabellen „Wortschatz“

Tab. g: Erste Parameterschätzung, z-Test, Subtest „Wortschatz“, ( $\alpha=0,01$ )

Item	z-Wert Score	p-Wert Score	z-Wert Geschlecht	p-Wert Geschlecht	z-Wert Alter	p-Wert Alter
Item1	1.199	0.231	-1.586	0.113	0.593	0.553
Item2	1.489	0.136	-1.456	0.145	0.855	0.393
Item3	0.114	0.909	2.138	0.032	-1.623	0.105
Item4	0.753	0.451	-0.037	0.971	0.538	0.590
Item5	2.191	0.028	-0.413	0.679	2.673	0.008
Item6	0.918	0.358	-0.200	0.841	0.229	0.819
Item7	1.805	0.071	3.276	0.001	0.926	0.355
Item8	-0.852	0.394	0.311	0.756	-0.450	0.652
Item9	-0.736	0.462	0.298	0.766	0.543	0.587
Item10	0.248	0.804	-0.026	0.979	-2.083	0.037
Item11	0.301	0.764	0.674	0.500	0.902	0.367
Item12	-0.358	0.720	-1.444	0.149	-0.025	0.980
Item13	0.772	0.440	-0.037	0.971	2.493	0.013
Item14	-1.180	0.238	0.626	0.531	0.167	0.867
Item15	-1.607	0.108	-1.455	0.146	0.975	0.330
Item16	-1.847	0.065	-1.404	0.160	-0.027	0.978
Item17	-0.848	0.396	-0.698	0.485	0.784	0.433
Item18	-1.703	0.089	-0.820	0.412	0.345	0.730
Item19	1.409	0.159	-0.229	0.819	0.429	0.668
Item20	-0.968	0.333	-1.735	0.083	1.824	0.068
Item21	1.183	0.237	-0.013	0.990	-0.010	0.992
Item22	-0.930	0.352	-2.300	0.021	1.841	0.066
Item23	1.627	0.104	-0.813	0.416	0.938	0.348
Item24	-1.858	0.063	1.547	0.122	0.050	0.960
Item25	-0.689	0.491	-0.455	0.649	-1.927	0.054
Item26	2.590	0.010	-0.835	0.404	3.765	0.000
Item27	-0.574	0.566	-0.976	0.329	1.292	0.196
Item28	0.940	0.347	-1.725	0.085	2.077	0.038
Item29	0.175	0.861	1.686	0.092	-1.494	0.135
Item30	-0.976	0.329	-2.400	0.016	2.245	0.025
Item31	-0.217	0.828	-0.624	0.533	0.404	0.686
Item32	-1.024	0.306	-1.049	0.294	0.439	0.661
Item33	4.120	0.000	0.345	0.730	4.739	0.000
Item34	-0.765	0.445	1.139	0.255	-0.815	0.415
Item35	-2.099	0.036	0.300	0.764	-1.691	0.091
Item36	-0.037	0.970	-0.518	0.604	1.631	0.103
Item37	-1.063	0.288	-0.252	0.801	1.181	0.237
Item38	0.970	0.332	-0.723	0.470	1.238	0.216
Item39	-0.449	0.653	0.187	0.852	0.876	0.381
Item40	0.964	0.335	-1.902	0.057	1.420	0.156
Item41	4.306	0.000	-1.183	0.237	2.056	0.040
Item42	0.272	0.785	-0.647	0.518	0.925	0.355
Item43	-1.054	0.292	1.440	0.150	-0.249	0.803
Item44	-1.433	0.152	-0.604	0.546	0.071	0.943
Item45	-1.718	0.086	-1.109	0.267	-2.862	0.004
Item46	0.406	0.685	-0.936	0.349	2.003	0.045

Item47	-1.301	0.193	-0.210	0.834	-0.572	0.567
Item48	0.098	0.922	-0.628	0.530	-1.140	0.254
Item49	0.882	0.378	0.949	0.342	-0.106	0.915
Item50	-0.645	0.519	1.307	0.191	-0.325	0.745
Item51	1.080	0.280	0.530	0.596	1.752	0.080
Item52	2.608	0.009	-0.861	0.389	1.342	0.180
Item53	-0.921	0.357	-1.867	0.062	0.546	0.585
Item54	-0.145	0.885	1.210	0.226	-2.385	0.017
Item55	-1.121	0.262	0.006	0.996	0.256	0.798
Item56	0.918	0.358	0.850	0.395	-0.188	0.851
Item57	-2.999	0.003	-0.022	0.982	-1.486	0.137
Item58	0.421	0.673	1.227	0.220	-2.105	0.035
Item59	-0.323	0.747	0.091	0.927	-1.002	0.316
Item60	0.535	0.593	1.495	0.135	0.166	0.868
Item61	-1.639	0.101	-1.506	0.132	0.254	0.800
Item62	-0.360	0.719	-0.368	0.713	-0.990	0.322
Item63	2.977	0.003	-0.173	0.862	0.115	0.908
Item64	-0.300	0.764	1.962	0.050	-1.536	0.124
Item65	0.549	0.583	0.977	0.328	-2.212	0.027
Item66	0.229	0.819	1.608	0.108	-1.166	0.244
Item67	-2.608	0.009	2.340	0.019	-4.528	0.000
Item68	-1.350	0.177	1.633	0.103	-0.681	0.496
Item69	0.215	0.829	0.279	0.780	-0.424	0.671
Item70	--	--	-0.786	0.432	-0.938	0.348
Item71	-0.262	0.794	0.409	0.683	-0.069	0.945
Item72	-1.550	0.121	1.862	0.063	-2.335	0.020
Item73	-0.074	0.941	0.177	0.859	-0.801	0.423
Item74	-2.005	0.045	1.135	0.257	-2.215	0.027
Item75	1.133	0.257	0.103	0.918	0.677	0.499
Item76	1.401	0.161	-1.104	0.270	0.586	0.558
Item77	-0.597	0.550	1.668	0.095	-2.522	0.012
Item78	0.450	0.653	0.908	0.364	-1.428	0.153



Tab. h: Itemfit-Statistik, vor Itemselektion, Subtest „Wortschatz“, ( $\alpha=0,01$ )

Item	$\chi^2$ (LRT)	df	p-Wert	Item	$\chi^2$ (LRT)	df	p-Wert
Item1	152.766	166	0.761	Item40	499.488	383	0.000
Item2	285.011	166	0.000	Item41	750.178	383	0.000
Item3	134.364	166	0.966	Item42	369.783	383	0.677
Item4	205.078	166	0.021	Item43	279.806	383	1.000
Item5	284.986	166	0.000	Item44	349.764	383	0.887
Item6	222.707	166	0.002	Item45	301.686	383	0.999
Item7	208.121	166	0.015	Item46	347.119	383	0.906
Item8	140.551	166	0.925	Item47	330.269	383	0.976
Item9	165.628	166	0.494	Item48	852.974	383	0.000
Item10	150.945	166	0.793	Item49	407.786	383	0.184
Item11	355.104	390	0.897	Item50	289.461	385	1.000
Item12	401.554	390	0.332	Item51	436.916	385	0.035
Item13	606.374	390	0.000	Item52	541.591	385	0.000
Item14	300.677	390	1.000	Item53	333.400	385	0.973
Item15	325.455	390	0.992	Item54	328.861	385	0.982
Item16	261.106	390	1.000	Item55	453.335	385	0.009
Item17	367.628	390	0.786	Item56	419.728	385	0.108
Item18	286.208	390	1.000	Item57	290.000	385	1.000
Item19	382.269	390	0.601	Item58	356.405	385	0.849
Item20	394.211	390	0.431	Item59	462.754	385	0.004
Item21	472.430	388	0.002	Item60	447.580	385	0.015
Item22	303.268	388	0.999	Item61	366.201	385	0.747
Item23	425.512	388	0.092	Item62	324.047	385	0.989
Item24	246.606	388	1.000	Item63	3338.683	385	0.000
Item25	246.825	388	1.000	Item64	401.599	385	0.270
Item26	797.740	388	0.000	Item65	431.114	385	0.052
Item27	210.093	388	1.000	Item66	596.184	385	0.000
Item28	477.217	388	0.001	Item67	293.879	385	1.000
Item29	450.962	388	0.015	Item68	284.970	385	1.000
Item30	398.451	388	0.346	Item69	339.904	385	0.952
Item31	338.142	383	0.952	Item70	55.674	164	1.000
Item32	287.714	383	1.000	Item71	138.679	164	0.925
Item33	634.755	383	0.000	Item72	119.876	164	0.996
Item34	331.718	383	0.972	Item73	175.224	164	0.260
Item35	246.628	383	1.000	Item74	98.856	164	1.000
Item36	418.996	383	0.099	Item75	178.748	164	0.204
Item37	292.112	383	1.000	Item76	136.294	164	0.944
Item38	417.426	383	0.109	Item77	385.997	164	0.000
Item39	319.156	383	0.992	Item78	160.585	164	0.561

Tab. i: Itemparameter (Leichtigkeitsparameter), vor Itemselektion, Subtest „Wortschatz“

Item Nr.	Itemparameter		Item-Nr.	Itemparameter		Item-Nr.	Itemparameter	
	Score<= Median	Score>= Median		Männlich	weiblich		8-12 Jahre	13-16 Jahre
1	1.8724885	1.2201327	1	1.3413680	2.1393876	1	1.770028	1.4201213
2	1.0143192	0.3401205	2	0.4386294	1.1040336	2	0.854323	0.4238600
3	0.5926987	0.5391099	3	1.0240872	0.0651635	3	0.307188	1.2151187
4	-2.4737378	-3.008793	4	-2.956995	-2.933844	4	-2.83847	-3.186997
5	0.2619991	-0.657659	5	-0.318927	-0.138283	5	0.159710	-1.082156
6	-1.934042	-2.476637	6	-2.444445	-2.335542	6	-2.37560	-2.504569
7	0.9620891	0.1636724	7	1.4243123	-0.070832	7	0.753382	0.2938678
8	1.3814973	1.9550778	8	1.5090155	1.3594379	8	1.377144	1.6534777
9	-1.736793	-1.337220	9	-1.410021	-1.549666	9	-1.41109	-1.677415
10	-1.052968	-1.169705	10	-1.175511	-1.163508	10	-1.61684	-0.592014
11	0.0720234	-0.026727	11	0.1251338	-0.088207	11	0.101507	-0.220642
12	-1.395983	-1.280561	12	-1.578331	-1.122309	12	-1.38740	-1.378705
13	2.1021673	1.7186559	13	2.0360037	2.0504447	13	2.268930	1.1522255
14	0.3001010	0.7395384	14	0.4989029	0.2962375	14	0.416876	0.3531329
15	0.1407284	0.7395384	15	0.0653284	0.5342521	15	0.374197	0.0171785
16	0.5046305	1.2927174	16	0.4345487	0.9000565	16	0.656343	0.6672191
17	1.0405858	1.4173771	17	0.9810413	1.2225581	17	1.169512	0.8458753
18	-0.710791	-0.149736	18	-0.619481	-0.364133	18	-0.47335	-0.593087
19	-3.166886	-3.859764	19	-3.809424	-3.704068	19	-3.72064	-3.922812
20	-2.030128	-1.691161	20	-2.111063	-1.543750	20	-1.63487	-2.272642
21	-0.000914	-0.345283	21	-0.135292	-0.131617	21	-0.12102	-0.117760
22	2.1245406	2.8180931	22	1.8131865	2.7648425	22	2.385041	1.4744060
23	3.0572462	2.1124918	23	2.7768905	3.1722156	23	3.081736	2.4476055
24	0.1359328	0.7958255	24	0.5272178	0.0660039	24	0.339526	0.3213076
25	-3.042306	-2.742677	25	-2.909356	-2.747773	25	-3.27180	-2.542155
26	0.1586325	-0.577435	26	-0.243596	-0.000578	26	0.19767	-0.990717
27	2.3880858	2.8180932	27	2.2608309	2.6757563	27	2.589639	1.8722545
28	0.4304129	0.1393128	28	0.1144143	0.6350963	28	0.525499	-0.182286
29	-1.697312	-1.752080	29	-1.527369	-2.044488	29	-1.97682	-1.488274
30	-1.653665	-1.350262	30	-1.820750	-1.107448	30	-1.25627	-1.964233
31	-3.427162	-3.316128	31	-3.451988	-3.205988	31	-3.29534	-3.457063
32	-0.895486	-0.609335	32	-0.887358	-0.604571	32	-0.70504	-0.836718
33	-0.866305	-1.972490	33	-1.458145	-1.554802	33	-1.01823	-2.429385
34	1.0678670	1.3843176	34	1.3077904	0.9548184	34	1.128061	1.5200246
35	-0.316893	0.334473	35	-0.050551	-0.132386	35	-0.19157	0.4108434
36	-1.738108	-1.726815	36	-1.799041	-1.650683	36	-1.55523	-2.044493
37	1.3453540	1.8726077	37	1.3911221	1.4730562	37	1.54895	1.048803
38	-2.703811	-3.090625	38	-3.130138	-2.868641	38	-2.82207	-3.275417
39	-0.671439	-0.547281	39	-0.583387	-0.633573	39	-0.52466	-0.787758
40	1.393814	1.049067	40	1.049697	1.651365	40	1.46253	0.926981
41	3.6016726	1.5747662	41	2.9570546	3.5257919	41	3.423348	2.2427271
42	-0.803052	-0.877519	42	-0.930872	-0.754820	42	-0.73683	-1.008297
43	0.9304293	1.3357786	43	1.2377576	0.8041883	43	1.035803	1.1331562
44	0.6992338	1.2324228	44	0.7355077	0.9127214	44	0.858914	0.8329019
45	-0.773898	-0.283090	45	-0.700846	-0.400041	45	-0.77948	0.1503175
46	2.0694502	1.8781689	46	1.8909891	2.2255045	46	2.227698	1.3711859
47	0.5172707	0.9667406	47	0.6036621	0.6640471	47	0.624665	0.8329019
48	-3.606759	-3.662828	48	-3.795735	-3.511283	48	-3.98813	-3.440085
49	0.837255	0.559533	49	0.906670	0.629748	49	0.79411	0.832901
50	-0.072526	0.1274665	50	0.1587820	-0.222050	50	-0.01810	0.0877330

51	2.8997412	2.2839529	51	2.9395949	2.7127390	51	3.016597	2.1492674
52	-0.568357	-1.321447	52	-1.065420	-0.811006	52	-0.77839	-1.202834
53	1.2890088	1.7003192	53	1.0971063	1.7125304	53	1.420234	1.2115417
54	-1.476778	-1.431204	54	-1.309673	-1.686725	54	-1.72369	-0.942091
55	1.4851547	2.0528499	55	1.5665273	1.5646620	55	1.604715	1.5010746
56	-0.568357	-0.836438	56	-0.595209	-0.844551	56	-0.69294	-0.633414
57	-0.006102	1.1197360	57	0.2666537	0.2732394	57	0.163069	0.6769786
58	-1.111298	-1.238476	58	-1.038777	-1.411055	58	-1.40290	-0.722562
59	-0.300444	-0.203436	59	-0.260530	-0.286964	59	-0.34084	-0.016317
60	2.3312946	2.0211695	60	2.6458912	2.0180672	60	2.327597	2.241627
61	0.9368939	1.7922041	61	0.8102856	1.3235667	61	1.076405	0.9766642
62	1.2747821	1.4430721	62	1.2380672	1.3668844	62	1.221853	1.6596671
63	-3.477415	-5.747974	63	-5.015279	-4.884317	63	-4.80155	-4.891550
64	-0.146782	-0.045938	64	0.147149	-0.465551	64	-0.25515	0.2943109
65	0.5003496	0.306429	65	0.586463	0.2738910	65	0.266257	1.1457437
66	-3.477415	-3.603308	66	-3.318479	-4.131187	66	-3.76900	-3.207474
67	-2.657748	-1.596994	67	-1.624246	-2.438278	67	-2.56096	-0.903036
68	1.7061969	2.7261855	68	2.1234795	1.5015384	68	1.740749	2.0738251
69	1.9064318	1.7922040	69	1.9574667	1.8497651	69	1.865426	2.0738251
70	--	--	70	2.3829615	2.8641552	70	2.479155	3.3710637
71	1.3530534	1.5150976	71	1.4789425	1.2786274	71	1.371075	1.4121491
72	-0.529940	0.2115382	72	0.1343072	-0.693966	72	-0.55205	0.6804135
73	2.1739867	2.2331757	73	2.2520490	2.1532842	73	2.100317	2.7610754
74	0.2783815	1.5150977	74	0.788726	0.2749438	74	0.28020	1.6560679
75	0.2295124	-0.269479	75	-0.002029	-0.047411	75	0.106049	-0.224936
76	2.1073244	1.279555	76	1.6787689	2.2758551	76	2.031366	1.6560679
77	-2.179021	-1.826264	77	-1.634184	-2.521311	77	-2.50093	-1.110659
78	1.796000	1.515097	78	2.008080	1.5367463	78	1.594015	2.7610754

## 15.6 Tabellen „Alltagswissen“

Tab. j: Erste Parameterschätzung, z-Test, Subtest „Alltagswissen“, ( $\alpha=0,01$ )

Item	z-Wert Score	p-Wert Score	z-Wert Geschlecht	p-Wert Geschlecht	z-Wert Alter	p-Wert Alter
Item1	2.830	0.005	-6.189	0.000	2.828	0.005
Item2	--	--	-1.874	0.061	0.758	0.448
Item3	1.067	0.286	0.803	0.422	-1.008	0.314
Item4	1.017	0.309	-2.518	0.012	-0.489	0.625
Item5	0.757	0.449	-3.109	0.002	4.380	0.000
Item6	0.126	0.899	0.749	0.454	0.638	0.523
Item7	-0.226	0.821	1.082	0.279	2.228	0.026
Item8	0.095	0.924	0.378	0.705	2.390	0.017
Item9	1.012	0.312	-0.902	0.367	1.831	0.067
Item10	2.373	0.018	-0.257	0.797	3.123	0.002
Item11	-0.370	0.711	-0.928	0.353	4.182	0.000
Item12	3.200	0.001	-0.941	0.347	3.562	0.000
Item13	1.381	0.167	-0.208	0.835	2.779	0.005
Item14	1.301	0.193	-2.226	0.026	3.780	0.000
Item15	2.541	0.011	-3.726	0.000	5.382	0.000
Item16	5.307	0.000	-1.481	0.139	5.652	0.000
Item17	--	--	1.654	0.098	0.749	0.454
Item18	1.393	0.164	3.082	0.002	0.861	0.389
Item19	4.541	0.000	-5.392	0.000	5.337	0.000
Item20	--	--	0.508	0.611	-0.202	0.840
Item21	0.127	0.899	-0.715	0.475	-3.649	0.000
Item22	0.303	0.762	0.773	0.439	-0.371	0.711
Item23	-0.187	0.852	0.278	0.781	0.306	0.760
Item24	-0.976	0.329	-1.986	0.047	-5.172	0.000
Item25	0.543	0.587	-0.757	0.449	-2.587	0.010
Item26	1.700	0.089	1.764	0.078	2.390	0.017
Item27	-0.839	0.401	-0.072	0.943	-2.908	0.004
Item28	--	--	-0.982	0.326	-1.099	0.272
Item29	-0.468	0.640	5.415	0.000	0.705	0.481
Item30	-0.725	0.468	0.622	0.534	-0.928	0.353
Item31	3.410	0.001	-1.620	0.105	2.204	0.028
Item32	-2.552	0.011	5.006	0.000	-2.690	0.007
Item33	2.763	0.006	-2.416	0.016	4.086	0.000
Item34	1.270	0.204	4.687	0.000	1.465	0.143
Item35	0.820	0.412	-0.961	0.337	-2.162	0.031
Item36	1.835	0.067	-1.694	0.090	-1.854	0.064
Item37	-1.538	0.124	1.371	0.170	-1.803	0.071
Item38	-2.506	0.012	3.199	0.001	-0.803	0.422
Item39	-2.958	0.003	-0.090	0.929	-5.820	0.000
Item40	-3.343	0.001	3.357	0.001	-0.609	0.543
Item41	-0.393	0.694	2.074	0.038	1.218	0.223
Item42	1.337	0.181	--	--	-0.044	0.965
Item43	-4.960	0.000	1.305	0.192	-4.035	0.000
Item44	1.276	0.202	0.649	0.517	0.090	0.928
Item45	1.738	0.082	-2.820	0.005	1.416	0.157
Item46	0.642	0.521	1.101	0.271	0.744	0.457

Item47	3.952	0.000	-1.112	0.266	2.631	0.009
Item48	-2.645	0.008	-1.165	0.244	-1.252	0.211
Item49	-1.129	0.259	1.171	0.242	-0.191	0.848
Item50	-0.903	0.367	0.146	0.884	-1.957	0.050
Item51	-3.200	0.001	-0.397	0.692	-4.048	0.000
Item52	-2.040	0.041	0.037	0.971	-0.973	0.331
Item53	-2.721	0.007	1.037	0.300	-2.061	0.039
Item54	-2.383	0.017	-0.807	0.420	-4.502	0.000
Item55	-3.077	0.002	2.094	0.036	-1.546	0.122
Item56	-1.482	0.138	0.591	0.554	0.612	0.541
Item57	--	--	--	--	--	--
Item58	-4.299	0.000	1.055	0.291	-2.958	0.003
Item59	--	--	--	--	--	--

Tab. k: Itemfit-Statistik, vor Itemselektion, Subtests „Alltagswissen“, ( $\alpha=0,01$ )

Item	$\chi^2$ (LRT)	df	p-Wert	Item	$\chi^2$ (LRT)	df	p-Wert
Item1	210.910	169	0.016	Item31	419.056	368	0.034
Item2	94.929	169	1.000	Item32	220.203	368	1.000
Item3	247.893	169	0.000	Item33	562.927	368	0.000
Item4	153.840	169	0.792	Item34	421.089	368	0.029
Item5	171.345	169	0.435	Item35	395.286	368	0.157
Item6	135.793	169	0.972	Item36	403.863	368	0.096
Item7	119.042	169	0.999	Item37	172.283	368	1.000
Item8	121.375	169	0.998	Item38	337.096	380	0.945
Item9	594.048	394	0.000	Item39	278.545	380	1.000
Item10	421.393	394	0.164	Item40	309.320	380	0.997
Item11	282.968	394	1.000	Item41	342.296	380	0.918
Item12	408.827	394	0.293	Item42	1354.789	380	0.000
Item13	315.092	394	0.999	Item43	209.766	380	1.000
Item14	332.808	394	0.989	Item44	419.565	380	0.079
Item15	489.016	394	0.001	Item45	373.952	384	0.634
Item16	1020.282	395	0.000	Item46	200.683	384	1.000
Item17	181.899	395	1.000	Item47	415.083	384	0.132
Item18	352.772	395	0.938	Item48	924.210	384	0.000
Item19	502.238	395	0.000	Item49	662.981	384	0.000
Item20	184.186	395	1.000	Item50	253.030	384	1.000
Item21	265.909	395	1.000	Item51	324.427	384	0.988
Item22	177.866	395	1.000	Item52	227.992	383	1.000
Item23	285.880	395	1.000	Item53	128.566	164	0.981
Item24	246.523	378	1.000	Item54	133.515	164	0.961
Item25	329.663	378	0.965	Item55	113.657	164	0.999
Item26	352.206	378	0.825	Item56	166.481	164	0.431
Item27	314.509	378	0.992	Item57	--	--	--
Item28	224.889	378	1.000	Item58	86.981	164	1.000
Item29	307.240	378	0.997	Item59	90.521	164	1.000
Item30	243.595	378	1.000				

Tab. I:Itemparameter(Leichtigkeitsparameter),vor Itemselektion, Subtest „Alltagswissen“




Item Nr.	Itemparameter		Item-Nr.	Itemparameter		Item-Nr.	Itemparameter	
	Score<= Median	Score>= Median		männlich	weiblich		8-12 Jahre	13-16 Jahre
1	0.4976842	-0.72721784	1	-1.81089382	1.03911047	1	0.5041459	-0.756238
2	--		2	1.30191009	2.25080395	2	2.3188555	1.892164
3	-1.713363	-2.36094858	3	-2.29341882	-2.79833591	3	-3.091200	-2.253632
4	0.2498801	-0.20732417	4	-0.82972881	0.27493698	4	-0.287772	-0.055371
5	2.4043662	1.88539004	5	1.12752185	2.75385176	5	3.1625597	0.927507
6	-0.149012	-0.20732417	6	-0.35153068	-0.68670329	6	-0.198333	-0.495852
7	-1.081888	-0.96684621	7	-1.07435792	-1.60199903	7	-0.477027	-1.551345
8	1.2043411	1.15251169	8	0.88729145	0.71354849	8	1.5471800	0.4525488
9	2.9227706	2.32958429	9	2.21757005	2.60748176	9	3.1004110	2.2286933
10	-0.457896	-1.25669714	10	-1.23872042	-1.15393223	10	-0.394674	-1.456135
11	1.3303999	1.49275063	11	0.75902493	1.07593397	11	1.8422506	0.3907819
12	-0.899807	-2.04036186	12	-2.03776884	-1.70830471	12	-0.921804	-2.209684
13	0.3096431	-0.15457067	13	-0.26576477	-0.19895372	13	0.4674650	-0.451207
14	0.4584891	0.01638594	14	-0.41288535	0.30229313	14	0.7928143	-0.451207
15	0.9966495	0.12823445	15	-0.24099047	0.98043826	15	1.4557477	-0.327270
16	2.4549974	0.68184581	16	1.39063784	1.88082117	16	2.6637889	0.7842209
17	--	--	17	4.92724538	3.54674031	17	4.5395773	3.9282265
18	-0.716932	-1.16701212	18	-0.88353093	-1.85196395	18	-0.931924	-1.211320
19	0.6116852	-0.68604738	19	-1.03413430	0.53239224	19	0.6892254	-0.837460
20	--	--	20	4.21215400	3.85445434	20	4.4131982	4.6289243
21	-1.122885	-1.16701216	21	-1.48666570	-1.26714791	21	-2.733209	-0.907821
22	-3.391102	-3.62403425	22	-3.60584170	-4.01709272	22	-3.981251	-3.681641
23	-3.391102	-3.24907050	23	-3.37260792	-3.50238883	23	-3.211816	-3.400372
24	-1.179671	-0.85034490	24	-1.51524642	-0.92907539	24	-3.681967	-0.472956
25	-0.140709	-0.29477532	25	-0.63838079	-0.42830269	25	-0.878399	-0.108770
26	1.126443	0.63394151	26	0.85456080	0.36272073	26	1.0422726	0.3852644
27	-0.735578	-0.47702126	27	-0.88352073	-0.86309268	27	-1.345740	-0.425003
28	--	--	28	-3.51845806	-3.06769031	28	-3.681967	-2.973937
29	-1.052398	-0.89938917	29	-0.55873372	-2.30781525	29	-0.962657	-1.174950
30	-1.486001	-1.22037332	30	-1.48237141	-1.67537198	30	-1.639805	-1.327498
31	1.164925	0.23409847	31	0.27012051	0.73004115	31	0.8773755	0.253331
32	-1.331107	-0.41244944	32	-0.34115394	-1.94619694	32	-1.418710	-0.549569
33	1.4460250	0.66306421	33	0.54361861	1.24684173	33	1.5140762	0.3366078
34	0.5299956	0.17832854	34	0.71344491	-0.65483565	34	0.3722509	-0.040240
35	0.3793787	0.15082059	35	-0.15531655	0.11510005	35	-0.267511	0.3647558
36	-1.177196	-1.83024106	36	-2.22671337	-1.65094714	36	-2.355389	-1.618541
37	-3.957009	-2.39279497	37	-2.57302328	-3.16466476	37	-3.449913	-2.422245
38	0.217931	1.01371877	38	0.86680276	-0.13854891	38	0.3521122	0.601693
39	0.315343	1.28904193	39	0.55736082	0.58520357	39	-0.229217	1.841939
40	-0.774584	0.29714383	40	0.09139869	-0.98229588	40	-0.381846	-0.191702
41	0.891066	1.01371875	41	1.14159153	0.48675698	41	1.0413730	0.663785
42	-3.826586	-5.08501296	42	--	--	42	-4.874680	-4.824653
43	-0.315556	1.39403059	43	0.43279512	0.02741542	43	-0.319864	0.999447
44	-3.415224	-4.37465402	44	-4.29791740	-4.85738372	44	-4.170907	-4.246531
45	1.787572	1.17197974	45	1.11892223	2.14996889	45	1.7729565	1.264662
46	1.712039	1.48080827	46	1.81550673	1.41369278	46	1.7438221	1.474154

47	2.521182	1.09908736	47	1.81550674	2.23192249	47	2.4083786	1.438344
48	0.0266568	0.95733473	48	0.22837191	0.63734555	48	0.1946566	0.6336882
49	0.8852324	1.28425467	49	1.20102723	0.78547114	49	0.9988471	1.0662447
50	1.6869253	2.02744495	50	1.84707800	1.79328621	50	1.5128403	2.2567272
51	-0.958580	0.21061306	51	-0.47796998	-0.33538222	51	-1.351486	0.2273431
52	-0.604258	0.11846883	52	-0.32223087	-0.33538224	52	-0.470732	-0.122812
53	1.4785563	2.81271500	53	2.21604724	1.74236417	53	1.5573618	2.5177634
54	1.4785563	2.61538557	54	1.78070352	2.14459391	54	0.8812321	3.1807322
55	0.5328348	2.06782615	55	1.72705397	0.72727291	55	0.9583092	1.6802408
56	1.5336188	2.20842117	56	1.94230504	1.67283382	56	1.9520272	1.6802408
57	--	--	57	--	--	57	--	--
58	0.7232615	3.04257394	58	1.94230505	1.45816973	58	1.0327007	2.4459425
59	--	--	59	--	--	59	--	--



# Curriculum Vitae

## Persönliche Daten

Name Petra Nicola Hofmayer  
Geburtstag/-ort 19. Februar 1984 in Wien  
Staatsangehörigkeit Österreich  
Kontaktmöglichkeiten  Wiedner Hauptstraße 135/16; 1050 Wien  
 0699/12853283  
 [petra.hofmayer@gmail.com](mailto:petra.hofmayer@gmail.com)

## Aus- und Weiterbildung

seit Oktober 2003 Studium der **Psychologie** an der Universität Wien.  
Russisch-Sprachkurse an der Slawistik  
seit Oktober 2009 Ausbildung zur multimedialen Kunsttherapeutin am ÖAGG  
Abschluss im Mai 2012  
2003 Matura an der AHS Franklinstraße 26, 1210 Wien – wirtschaftskundliches Realgymnasium

## Arbeitserfahrung

Februar 2011 bis Februar 2012 **Studienabschlussstipendium**  
Erhalt eines Stipendiums des Europäischen Sozialfonds für die Studienabschlussphase

September 2011 bis Dezember 2011 **Kunsttherapeutisches Praktikum**  
Kunsttherapeutische Arbeit unter Supervision im gerontopsychologischen Bereich in einer Wohngemeinschaft für Demenzkranke bei der Caritas Socialis

Juli 2010 bis Dezember 2010 **AMS Niederösterreich**  
Durchführung eines Pilotprojekts im Bereich Eignungsdiagnostik mit Persönlichkeitsschwerpunkt (HILL Best Fit), Karrierecoaching und Einzelberatung

Feb. 2009 bis September 2010 **Feinkost MEDIA GmbH**  
*Projektmanagement*  
Konzeption und Umsetzung von Projekten im klassischen PR und Video-PR Bereich, journalistische Tätigkeit, Eventmanagement, Administrations- und Organisationagenden

Oktober 2008 bis April 2009 **Psychologisches Pflichtpraktikum beim AMS Niederösterreich**  
Psychologische Testung von arbeitssuchenden Personen und Beratung; Gutachtenerstellung für AMS interne Verwendung und Versicherungen

März 2007 bis November 2008 **ALARIS Projects GmbH**  
*Projektleitung* eines EU kofinanzierten Arbeitsmarktintegrationsprojekts für Jugendliche und KMUs; psychosoziale Betreuung von Jugendlichen, Coaching- und Beratungstätigkeit

November 2004 bis März 2007 **Zeit!Raum – Verein für soziokulturelle Arbeit**  
*Projekt Management* "bz-future – Berufsinformationszentrum für künftige Flächenberufe"  
*Teamleitung* in der aufsuchenden Kinder- und Jugendarbeit im 22. Bezirk, pädagogische Konzeption, Beratungstätigkeit