



universität
wien

Diplomarbeit

Titel der Arbeit

Aktualisierung und Itemkonstruktion zu den Untertests
U3 „Angewandtes Rechnen“ und U9 „Funktionen
Abstrahieren“ der Intelligenztestbatterie AID 3

Verfasserin

Nicole Görner

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im April 2013

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Dr. Stefana Holocher-Ertl

Danksagung

An erster Stelle möchte ich meinem Mann, **Werner Schostal**, und meinen Kindern, **Lukas** und **Lena**, einen besonders großen Dank aussprechen, die es mir ermöglicht haben, neben den familiären und beruflichen Pflichten studieren und letztendlich diese Arbeit abfassen zu können.

Mein großer Dank gilt **Mag. Dr. Stefana Holocher-Ertl**, die mir das Thema der vorliegenden Diplomarbeit vorgeschlagen und schmackhaft gemacht hat, was für mich der Anstoß war, überhaupt mit dieser Arbeit zu beginnen, obwohl ich erst am Anfang des zweiten Studienabschnittes gewesen bin. Sie stand mir bei Unklarheiten, Problemen und der gleichen jederzeit mit ihrem Fachwissen und ihrer Hilfsbereitschaft zur Verfügung. Meinem Kollegen Herrn **Thomas Scheck** danke ich für die große Unterstützung bei der Auswertung der Daten mit dem Programm R.

Weiters möchte ich meinen Kolleginnen **Melanie Hellebart** und **Eva Karesch** für ihre Unterstützung während des gesamten Studiums und meiner Kollegin **Eva Klug** für die Begleitung während der Diplomarbeitsphase unter anderem für ihre interessanten Anregungen, Hilfestellungen, Diskussionen sowie fürs Korrekturlesen danken.

Danken möchte ich auch allen, an dieser Studie teilnehmenden Kindern und Jugendlichen, mit denen mir die Zusammenarbeit großen Spaß gemacht hat und für mich sehr wertvoll war. Weiters möchte ich mich bei den DirektorInnen und LehrerInnen für ihre Kooperation und ihr Interesse bedanken. Besonderer Dank gilt an dieser Stelle den Lehrerinnen **Isabella Nistelberger** und **Esther Wertitsch**, welche mir die Kooperative Mittelschule und die Ganztagsvolksschule vermittelt haben.

Zu guter letzt möchte ich mich bei all jenen Personen bedanken, die mich bei dem gesamten Forschungsprojekt und bei der Erstellung dieser Diplomarbeit unterstützt und begleitet haben.

Abstract Deutsch

Das Adaptives Intelligenz Diagnostikum - aktuelle Version 2.2 (AID 2, Kubinger, 2009) ist eine Intelligenztestbatterie, welche basale und komplexe Fähigkeiten von Kindern und Jugendlichen im Alter von 6;0 bis 15;11 Jahren erfasst. Sie wird häufig in der Diagnostik bei Lern- und Verhaltensauffälligkeiten, sowie in der Schulpsychologie angewendet. Auf Grund dessen und der von der DIN 33430 alle acht Jahre geforderten Überprüfung von psychologisch diagnostischen Verfahren und der Notwendigkeit der Aktualisierung von Aufgaben, wurde eine Revision des AID 2 zum Adaptiven Intelligenz Diagnostikum – Version 3.1 (AID 3, Kubinger & Holoher-Ertl, in Druck) durchgeführt. Die vorliegende Arbeit befasst sich mit der Revision der Untertests „Angewandtes Rechnen“ und „Funktionen Abstrahieren“. Viele Items wurden aus dem AID 2 (Version 2.2) übernommen, einige von ihnen wurden überarbeitet und aktualisiert und andere ganz entfernt und durch neue Items ersetzt. Von den 70 für den Untertest „Angewandtes Rechnen“ in die Datenüberprüfung einbezogenen Items, sind 16 neu konstruiert worden und von den 69 Items des Untertests „Funktionen Abstrahieren“ sind 18 Items neu konstruiert worden. Die entstandene Rohversion des AID 3 bestand, inklusive der zwei genannten Tests, aus 10 Untertests, welche von mehreren Diplomanden (siehe Hagenmüller, 2011; Hellebart, 2013; Weber, 2011) einer Stichprobe von 711 Kindern und Jugendlichen vorgegeben wurde. Die anschließende Datenanalyse der Untertests „Angewandtes Rechnen“ und „Funktionen Abstrahieren“ auf Geltung des Rasch-Modells zeigte, dass für den Untertest „Angewandtes Rechnen“ nach Ausschluss von 11 Items a posteriori die Gültigkeit des Rasch-Modells für die Teilungskriterien *Geschlecht* und *Alter* angenommen werden kann. Aber nicht in Bezug auf die Teilungskriterien *Rohscore* und *Muttersprache*. Dieser benachteiligt demnach keine Testpersonen hinsichtlich ihres Geschlechts und ihres Alters. In Bezug auf diese Teilungskriterien ist anzumerken, dass der graphische Modelltest eine relativ gute Modellpassung zeigt und dass beliebig kleine Unterschiede bei der Modellprüfung für hinreichend große Stichproben signifikant werden, ohne dass diese Unterschiede inhaltlich relevant sein müssen. Nach dem Ausschluss von 14 Items beim Untertest „Funktionen Abstrahieren“ konnte a posteriori die Gültigkeit des Rasch-Modells für alle Teilungskriterien *Rohscore*, *Geschlecht*, *Muttersprache* und *Alter* angenommen werden, was bedeutet, dass dieser Untertest keine Personen aufgrund ihrer Geschlechtszugehörigkeit, ihrer Muttersprache und ihres Alters benachteiligt und die Testergebnisse der Testpersonen zu Testwerten verrechnet werden können. Offen bleibt trotzdem die Frage, inwieweit die Items Kinder und Jugendliche mit nicht deutscher Muttersprache benachteiligen.

Abstract English

The *Adaptives Intelligenz Diagnostikum* - version 2.2 (*AID 2*, Kubinger, 2009) is an intelligence test which allows the measurement of basic and complex cognitive skills of children at the age of 6;0 to 15;11 years. It is mainly applied to analyze and diagnose behavioral and learning disorders and is often used in the field of school psychology. Because of the necessary update of psychological tests, which is claimed from the DIN 33430 every eight years, a revision of the *AID 2* into the *Adaptive Intelligenz Diagnostikum* – version 3.1 (*AID 3*, Kubinger & Holocher-Ertl, in press) was developed.

This paper focuses on the two constructs “Angewandtes Rechnen” (“Numeracy”) and “Funktionen abstrahieren” (“Prescind features”). Many Items of these two constructs were taken from the *AID 2* (version 2.2), some were revised and adapted and others were created new to replace old items. The construct “Angewandtes Rechnen” (“Numeracy”) contains 70 items of which 16 were new items, which replaced some inadequate ones of version 2.2. The construct “Funktionen abstrahieren” (“Prescind features”) contains 69 items with 18 new ones. The first draft of the *AID 3* contains 10 sub-scales, which were tested on 711 children and adolescents.

The data were analyzed and a validity check for the Rasch-Model was calculated for the two constructs named above. After 11 items were excluded from the scale “Angewandtes Rechnen” (“Numeracy”) a posteriori, the validity of the Rasch-Model can be accepted for the criteria Sex and Age, but not for Score and Mother tongue. This means that no participants were discriminated against on account of sex and age, but it is nor sure if it is a fair measurement on account of mother tongue. The graphical model test shows a relatively good model fit for the criteria. After 14 items were excluded from the scale “Funktionen Abstrahieren” (“Prescind features”) a posteriori, the validity of the Rasch-Model can be accepted for all the criteria used, *Score*, *Mother tongue*, *Sex* and *Age*. This means that no participants were discriminated against on account of their mother tongue, sex, age and the overall score in the test of a person can be interpreted as his or her ability.

The question remains open, if participants are disadvantaged in items measuring numeracy when their mother tongue is not German.

Inhaltsverzeichnis

I. Einleitung	3
II. Theoretischer Teil	7
1. Definition von Intelligenz?.....	7
2. Intelligenzmodelle	7
2.1. Intelligenzkonzept von Wechsler	8
2.2. Zwei-Faktoren-Theorie (nach Spearman, 1904).....	9
2.3. Multiple Faktorentheorie (nach Thurston, 1938).....	9
2.4. Theorie der fluiden und kristallinen Intelligenz (nach Horn und Cattell, 1966)	9
2.5. Three Stratum Theorie (nach Carroll, 1993).....	10
2.6. CHC-Theorie (nach Mc Grew, 2005).....	10
3. Der Begriff „Intelligenz“ im AID 2	12
4. Der AID und seine Revisionen.....	13
4.1. Der AID 2	14
4.2. Adaptives Testen – branched testing.....	17
4.3. Angewandtes Rechnen	18
4.3.1. Wie wird die Fähigkeit in anderen Tests abgeprüft?	19
4.4. Funktionen Abstrahieren	21
4.4.1. Wie wird die Fähigkeit in anderen Tests abgeprüft?	22
5. Das Dichotom logistische Modell von Rasch	23
III. Empirischer Teil	27
6. Hintergrund und Ziel der Untersuchung	27
7. UT 3 Angewandtes Rechnen	27
7.1. Theoretischer Hintergrund.....	27
7.2. Definition der zu messenden Fähigkeit	32
7.3. Testart und Aufgabentyp	32
7.4. Das Item.....	33
7.5. Zielgruppe.....	34
7.6. Itemkonstruktion.....	34
8. UT 9 Funktionen Abstrahieren.....	36
8.1. Theoretischer Hintergrund.....	36
8.2. Definition der zu messenden Fähigkeit	38
8.3. Testart und Aufgabentyp	39
8.4. Das Item.....	39

8.5. Zielgruppe.....	40
8.6. Itemkonstruktion.....	40
9. Methode	43
9.1. Hypothesen.....	43
9.2. Untersuchungsplan	44
9.3. Erhebungsinstrument.....	46
9.4. Stichprobe.....	47
10. Ergebnisse.....	52
10.1. Überprüfung der Hypothesen des Untertests Angewandtes Rechnen	52
10.1.1. Erster Berechnungsdurchgang.....	53
10.1.2. Letzter Berechnungsdurchgang	62
10.2. Überprüfung der Hypothesen des Untertests Funktionen Abstrahieren	68
10.2.1. Erster Berechnungsdurchgang.....	69
10.2.2. Letzter Berechnungsdurchgang	78
11. Diskussion und Ausblick	85
12. Zusammenfassung	90
Literaturverzeichnis.....	93
Tabellenverzeichnis.....	97
Abbildungsverzeichnis	101
ANHANG	
LEBENS LAUF	

I. Einleitung

Das *Adaptive Intelligenz Diagnostikum* – aktuelle Version 2.2 (AID 2, Kubinger, 2009) stellt eine Intelligenz-Testbatterie für Kinder und Jugendliche im Alter von 6;0 bis 15;11 Jahren dar, die aufgrund ihrer hohen Güte in der psychologisch-diagnostischen Praxis (Schulpsychologie, Diagnostik bei Lern- und Verhaltensauffälligkeiten) häufig angewendet wird. Bearbeitet ein Kind oder Jugendlicher die Aufgaben des AID 2, so können daraus wesentliche Informationen abgeleitet werden:

- Verbale Fähigkeiten
- Rechnerische Fähigkeiten
- Visuell-gestalthaft intellektuelle Fähigkeiten
- Akustische Merkfähigkeit
- Visuelle Merkfähigkeit
- Informationsverarbeitungsgeschwindigkeit (Auffassungsgabe)
- Verständnis in sozialen Situationen

Zusätzlich können wesentliche Teilleistungsfähigkeiten mit erfasst werden. Dies sind basale neurologische Funktionen welche unabhängig vom Intelligenzniveau Störungen komplexer kognitiver Vorgänge und des kindlichen Verhaltens nach sich ziehen können. Der AID 2 ist vielseitig einsetzbar im Zusammenhang mit schulpsychologischen Fragestellungen, wie unter anderem schulischer Unter- bzw. Überforderung, Verhaltensproblemen im schulischen Kontext, Abklärung einer Aufmerksamkeitsdefizitstörung, Abklärung einer möglichen Legasthenie/Dyskalkulie oder Schullaufbahnberatung.

Eine Revision steht nun aus folgenden Gründen an (welche dem Brief an das Präsidium des Stadtschulrates für Wien vom 19.10.2009 entnommen sind, dieser ist im Anhang vollständig angeführt):

1. Die zur Messung der intellektuellen Fähigkeiten eines Kindes vorgegebenen Aufgaben müssen *aktualisiert* werden, so dass jene *Fähigkeiten und jenes Wissen* überprüft wird, welches *in der heutigen Zeit tatsächlich von Relevanz* ist. So werden z.B. in demjenigen Untertest, welcher das bisher angeeignete Wissen eines Kindes überprüft, neue Wissensgebiete (wie z.B. Informationstechnologien) aufgenommen und historische und geographische

Entwicklungen der letzten 25 Jahre in den Fragen berücksichtigt. Aufgaben die den Wortschatz und das Sprachverständnis der Kinder überprüfen, werden an den heutigen Sprachgebrauch angeglichen.

2. Die Aufgaben werden je gemessener Fähigkeit dahingehend erweitert, dass sowohl *genaue und differenzierte Diagnosen* über intellektuelle Stärken und Schwächen bei *leistungsschwachen* als auch bei sehr *leistungsstarken* Kindern und Jugendlichen möglich sind.
3. Das Konzept der *Teilleistungsstörungen* wird durch die Entwicklung zusätzlicher Aufgaben ausgebaut, so dass mögliche Teilleistungsstörungen noch besser erkannt werden und hilfreiche Interventionsmaßnahmen geplant werden können.
4. Neben der Erfassung der *förderungsabhängigen Intelligenz* (also jener Fähigkeiten, die in ihrer Entwicklung stark von entsprechender familiärer und institutioneller Förderung abhängig sind) wird auch die Erfassung der so genannten „Grundintelligenz“, darunter versteht man die relativ *förderungsunabhängige Fähigkeit* des formal-logischen Denkens, durch die Entwicklung eines neuen Untertests möglich sein. Damit kann zwischen grundsätzlich gut begabten, aber wenig geförderten, und gut begabten und entsprechend gut geförderten Kindern unterschieden werden.
5. Zur möglichst *objektiven und gültigen Einschätzung des Arbeitsverhaltens* eines Kindes (also z.B. seiner Leistungsmotivation, Frustrationstoleranz, Ausdauer und Sorgfalt) wird die Testbatterie um Beobachtungsrichtlinien und -anleitungen für die Testleiter(innen) ergänzt.

In der vorliegenden Arbeit, geht es um die Revision der beiden Untertests „Angewandtes Rechnen“ und „Funktionen Abstrahieren“. Für diese aktualisierten und erweiterten Untertests soll mit Hilfe der pro Untertest aufgestellten Hypothesen überprüft werden, ob sie dem Rasch-Modell entsprechen.

Die Arbeit gliedert sich in einen theoretischen und praktischen Teil. Im theoretischen Teil werden Fragen beantwortet z.B. was man unter Intelligenz versteht und wie sie gemessen wird. Weiters wird kurz der AID 2 und das Dichotom logistische Modell von Rasch vorgestellt. Der empirische Teil beschäftigt sich mit dem Hintergrund und Ziel

der Untersuchung, es werden die erwähnten Untertests genauer vorgestellt, die Methode beschrieben und die Ergebnisse mit einer anschließenden Diskussion dargestellt.

II. Theoretischer Teil

1. Definition von Intelligenz?

Mit dem Begriff „Intelligenz“ werden die geistigen Fähigkeiten, die mit dem Verstand verbunden sind gekennzeichnet. Es existieren recht unterschiedliche Definitionen, welche unter anderem einem geschichtlichen Wandel unterliegen. Intelligenz wird charakterisiert als die Fähigkeit, gut urteilen, verstehen und denken zu können (Binet & Simon, 1905; zitiert nach Daseking, Janke & Petermann, 2006; S. 314). Für Wechsler ist es eine zusammengesetzte und globale Fähigkeit „zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umwelt wirkungsvoll auseinander zu setzen.“ (1956; zitiert nach Daseking et al., 2006; S. 314). Für Sternberg ist Intelligenz ein dynamisches Konstrukt in Abhängigkeit von Umgebungsvariablen. Allen gemeinsam ist, dass sie als das Wesentliche der Intelligenz die Fähigkeit bezeichnen, sich auf Grund von Einsichten in neuen Situationen zurechtzufinden oder mit Hilfe des Denkens Aufgaben zu lösen, ohne dass dafür die Erfahrung eine Rolle spielt, sondern eher die Erfassung von Beziehungen (Häcker & Stapf, 2009).

Grundsätzlich ist zu sagen, dass Intelligenz das am meisten untersuchte psychologische Merkmal ist und den Schwerpunkt in der psychologischen Leistungsdiagnostik bildet.

Bereits Cattell (1987) beschrieb, dass verschiedene Forscher zwar unterschiedliche Theorien zum Konstrukt „Intelligenz“ hervorbrachten, aber grundsätzlich bestand Einigkeit darüber, dass eines der ersten Ziele sei, die Struktur der Fähigkeiten die für intelligentes Verhalten wichtig sind, zu entdecken. Der Fortschritt der Wissenschaft zeigt sich, wie bereits erwähnt in genaueren Beschreibungen und in der Entdeckung neuer Erhebungsmethoden, daher soll im Punkt 2. auf den methodischen Aspekt der Messung von Intelligenz eingegangen werden.

2. Intelligenzmodelle

Über den Aufbau der Intelligenz bestehen verschiedene Theorien. Ganz allgemein werden die Dimensionen bzw. Faktoren intelligenten Verhaltens, da sie nicht direkt beobachtbar sind, aus Tests, Experimenten und Leistungsproben gewonnen und unter Zuhilfenahme der Methode der Faktorenanalyse extrahiert. Innerhalb dieser

faktorenanalytischen Betrachtung der Intelligenz gibt es nun unterschiedliche Modelle, die unter anderem die historische Entwicklung der Intelligenz widerspiegeln, aber auch auf rein methodisch begründbare Unterschiede zurückgehen, wie unter anderem die Zahl der angenommenen Fähigkeiten und deren Beziehungen zueinander. (Dorsch u. Friedrich, 2009). Grundsätzlich wird bei den Faktoriellen Theorien zwischen *Hierarchischen* und *Mehrdimensionalen Modellen* unterschieden. Sie bilden die Basis für einige Intelligenztests für das Kinder- und Jugendalter. Im Weiteren wird zur Illustration nur auf die Ersteren eingegangen.

2.1. Intelligenzkonzept von Wechsler

David Wechsler (1944; S. 3) definierte Intelligenz wie zuvor erwähnt als die „Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen“. Intelligenz ist einerseits ein *globales* Konstrukt und andererseits *spezifisch*, da Intelligenz aus mehreren Faktoren zusammengesetzt ist. Da kein Intelligenztest in der Lage ist, alle Bereiche der kognitiven Funktionen zu erfassen, wählte Wechsler eine Form der Messung aus, die diejenigen Bereiche abdecken soll, die er als bedeutsame kognitive Aspekte der Intelligenz erachtete, wie *Sprachliches Verständnis*, *abstraktes logisches Denken*, *Wahrnehmungsorganisation*, *mengenbezogenes Denken*, *Gedächtnis* und *Verarbeitungsgeschwindigkeit*. Er wählte eine hinreichend große Anzahl von Untertests aus, um klinisch relevante Informationen über das kognitive Niveau eines Menschen in einer angemessenen Zeitspanne zu erhalten. Letztendlich reflektieren aber die Leistungen in einem Test nur einen Anteil dessen, was Intelligenz ausmacht. Wechsler vermied es, Intelligenz in rein kognitiven Begriffen zu definieren, weil er glaubte, dass diese Faktoren nur einen Ausschnitt der Intelligenz erfassen. Er nahm an, dass weitere Eigenschaften, wie *Planung* und *Zielbewusstsein*, *Begeisterungsfähigkeit*, *Feldabhängig/-unabhängigkeit*, *Impulsivität*, *Ängstlichkeit* und *Ausdauer* zu intelligentem Verhalten beitragen, die die Leistung eines Kindes bzw. Erwachsenen bei einem Test wie auch seine Leistungsfähigkeit im tägliche Leben beeinflussen. Nicht kognitive Merkmale könnten unter Umständen dazu führen, dass Personen mit gleichen Leistungen in einem Test alltägliche Anforderungen unterschiedlich meistern (Petermann & Petermann, 2007). Für die Praxis ist deshalb wichtig, bei der Interpretation von Testleistungen, neben den Intelligenzleistungen auch diese Eigenschaften zu berücksichtigen.

2.2. Zwei-Faktoren-Theorie (nach Spearman, 1904)

Spearman stellte diese Theorie auf Grund von Interkorrelationsstudien von kognitiven Leistungsvariablen, Schulleistungen und experimentell gewonnenen Parametern auf. Er nahm an, dass die Leistungen durch zwei Komponenten erklärbar sind. Einmal durch einen *general factor* „g“, welcher sich aus allen intellektuellen Leistungen ableiten lässt und zusätzlich durch viele *spezifische Faktoren* „s“, welche jeweils nur einer Funktion zuzuordnen sind (Dorsch u. Friedrich, 2009; Süß, 2003).

2.3. Multiple Faktorentheorie (nach Thurston, 1938)

Thurston entwickelte die von Spearman aufgestellte Theorie weiter. Er hatte beobachtet, dass immer noch ein beachtlicher Rest an gemeinsamer Varianz, welcher nicht auf die *spezifischen Faktoren* zurückzuführen ist, nach der Extraktion des *general factors* übrig blieb. So extrahierte Thurston sieben weitere Faktoren: Faktor S (*space*) dieser stellt den Faktor der räumlichen Vorstellung bei zwei und dreidimensionalen Objekten dar, Faktor V (*verbal comprehension*), welcher die Fähigkeit sprachliche Bedeutung zu erfassen widerspiegelt, Faktor W (*word fluency*) stellt den Faktor der Wortflüssigkeit dar, Faktor N (*number*) repräsentiert die Gewandtheit (Schnelligkeit und Präzision) bei der Ausführung einfacher Rechenoperationen, den Faktor M (*memory*), der das Kurzzeitgedächtnis verkörpert, Faktor P (*perceptual speed*), die Fähigkeit visuelle Details schnell und genau zu erfassen und den Faktor R (*reasoning*), den allgemeinen Denkfaktor, der Fähigkeit zum induktiven und deduktiven Denken. Dieses Modell von Thurston mit ursprünglich sieben unabhängigen Fähigkeiten (Primary Mental Abilities) wurde später zu einem hierarchischen Modell modifiziert, da die Faktoren nicht als voneinander unabhängig betrachtet werden konnten (Dorsch u. Friedrich, 2009; Süß, 2003).

2.4. Theorie der fluiden und kristallinen Intelligenz (nach Horn und Cattell, 1966)

Diese Theorie zählt zu den einflussreichsten der Intelligenzforschung, welche auf einer Faktorenanalyse höherer Ordnung beruht. Cattell (1987) beschreibt in seiner Investmenttheorie das Verhältnis zwei generell voneinander unabhängiger Faktoren zueinander, der *fluiden* und *kristallinen Intelligenz*.

Fluide Intelligenz ist relativ frei von Bildungseinflüssen, genetisch bedingt, d.h. von Geburt an angelegt und betrifft hauptsächlich die Basisprozesse des Denkens. Sie ist die

Fähigkeit, sich neuen Situationen anzupassen und neuartige Probleme zu lösen, ohne dass dafür erlerntes Wissen benötigt wird. Die fluide Intelligenz ist daher eher kulturunabhängig. Sie wird mit induktiven und deduktiven Aufgabenstellungen erfasst, welche das Erkennen und Anwenden von Regeln erfordern.

Kristalline Intelligenz ist die Fähigkeit, welche durch Lernerfahrung und Faktenwissen erworben wurde, sie ist dem zur Folge ein Indikator für erworbenes Wissen, welches die Person bei der Lösung von Problemen anwendet. Die kristalline Intelligenz ist eher abhängig von kulturellen und gesellschaftlichen Einflüssen. Sie wird z.B. mit Aufgaben zum Wortschatz oder Allgemeinwissen erfasst (Süß, 2003).

2.5. Three Stratum Theorie (nach Carroll, 1993)

Die Theorie von Carroll kann als Erweiterung der bereits erwähnten Modelle gesehen werden, welche einen Rahmen für die Anordnung der meisten bisher erforschten Fähigkeiten der Intelligenz bildet. Die Three Stratum Theorie ist ein auf drei Hierarchieebenen aufgebautes Strukturmodell. Auf der obersten Ebene ist die allgemeine Intelligenz, die durch komplexe kognitive Prozesse höherer Ordnung bestimmt wird, im Sinne des Generalfaktors von Spearman zu finden. Auf der zweiten Ebene sind acht Faktoren angeordnet: „*Fluid Intelligence, Crystallized Intelligence, General Memory and Learning, Broad Visual Perception, Broad Auditory Perception, Broad Retrieval Ability, Broad Cognitive Speediness und Processing Speed*“ (Süß, 2003; S. 220). Auf der dritten Ebene befindet sich eine Vielzahl von spezifischen Fähigkeiten, welche die acht Intelligenzfähigkeiten beeinflussen.

2.6. CHC-Theorie (nach Mc Grew, 2005)

Das Cattell-Horn-Carroll Modell (CHC-Modell) ist die aktuellste Theorie kognitive Fähigkeiten zu klassifizieren und bietet bei der Vorhersage von schulischem oder beruflichem Erfolg eine Reihe von Vorteilen. Es ist ein dreistufiges, hierarchisches Modell und integriert das Modell der fluiden und kristallinen Intelligenz (Gf – Gc) von Cattell sowie die Three Stratum Theorie von Carroll, da sie gleichfalls von einer hierarchischen Struktur der Intelligenz ausgehen und viele Gemeinsamkeiten aufweisen, bis auf die Frage, ob ein Generalfaktor der Intelligenz anzunehmen ist. Auf der untersten Ebene (Stratum I) des Modells sind spezifische Fähigkeitsmerkmale (Primärfaktoren) angesiedelt, die sich durch breiter angelegte Sekundärfaktoren auf der zweiten Ebene (Stratum II) erklären lassen. Sie weisen aufgrund ihrer inhaltlichen Breite

ein höheres Abstraktionsniveau auf. Zu ihnen zählen die *fluide Intelligenz* (Gf), *kristalline Intelligenz* (Gc), *visuelle Verarbeitung* (Gv), *quantitatives Denken* (Gq), *Kurzzeitgedächtnis* (Gstm), *Langzeitgedächtnis* (Gltm), *Verarbeitungsgeschwindigkeit* (Gs) und *Entscheidungssicherheit* (Gd). Die Zusammenhänge zwischen diesen können wiederum durch einen Generalfaktor der Intelligenz (G), welcher sich auf der obersten Ebene (Stratum III) befindet, erklärt werden.

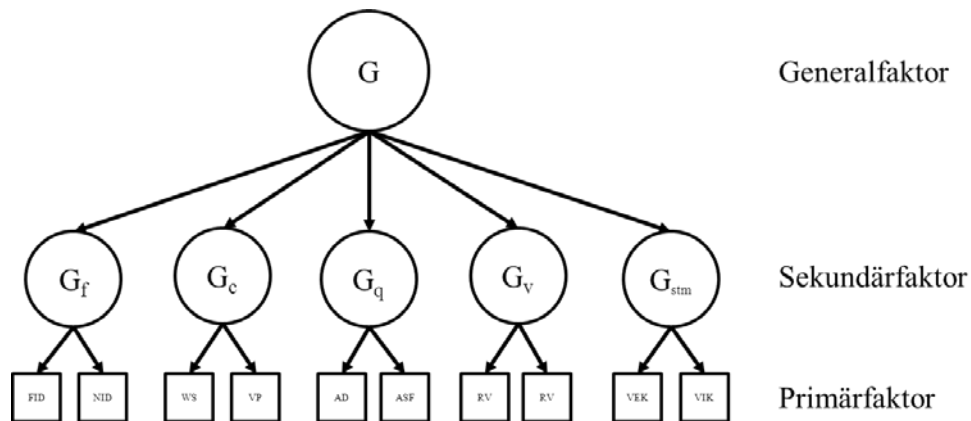


Abbildung 1: Struktur des CHC-Modells - adaptiert nach Arendasy et al. (2012)

Es konnte gezeigt werden, dass sich breiter und abstrakter angelegte Kriterien, wie z.B. der Schulerfolg besser aus dem Generalfaktor vorhersagen lassen als aus den Primärfaktoren der untersten Schicht. Ganz allgemein lässt sich aus dem Modell ableiten, dass die Prädiktoren, umso breiter und abstrakter angelegt sein sollten, je breiter und abstrakter die vorherzusagende Kriteriumsvariable formuliert ist (Petermann, 2006; Arendasy, 2012).

Bei der Entwicklung eines Intelligenzverfahrens ist es sowohl für die Diagnostik als auch für die Interpretation besonders wichtig zu wissen, auf welchem Modell der Test aufbauen soll und welche Faktoren der Intelligenz er misst. Im Zusammenhang mit dieser Arbeit soll kein Vergleich der zur Zeit am Markt gängigen Intelligenztests erfolgen, vielmehr wird im weiteren sehr ausführlich auf den AID 2 und die in ihm verankerten Untertests „Angewandtes Rechnen“ und „Funktionen Abstrahieren“ mit einem Vergleich von Testverfahren, die ähnliche Aufgabenstellungen und Messintensionen haben, eingegangen.

3. Der Begriff „Intelligenz“ im AID 2

An Cattell (1987) und seine Investmenttheorie angelehnt, kann der Begriff „Intelligenz“ in Zusammenhang mit dem AID 2 als „die Gesamtheit aller kognitiven Voraussetzungen, die notwendig sind, um Wissen zu erwerben und Handlungskompetenzen zu entwickeln“ (Kubinger, 2009; S. 23) definiert werden. „Kognitionen“ beziehen sich dabei auf „jeden Prozess, durch den das Lebewesen Kenntnis von einem Objekt erhält oder sich seiner Umwelt bewusst wird...: Wahrnehmung, Erkennen, Vorstellen, Urteilen, Gedächtnis, Lernen, Denken, ...Sprache“ (Lexikon der Psychologie – Arnold, Eysenck & Meili, 1980; S. 1086; zitiert nach Kubinger, 2009, S. 23). Der AID 2 prüft ohne Altersdifferenzierung Intelligenz in bestimmten Dimensionen: „*Informationsverarbeitung in der gesellschaftlichen Umwelt*“, „*Informationsverarbeitung neuer Inhalte*“, „*Auffassungskapazität*“ und „*(Re-)Produktionsfähigkeit durch Strukturierung*“. Die meisten Untertests lassen sich deutlich einer der genannten Dimension zuordnen. Intelligenz, wie sie im AID 2 erfasst wird, begründet kein Kompensationsmodell aus dem ein „Intelligenzquotient“ resultiert. Fähigkeitsdefizite in bestimmten Bereichen können nicht durch einen Fähigkeitsüberschuss in anderen Bereichen aufgehoben werden. Der im AID 2 verwendete Begriff zeigt eher ein Defizitmodell auf, was bedeutet, dass die minimale Testleistung eines Kindes oder Jugendlichen am meisten interessiert, so dass, wenn ein bestimmtes Fähigkeitsdefizit diagnostiziert worden ist und es durch andere Teilleistungen, welche ebenfalls mit dem AID 2 gemessen wurden, nicht auszugleichen ist, man sofort Maßnahmen bzw. Interventionen ableiten kann. „Aus diesem Grund wurde bereits für den AID als globales Maß der kognitiven Fähigkeiten einer Testperson die „(untere Grenze der) Intelligenzquantität“ definiert; sie gibt jeweils in Prozenträngen den Anteil der Referenzpopulation an, mit dem andere Personen eine weniger niedrige *schlechteste Untertestleistung* (= der minimale T-Wert in allen (Untertest-) Testkennwerten) erbringen als die betroffene Person.“ (Kubinger, 2000; S. 31). Die „untere Grenze der Intelligenzquantität“ ist als die kognitive „Mindestfähigkeit“, mit der bei einer bestimmten Person gerechnet werden kann, zu interpretieren. Wobei ihre kognitiven Fähigkeiten in den anderen zu messenden Untertest höher liegen können. Die Zweckmäßigkeit dieses globalen Maßes ist allein aus der Kritik am Kompensationsmodell abzuleiten, seine Sinnhaftigkeit wird jedoch an die Voraussetzungen geknüpft, dass allen erfassten Fähigkeiten (allen Untertests)

dieselbe Bedeutung (in Bezug auf alltägliche berufliche wie schulische Leistungsanforderungen) beizumessen ist. Im AID wurde weiters die Schwankungsbreite der Untertestleistungen normiert und zwar der *Range* der „Intelligenz“. Dieser gibt jeweils in Prozenträngen den Anteil der Referenzpopulation an, welcher noch größere Differenzen zwischen bester und schlechtester Untertestleistung [minimalster bzw. maximalster T-Wert in allen (Untertest-) Testkennwerten] aufweist, als die betroffene Testperson. Für die psychologische Diagnostik weitaus wichtiger ist, dass mit dem *Range* der „Intelligenz“ beurteilt werden kann, wie homogen bzw. wie differenziert das mit dem AID 2 erfasste Fähigkeitsspektrum bei einer bestimmten Person ist. Das heißt, wie weit die einzelnen Untertestwerte streuen (Kubinger, 2000).

4. Der AID und seine Revisionen

Die Testbatterie AID (*Adaptives Intelligenz Diagnostikum*, Kubinger & Wurst) erschien erstmals 1985, es folgten 2 weitere Auflagen 1988 und 1991. Die „2. Generation“, die Testbatterie AID 2 als Version 2.1 kam, als Neunormierung zum AID, welche auch inhaltliche Veränderungen enthielt, 2000 auf den Markt. 2009 wurde eine neu geeichte und überarbeitete Auflage samt AID 2-Türkisch vorgestellt, der AID 2, Version 2.2. Zwei Umstände machten diese neue Auflage erforderlich. Erstens die DIN 33430, die nachdrücklich fordert, dass einschlägige Verfahren spätestens alle 8 Jahre auf Aktualität ihrer Eichung überprüft werden müssen und zweitens die Globalisierung im deutschsprachigen Teil Europas.

Im Vergleich zur ersten Auflage (Version 2.1) gab es einige Veränderungen wobei zu erwähnen ist, dass bei den Aufgaben selbst nur minimale Korrekturen vorgenommen wurden. Für eine noch bessere Illustration bei der Anwendung des AID 2 war ein zweites Fallbeispiel im Testhandbuch enthalten. Eine der sprachfreien Instruktionen wurde komplett geändert, da sie für viele Kinder nicht verständlich war. Mit dem AID 2 war es nun auch möglich nicht-kognitive Eigenschaften zu erfassen wie die Polarisierung in den „hilflosen“ und den „offensiv-zielorientierten“ Verhaltenstyp. Neu war weiters, dass „der standardisierte Faktorscore des 1. Faktors der 4-Faktorenlösung des AID 2 als „Primär“-Intelligenzquotient angeboten“ (Kubinger, 2009; S. III) wurde.

4.1. Der AID 2

Die Vorstellung der Testbatterie erfolgt angelehnt an das Manual der Testbatterie Adaptives Intelligenz Diagnostikum aktuelle Version 2.2 (AID 2, Kubinger, 2009). Der AID 2 ist ein mehrdimensionales Verfahren und hat damit den Charakter einer Testbatterie, welche als Individualtestung, im Bereich der Schulpsychologie, Klinischen Psychologie aber auch in der Berufs- und Bildungsberatung Anwendung findet. Er ist inhaltlich am weltweit verbreiteten, bis auf das Jahr 1939 zurückgehenden Testkonzept von David Wechsler orientiert. Thematisch finden sich im AID 2 an den Hamburg-Wechsler-Intelligenztest-IV (HAWIK IV, Petermann & Petermann, 2007) angelehnte Untertests, jedoch mit deutlichen konzeptionellen Veränderungen. Der AID 2 realisiert das in der „modernen“, in der sog. „*Probabilistischen Testtheorie*“ verankerte „adaptive Testen“, welches unter Abschnitt 4.2. noch näher erläutert wird. Der AID 2 beinhaltet mit wenigen Ausnahmen reine *power*-Tests (nähere Erläuterungen dazu siehe Abschnitt 7.3.), welche in zwei Gruppen geteilt werden können. Einerseits Untertests, die den „verbal-akustischen“ Bereich betreffen und zum anderen Aufgabenstellungen, die den „manuell-visuellen“ Bereich betreffen. Er verwendet ein freies Antwortformat (Erläuterungen siehe Abschnitt 8.4.), wobei die Auswertung zweikategoriell („richtig“ oder „falsch“) geschieht, mit Ausnahmen der Untertests: „*Unmittelbares Reproduzieren*“, „*Kodieren und Assoziieren*“ und „*Antizipieren und Kombinieren*“. Ein wesentlicher Unterschied des AID 2 gegenüber Gruppentests ist aber die besondere Vorgabe-, Prüf- und Beobachtungsmöglichkeit sowie Materialbeschaffenheit der einzelnen Untertests. Der AID 2 hat eine „förderungsorientierte“ Diagnostik zum Ziel. Was bedeutet, dass mit dem AID 2, wie bereits unter Abschnitt 3 erwähnt, versucht wird, „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 und Wurst, 2000; S. 13). Der AID 2 versucht dem Trend nachzukommen, dass in der Praxis ein immer größerer Bedarf an Instrumenten besteht, die zur Erfassung von Teilleistungsstörungen geeignet sind. So kommt er auch bei Fragestellungen, welche Lern- oder Schul-„Schwierigkeiten“ (wie Legasthenie/Dyskalkulie), Konzentrationsstörungen und Verhaltensauffälligkeiten betreffen, immer häufiger zum Einsatz. Dem AID 2 liegt der pragmatische Standpunkt zugrunde, ziemlich viele komplexe und basale Fähigkeiten zu erfassen, welche für „intelligentes“ Verhalten

verantwortlich zu sein scheinen. Im Sinne des pragmatischen Ansatzes wird mit dem AID 2

„eigentlich eine eigene, faktorenanalytisch fundierte Intelligenz-„Theorie“ begründet: Wenn auch die resultierenden Faktoren einer Faktorenanalyse niemals die gesamte beinhaltete Varianz wiedergeben und also für die Einzelfalldiagnostik nicht informationserschöpfend sind, überzeugt doch, theoretisch, die vom AID auf den AID 2 weitergegebene Faktorenstruktur.... Jedenfalls erteilt der AID 2 der Generalfaktortheorie im Sinne Charles E. Spearman eine Absage: Ein Intelligenzquotient, definiert als der Durchschnitt aller geprüften Fähigkeiten, ist intelligenztheoretisch nicht mehr vertretbar und förderungsorientiert nicht zielführend.“ (Kubinger, 2000; S. 14).

Im Gegensatz dazu wird im AID 2 in Bezug auf die Leistungen der Testperson in den einzelnen Untertests, die *Profilinterpretation* propagiert. Das heißt, es werden die Leistungshöhen und Leistungstiefen einer Testperson sowie deren Ergebnisse in den einzelnen Untertests zueinander interpretiert. Für den AID 2 konnte eine Einteilung der Untertests in eher förderungsabhängige und eher förderungsunabhängige Untertests nachgewiesen werden (Kubinger und Wurst, 2000). Es folgt nun eine vollständige Auflistung aller Subtests des AID 2, welche dem Manual entnommen wurden.

Tabelle 1: Vollständige Übersicht der Subtests des AID 2 (Kubinger, 2009)

UT 1 Alltagswissen	Der Untertest 1 <i>Alltagswissen</i> soll die Fähigkeit prüfen, sich Sachkenntnisse über Inhalte anzueignen, die in der heutigen Gesellschaft alltäglich sind.
UT 2 Realitätssicherheit	Der Untertest 2 <i>Realitätssicherheit</i> soll prüfen, inwieweit die Wirklichkeit um Dinge des Alltags verstanden wird bzw. kontrolliert werden kann.
UT 3 Angewandtes Rechnen	Der Untertest 3 <i>Angewandtes Rechnen</i> soll weitgehend unabhängig von schulischen Rechenfertigkeiten prüfen, inwieweit die Testperson bei der Problemlösung alltäglicher Aufgabenstellungen durch entsprechende Schlussfolgerungen die passenden Rechenoperationen anzuwenden imstande ist.

UT 4 Soziale und sachliche Folgerichtigkeit	Mit dem Untertest 4 <i>Soziale und Sachliche Folgerichtigkeit</i> soll die Fähigkeit erfasst werden, die Abfolge sozialen Geschehens bzw. alltäglicher Sachgegebenheiten zu verstehen und zu kontrollieren.
UT 5 Unmittelbares Reproduzieren	Der Untertest 5 <i>Unmittelbares Reproduzieren-numerische</i> soll die Kapazität der seriellen Informationsverarbeitung (im verbal-akustischen Bereich) messen.
UT 6 Synonyme Finden	Der Untertest 6 <i>Synonyme Finden</i> soll das elementare Sprachverständnis prüfen, nämlich inwieweit die Testperson die Bedeutung sprachgebundener Begriffe erfasst bzw. über einen Wortschatz verfügt, der solche Begriffe alternativ ausdrücken lässt.
UT 7 Kodieren und Assoziieren	Mit dem Untertest 7 <i>Kodieren und Assoziieren</i> sollen zwei voneinander partiell unabhängige Fähigkeiten erfasst werden: Die Informationsverarbeitungsschnelligkeit und die Fähigkeit zum inzidentellen Lernen.
UT 8 Antizipieren und Kombinieren	Der Untertest 8 <i>Antizipieren und Kombinieren- figural</i> soll schlussfolgerndes Denken in der Hinsicht prüfen, Teile eines (konkreten) Ganzen erkennen und dieses Ganze gestalten zu können.
UT 9 Funktionen Abstrahieren	Mit dem Untertest 9 <i>Funktionen Abstrahieren</i> soll die Fähigkeit erfasst werden, durch Abstraktion zu einer Begriffsbildung zu gelangen.
UT 10 Analysieren und Synthetisieren-abstrakt	Der Untertest 10 <i>Analysieren und Synthetisieren-abstrakt</i> soll die Fähigkeit prüfen, komplexe abstrakte Gestalten durch eine geeignete Strukturierung reproduzieren zu können.
UT 11 Soziales Erfassen und Sachliches Reflektieren	Mit dem Untertest 11 <i>Soziales Erfassen und Sachliches Reflektieren</i> soll geprüft werden, inwieweit die Testperson Sachzusammenhänge der „gesellschaftlichen“ Umwelt begreift bzw. inwieweit sie sozialisiert in dem Sinn ist, dass

	sie über sozial angepasste Verhaltensweisen und gesellschaftliche Bedingungen Bescheid weiß.
5a Unmittelbares Reproduzieren-figural/abstrakt	Der Zusatztest 5a <i>Unmittelbares Reproduzieren-figural/abstrakt</i> soll die Kapazität der seriellen Informationsverarbeitung (im visumotorischen Bereich) messen.
5b Merken und Einprägen	Mit dem Zusatztest 5b <i>Merken und Einprägen</i> soll die Behaltenskapazität erfasst werden, wie sie durch eine einmalige Wiederholung der Reizdarbietung erreichbar ist.
10a Strukturieren-visumotorisch	Der Zusatztest 10a <i>Strukturieren-visumotorisch</i> soll die Fähigkeit erfassen, komplexe (abstrakte) Gestalten in elementare Teilkomponenten zerlegen zu können.

4.2. Adaptives Testen – branched testing

Beim *adaptiven Testen* geht es darum, nicht allen Testpersonen die gleichen Items vorzugeben, sondern nur diejenigen auszuwählen, welche der individuellen Fähigkeit der Testperson entsprechen. Es sollten also nur diejenigen Items vorgegeben werden, welche einen Informationsgewinn liefern, d.h. Items bei denen die Wahrscheinlichkeit gleich hoch ist das Item zu lösen oder auch nicht. Die Itemauswahl muss dem zur Folge dem jeweiligen Leistungsvermögen der Testperson adaptiert, d.h. angepasst sein. Dadurch reduziert sich die Testlänge drastisch. Die Vorteile des adaptiven Testens sind somit die Erhöhung der Testökonomie. Das heißt, es ermöglicht, dass ein Test trotz geringer Testlänge genau misst. Die Leistungsmotivation kann durch diese Art der Vorgabe der Items (sie sind vom Schwierigkeitsgrad her nicht stetig ansteigend) erhöht werden und es müssen die Untertests nicht frühzeitig abgebrochen werden, weil sie der Testperson zu schwer fallen. Nachteile des adaptiven Testens sind, dass bei der Testkonstruktion eine größere Anzahl an Items erstellt werden muss und dass testtheoretische Analysen notwendig sind (Kubinger, 2000, 2006). Adaptiven Tests liegt ein IRT-Modell (auf welches im Abschnitt 5 noch näher eingegangen wird) zugrunde, welches eine sinnvolle Interpretation und interindividuelle Vergleiche von Testwerten erlaubt, unabhängig davon welche Items die Testperson beantwortet hat (Frey, 2006).

Das *branched-testing*, wie es beim AID 2 Anwendung findet, ist eine suboptimale Strategie des adaptiven Testens, bei dem die Items in Gruppen zu einem so genannten Block zusammengefasst werden. Es gibt verschiedene Einstiegsblöcke, bei welchem Block eingestiegen wird, richtet sich nach dem Alter des Kindes bzw. Jugendlichen. Die Verzweigungswege sind beschränkt und von vornherein entsprechend des Leistungsniveaus der Testperson festgelegt. Auf Grund der Itemgruppen sind weniger Entscheidungsschritte notwendig, was zu Folge hat, dass diese Art der Vorgabe auch ohne Computer angewendet werden kann und vom Testleiter ohne technische Hilfsmittel bewältigbar ist (Kubinger, 1989, 2006; Kubinger & Jäger, 2003).

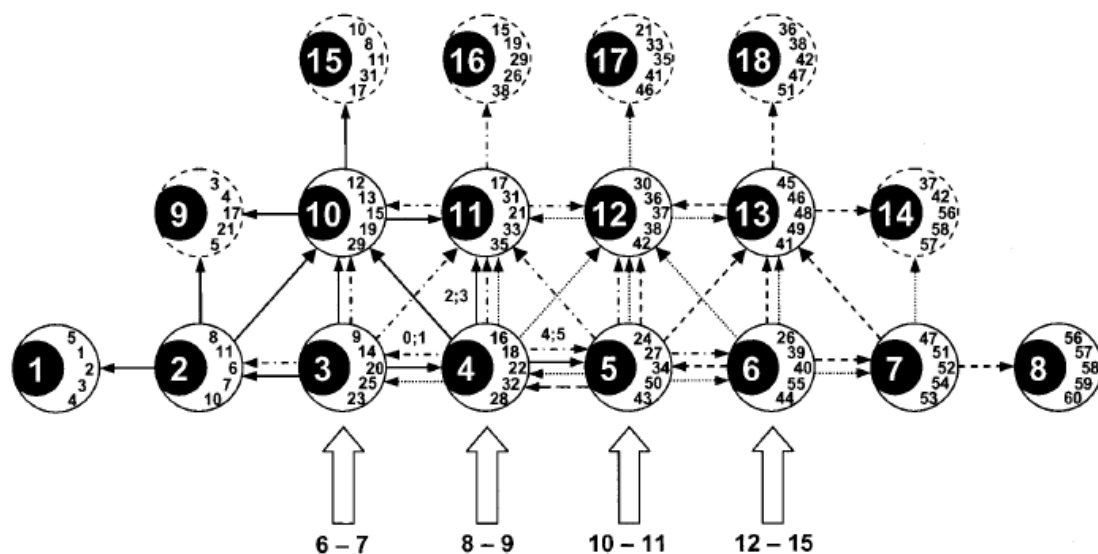


Abbildung 2: Prinzip der Testvorgabe nach dem branched-testing-Design - AID2 Manual (Kubinger, 2009; S. 37)

Der weitere Teil der Diplomarbeit wird sich, wie bereits erwähnt, nur mehr auf die Untertests U3 *Angewandtes Rechnen* und U9 *Funktionen Abstrahieren* beziehen. Dazu wird noch einmal speziell auf diese zwei Untertests eingegangen.

4.3. Angewandtes Rechnen

UT3 Angewandtes Rechnen: Die Testperson muss durch mathematisches Schlussfolgern Textrechnungen lösen, die sie vorgelesen bekommt. Sie muss die passenden Rechenoperationen anwenden, um die Problemstellungen des Alltags zu bewältigen. Ab einem bestimmten Alter werden ihr die Aufgaben nicht nur vorgelesen, sie bekommt zudem die Möglichkeit die Aufgaben in einem Textheft mitzulesen. Grundsätzlich sind

die Rechenoperationen numerisch relativ einfach gestellt, so dass auch bei inhaltlich komplizierteren Problemstellungen die Lösungen der Aufgaben rechnerisch im Kopf erfolgen können. Mit diesem Untertest wird die „rechnerische Fähigkeit unabhängig von schulischen Rechenfertigkeiten gemessen“ (Kubinger & Wurst, 2000; S. 16). Dieser Untertest (folglich nur mehr kurz UT) wird adaptiv vorgegeben. Er besteht aus 18 Blöcken mit je fünf Aufgaben, wobei jeder Testperson standardmäßig drei Blöcke mit insgesamt 15 Aufgaben vorgegeben werden. Die Auswahl der Aufgaben richtet sich wie bereits unter 4.2. ausgeführt nach den Testleistungen der Testperson. Die Verrechnung der Testleistungen der jeweiligen Testperson zu Testwerten erfolgt, auch wenn Items mit einer Zeitbeschränkung versehen wurden, zweikategorial („richtig“ oder „falsch“), das bedeutet, dass lediglich gezählt wird, wie viele Aufgaben von der Testperson (bei Items mit Zeitbeschränkung, in der dafür vorgegebenen Zeit) gelöst wurden. Der Untertest misst nachweislich eindimensional. Er lädt relativ hoch auf den Faktor „Informationsverarbeitung in der gesellschaftlichen Umwelt“ (Kubinger & Wurst, 2000).

Der Untertest „Angewandtes Rechnen“ wird zu den förderabhängigen Untertest gezählt, das heißt, je mehr ein Kind in diesem Bereich gefördert wurde und geübt hat, desto höher ist die Chance in diesem Untertest ein besseres Ergebnis zu erzielen. Da es sich in diesem Zusammenhang um Wissen handelt, was es im Laufe seines Lebens erworben hat, nämlich das Verständnis von Zahlen, das Anwenden von den verschiedenen Rechenoperationen usw., kann es der *Kristallinen Intelligenz* zugeschrieben werden. In Bezug auf das CHC-Modell gliedert sich dieser Untertest daher auf der zweiten Ebene (Stratum II) bei den Sekundärfaktoren ein.

4.3.1. Wie wird die Fähigkeit in anderen Tests abgeprüft?

Zum Vergleich werden der Hamburger-Wechsler-Intelligenztest für Kinder IV (HAWIK-IV, Petermann & Petermann, 2007) und die Kaufmann Assessment Battery for Children, dt. Version (K-ABC, Melcher & Preuß, 2009) herangezogen. Für beide Tests stellte das Testkonzept von Wechsler den theoretischen Hintergrund dar bzw. fanden ihren Ursprung in ihm. Der HAWIK-IV ist ein Individualtest für Kinder und Jugendliche im Alter von 6;0 bis 16;11 Jahren zur bildungsabhängigen, klinischen und diagnostischen Untersuchung von kognitiven Fähigkeiten (Petermann & Petermann, 2007). Die K-ABC ist eine Testreihe für Kinder im Alter von 2;6 bis 12;5 Jahren zur

psychologisch-diagnostischen Prüfung der Intelligenz und des Niveaus der Fertigkeiten sowie zur erziehungs-psychologischen Untersuchung von lernbehinderten oder anderen auffälligen Kindern (Melchers & Preuß, 2009). Als Kriterium für die Gegenüberstellung gilt wie in der Überschrift bereits angedeutet die Ähnlichkeit der Aufgabenstellung sowie ähnliche Messintention. Interessierte Leser seien an dieser Stelle für ausführlichere Ausführungen auf Petermann und Petermann (2009) sowie auf Melchers und Preuß (2009) verwiesen.

HAWIK-IV

Der Untertest *Rechnerisches Denken (RD)* im HAWIK-IV ist ein optionaler Untertest, der die Auswahl der erfassten kognitiven Fähigkeit erweitern und zusätzliche klinisch relevante Informationen und weitere Analysen ermöglichen soll. Er wird dem Indizes *Arbeitsgedächtnis* zugeschrieben, welcher Aufmerksamkeit, Konzentration und Arbeitsgedächtnis erfasst (Petermann & Petermann, 2007).

Inhaltliche Beschreibung: Das Kind soll eine Serie von mündlich vorgegebenen Rechenaufgaben im Kopf lösen.

Beispielitem: In einem Papiersack sind 8 Äpfel. Wie viele Äpfel sind in 3 Papiersäcken?

Gemessene Fähigkeit: Akustische Merkfähigkeit, Arbeitsgedächtnis, Konzentrationsvermögen sowie die Fähigkeit zur Lösung einfacher Rechenaufgaben im Kopf und unter Zeitdruck.

Es gibt je nach Alter der Kinder drei unterschiedliche Startaufgaben. Für jüngere Kinder werden die ersten fünf Aufgaben visuell veranschaulicht und nur die schwierigsten Aufgaben (Aufgabe 17-24) dürfen mitgelesen werden. Es dürfen keine Hilfsmittel wie Papier und Bleistift verwendet werden. Eine richtige Aufgabe wird mit einem Punkt und eine falsche Aufgabe mit 0 Punkten bewertet. Die auditive Merkfähigkeit spielt bei diesem Untertest eine große Rolle. Es gibt eine Umkehrregel als auch die Möglichkeit des Abbrechens, wenn ein Kind vier aufeinander folgende Aufgaben nicht lösen kann. Eine Aufgabe darf nur einmal wiederholt vorgelesen werden (Preusche und Leiss, 2003).

K-ABC

Der Untertest *Rechnen* im K-ABC wird der Fertigkeitsskala zugeordnet (Melchers & Preuß, 2009).

<u>Inhaltliche Beschreibung:</u>	Das Kind muss auf Bildern Zahlen erkennen, zählen und dazugehörige Rechenaufgaben lösen.
<u>Beispielitem:</u>	Eine Familie geht in den Zoo. Hier sind Vater und Mutter. Zähle sie.
<u>Gemessene Fähigkeit:</u>	Grundlegende mathematische Konzepte und Rechenfertigkeiten.

Es gibt je nach Alter des Kindes fünf unterschiedliche Startaufgaben, wobei ältere Kinder, die bei einer späteren Aufgabe beginnen in das Thema Zoo, welches sich durch alle Aufgaben zieht, eingeführt werden müssen. Jede der 37 Aufgaben wird visuell veranschaulicht, wobei nicht allen Kindern alle Aufgaben vorgegeben werden. Dies richtet sich letztendlich nach dem Alter des Kindes. Zur Lösung der Aufgaben dürfen die Kinder, bis auf ihre Finger, keine Hilfsmittel verwenden. Die Aufgabe kann gegebenenfalls auch mehrmals wiederholt werden. Zu beachten sind die besonderen Bewertungskriterien bei einzelnen Aufgaben, da diese aus zwei Teilen a und b bestehen und die Aufgabe nur dann als richtig gelöst bewertet wird, wenn beide Teile richtig beantwortet wurden.

4.4. Funktionen Abstrahieren

UT9 *Funktionen Abstrahieren*: Die Testperson muss die gemeinsame wesentliche Funktion von zwei Dingen, die sie genannt bekommt erkennen und beschreiben. Bei diesem Untertest wurde darauf geachtet, dass „die Lösungen geringe Ansprüche an den Wortschatz“ (Kubinger & Wurst; S. 18) stellen. Mit diesem Untertest wird die Fähigkeit gemessen, durch logisch-schlussfolgerndes Denken Funktionen zu abstrahieren und diese auch sprachlich ausdrücken zu können. Dieser UT wird adaptiv vorgegeben. Er besteht ebenfalls aus 18 Blöcken mit je fünf Aufgaben. Die Vorgabe erfolgt analog dem UT „Angewandtes Rechnen“. Die Verrechnung der Testleistungen der jeweiligen Testperson zu Testwerten erfolgt ohne Zeitbegrenzung zweikategoriell. Der UT misst nachweislich eindimensional, was auch hier als gegeben angenommen wurde. Er lädt ebenfalls relativ hoch auf den Faktor „Informationsverarbeitung in der gesellschaftlichen Umwelt“ (Kubinger & Wurst, 2000).

Der Untertest „Funktionen Abstrahieren“ gehört gleichfalls zu den förderabhängigen Untertest, was auch hier bedeutet, umso mehr ein Kind gefördert wurde, desto höher ist

die Möglichkeit einen höheren Testwert zu erzielen. Auch dieser Untertest gliedert sich in Bezug auf das CHC-Modell auf der zweiten Ebene bei den Sekundärfaktoren ein.

4.4.1. Wie wird die Fähigkeit in anderen Tests abgeprüft?

Diese Fähigkeit wird im K-ABC nicht so isoliert gemessen, so dass zum Vergleich nur der HAWIK-IV herangezogen wurde.

HAWIK-IV

Der Untertest *Gemeinsamkeiten finden (GF)* im HAWIK-IV ist einer von drei Kerntests, aus denen der Indizes *Sprachverständnis* gebildet wird. Er erfasst die sprachliche Begriffsbildung, das sprachliche Schlussfolgern und das erworbene Wissen. (Petermann & Petermann, 2007)

Inhaltliche Beschreibung: Das Kind soll das Gemeinsame von mündlich vorgegebenen Begriffspaaren benennen oder beschreiben. Die Begriffe beziehen sich auf Konzepte oder Gegenstände des Alltags.

Beispielitem: Was ist das Gemeinsame bei Klavier und Gitarre?

Gemessene Fähigkeit: Das Erkennen von unmittelbaren, funktionalen oder abstrakten Beziehungen zwischen zwei Begriffen auf der Grundlage von Analogiebildung, verbales Schlussfolgern, sprachliche Konzeptbildung und Denken in abstrakten logischen Kategorien.

Es gibt je nach Alter der Kinder drei unterschiedliche Startaufgaben. Die ersten beiden Aufgaben werden unter Zuhilfenahme der Bewertungskriterien und der Antwortbeispiele mit 1 oder 0 Punkten bewertet, die Aufgaben 3 bis 24 werden entweder mit 2 Punkten, 1 Punkt oder 0 Punkten bewertet. Es gibt eine Umkehrregel wie auch die Möglichkeit des Abbrechens, wenn ein Kind fünf aufeinander folgende Aufgaben nicht lösen kann. In Bezug auf die Hilfestellung ist zu erwähnen, dass es dem Testleiter beim HAWIK-IV nicht erlaubt ist, das Kind mehr als einmal darauf hinzuweisen, dass „das Gemeinsame an den beiden Begriffen“ gesucht ist.

5. Das Dichotom logistische Modell von Rasch

Das Rasch-Modell beruht auf der Grundlage der *probabilistischen Testtheorie* auch Item-Response-Theorie (kurz IRT genannt). Dabei geht es um die stochastische Beziehung zwischen dem Antwortverhalten einer Person und der latenten (nicht beobachtbaren) Variable. Wenn man adaptiv testen möchte, ist es unumgänglich diesen Test nach den Methoden der IRT zu konstruieren (Frey, 2006). Auch der AID wurde mit deren Hilfe entwickelt. Mit der Item-Response-Theorie ist eine Überprüfung des Gütekriteriums *Skalierung* möglich. Laut Kubinger (2006) ist das Gütekriterium *Skalierung* dann erfüllt, „wenn die laut Verrechnungsvorschrift resultierenden Testwerte die empirische Verhaltensrelation adäquat abbilden“ (S. 79). Es geht also darum, ob der Verrechnungsmodus der Testleistungen (beim AID das Zusammenzählen der richtig gelösten Aufgaben einer bestimmten Testperson) zu Testwerten empirisch gerechtfertigt ist. Letztendlich geht es um die *Verrechnungsfairness* eines Tests, was bei Verfahren, die der *klassischen Testtheorie* entstammen, nicht gegeben ist (nähere Erläuterungen dazu siehe Kubinger, 2006). Wichtig für einen Test (bzw. einen Untertest im AID) ist, dass nur eine Fähigkeitsdimension erfasst wird und somit der Verrechnungsmodus im Sinne des Gütekriteriums „Skalierung“ fair ist.

In der Literatur werden je nach Autor verschiedene Formen der dem Rasch-Modell zu Grunde liegenden Gleichung verwendet, die sich jedoch alle auf die von Rasch vorgeschlagene Form zurückführen lassen. Man findet einerseits unterschiedliche Benennungen der Variablen und andererseits drei verschiedene Varianten der Grundgleichung, wobei im Zusammenhang mit dieser Arbeit nur auf die erste Form der Grundgleichung, dem „1-PL Modell (ein-(item)parametrig-logistisches Modell)“ (Kubinger, 2006; S. 101) oder Rasch-Modell eingegangen wird. Zur Verdeutlichung dieser Grundgleichung stelle man sich nun vor, dass eine einzige Person v mit der Fähigkeit ξ_v ein einziges dichotomes (zweikategoriell) Item i mit der Schwierigkeit σ_i bearbeitet. A_{vi} kennzeichnet das Antwortverhalten der Person v beim Item i . Mit ξ_i ist der Ausprägungsgrad der zu messenden latenten Merkmalsdimension einer Person v gemeint. In Bezug auf die vorliegende Arbeit ist ξ_v ein reeller Wert für die nicht beobachtbare eindimensionale Dimension z.B. im UT3 die Fähigkeit durch schlussfolgerndes Denken, die passenden Rechenoperationen anzuwenden und im UT9 die Fähigkeit durch Abstraktion zu einer Begriffsbildung zu kommen, welche direkt mit den entsprechenden Fähigkeiten einhergeht. Je größer das Ausmaß dieser

entsprechenden Fähigkeiten, desto größer ξ_v . ξ_v wird als *Fähigkeitsparameter* der jeweiligen Person bezeichnet und analog dazu σ_i als *Schwierigkeitsparameter* des Items i , welcher ebenso einem direkten Verhältnis folgt, dass σ_i umso größer wird, je weniger Testpersonen die Aufgabe lösen. Der Hauptgedanke der Grundgleichung lässt sich mit dem folgenden Satz beschreiben: „Je fähiger eine Person v und je leichter ein Item i , desto größer ist die Wahrscheinlichkeit P einer „richtigen“ Lösung ($A_{vi} = 1$).“ (Brandstätter, 2001; S. 24).

Modellgleichung nach Kubinger (1989):

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

Das Rasch-Modell beschreibt damit die Wahrscheinlichkeit, dass eine Testperson v Item i löst, in Abhängigkeit von der Fähigkeit der Testperson (Personenparameter ξ_v) und der Schwierigkeit des Items (Itemparameter σ_i) (Kubinger, 2006).

Das Rasch-Modell geht weiters davon aus, dass eine Person mit einem Fähigkeitsgrad $\xi_v < x$ eine richtige und eine Person mit $\xi_v > x$ eine falsche Antwort geben könnte, das aber mit unterschiedlicher Wahrscheinlichkeit. Daraus ergibt sich die *probabilistische Itemcharakteristik*. Die Item Charakteristik Kurve (kurz ICC) eines Items stellt die Lösungswahrscheinlichkeit als Funktion der latenten Fähigkeitsdimension ξ graphisch dar.

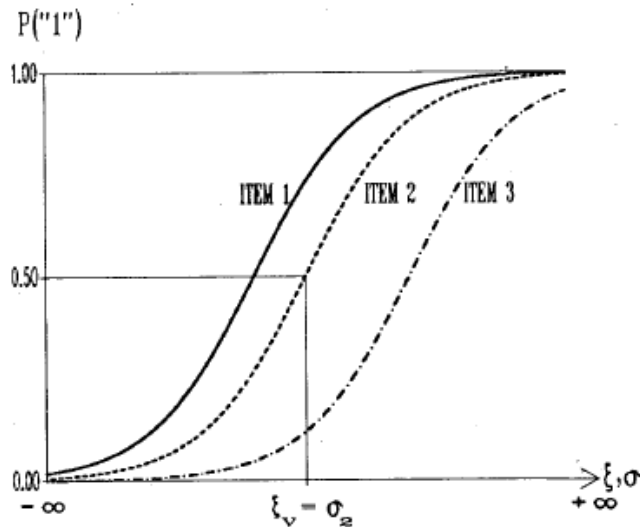


Abbildung 3: Itemcharakteristikkurve (ICC) nach dem dichotom logistischen Modell von Rasch für drei verschiedene Itemschwierigkeiten (Kubinger, 1989)

Eigenschaften des Modells:

Von insgesamt fünf Eigenschaften des Modells, welche alle auf Rasch (1960) zurückgehen, werden exemplarisch drei vorgestellt.

- Homogenität der Aufgaben

Diese Forderung impliziert die Eigenschaft, dass zur Bearbeitung aller Items eine einzige Eigenschaft benötigt wird (Brandstätter, 2001). Der Test misst dem zu Folge eindimensional.

- Spezifische Objektivität

Damit ist die Generalisierbarkeit von Aussagen gemeint. Der Fähigkeitsparameter der jeweiligen Person und der Schwierigkeitsparameter der Items werden separiert betrachtet. Das Rasch-Modell erfüllt den Anspruch „spezifisch objektive Vergleiche“ zu ermöglichen (Kubinger, 1989). Das bedeutet, dass der Unterschied in den Fähigkeiten zwischen zwei Personen unabhängig davon bestimmt werden kann, welche Items des Tests für die jeweilige Person herangezogen wurden. Daraus ergibt sich ein noch weitaus wichtigerer Schluss, dass der Vergleich zweier Items bezüglich ihrer Schwierigkeit unabhängig davon ist, welche Stichprobe dafür verwendet wurde. Was bedeutet, dass die Schätzungen der Itemparameter stichprobenunabhängig sind.

- Lokal stochastische Unabhängigkeit und Summenwerte als erschöpfende Statistik

Ob eine Testperson ein Item löst oder nicht hängt, wenn man den Zufall außen vor lässt, nur von ihrer Fähigkeit und von der Schwierigkeit des Items ab und nicht davon, welche anderen Items sie schon gelöst hat oder noch lösen wird. (Kubinger 1989, 2006; Brandstätter 2001). Dass beim Bearbeiten eines Tests keine Lern-, Ermüdungs-, Positions- oder sonstigen Effekte auftreten, wird ebenfalls vom Modell postuliert. Wenn das Modell gilt, messen alle Items auf einer Dimension, was bedeutet, dass keine zusätzlichen Faktoren die Höhe des Rohwertes beeinflussen. Wenn dem zu Folge zwei Personen den gleichen Rohwert haben, kann ihnen auf Grund der Itemcharakteristik der gleiche Fähigkeitsparameter ξ zugeordnet werden. Damit genügt allein der Rohwert, um den Fähigkeitsparameter einer Person zu bestimmen, unabhängig welche Items von ihr gelöst wurden. Der Rohwert stellt somit eine erschöpfende Statistik für den Fähigkeitsparameter ξ dar (Fischer, 1989).

III. Empirischer Teil

6. Hintergrund und Ziel der Untersuchung

Der Hintergrund dieser Untersuchung ist, wie bereits eingangs erwähnt, dass psychologische Tests grundsätzlich in gewissen Abständen einer Revision bedürfen. Der Abstand ist in der DIN 33430 festgelegt und beträgt für psychologische Verfahren 8 Jahre. Das bedeutet, dass auch der AID 2 solch einer Revision unterzogen werden musste, denn dieses Verfahren wird häufig in der Schulpsychologie und Bildungsberatung eingesetzt und soll dem zur Folge natürlich jenes Wissen überprüfen, das in der heutigen Zeit relevant ist.

Das Ziel der Untersuchung war nun für den Untertest „Angewandtes Rechnen“ neue Items zu entwickeln. Anlass dafür waren unter anderem Rückmeldungen von praktisch tätigen PsychologInnen, dass teilweise die Itemzusammenstellungen in den verschiedenen Aufgabengruppen nicht optimal waren. Teilweise waren in den Aufgabengruppen Items enthalten, welche die Kinder auf Grund ihres Alters bzw. ihrer Schulstufe und dem damit zusammenhängenden bereits erlernten Schulstoff nicht lösen konnten. Wie z.B. ein Item aus dem Block 3, welches für Kinder von 6 -7 Jahren gedacht ist, bei dem sie eine Aufgabe zu lösen hatten, welche sich im Zahlenraum 1000 bewegt, obwohl ein sehr großer Prozentsatz der Kinder in diesem Alter allerhöchstens im Zahlenraum 100 rechnet. Zum anderen ging es auch um eine inhaltliche aktuelle „Einkleidung“ der Aufgaben. Für den Untertest „Funktionen Abstrahieren“ war es notwendig, Worte, welche nicht mehr so aktuell im Gebrauch sind zu ersetzen und in manchen Aufgabengruppen das Schwierigkeitsniveau zu erhöhen. Zu den zu ersetzenden Wörtern gehören z.B. „Schreibmaschine“ oder „Weinpresse“, diese sind im heutigen Sprachgebrauch sowie in der aktuellen Kinder- und Jugendliteratur kaum noch zu finden.

7. UT 3 Angewandtes Rechnen

7.1. Theoretischer Hintergrund

Wenn man die Entwicklung mathematischer Kompetenzen betrachtet, ist festzustellen, dass das Rechnen nicht erst mit dem Eintritt in die Schule beginnt. Antell und Keating

(1983, zit. nach Jacobs & Petermann, 2007; S. 24) „konnten belegen, dass Kinder bereits in der ersten Lebenswoche dazu in der Lage sind, zwischen verschiedenen kleinen Mengen (zwei oder drei Objekte oder Punkte) zu unterscheiden“. Grundsätzlich erwerben Kinder im Alltag schon sehr früh Kenntnisse und Fertigkeiten im Umgang mit Zahlen. Siegler und Robinson (1982, zit. nach Siegler, 1991, zit. nach Oerter & Dreher 2002) unterscheiden bei der Untersuchung von vier- und fünfjähriger Kindern, denen in Geschichten eingekleidete einfache Additionsaufgaben vorgelegt wurden, vier Strategien: *Abrufen aus dem Gedächtnis*, *Zahlendarstellen durch Finger*, *Zählen* und *Fingerzählen*. Die Kinder können, sofern sie alle vier Strategien beherrschen, zwischen ihnen wählen und die jeweils für sie Geeignetste aussuchen. Es kommt auch vor, dass sie mehrere dieser Strategien anwenden.

Die Rechenfertigkeiten, welche die Kinder in den ersten vier Schuljahren erlernen, sind mit enormen Fortschritten verbunden. Eines der Hauptanliegen ist das Erlernen der Grundrechenarten. Hierbei wird mit sehr einfachen Aufgaben, der Addition und Subtraktion wie „5+2“ oder „9-5“ begonnen, die mit der Zeit immer komplexer werden. Zu Beginn beschäftigen sich die Kinder wie erwähnt mit Addition und Subtraktion und später mit Multiplikation und Division. Der Zahlenraum, mit dem die Kinder rechnen, wird zunehmend größer, wobei das Verständnis des Zehnersystems einen wichtigen Meilenstein darstellt. Der zeitliche Erwerb der Rechenfertigkeiten wird im Großen und Ganzen durch Curricula und Lehrpläne festgelegt. Dabei sind aber auch Unterschiede zwischen den Kindern zu berücksichtigen wie z.B. im Leistungsniveau als auch im Leistungsverlauf. Laut Helmke (1997, zit. nach Grube, 2005) scheinen kognitive Determinanten für den Erwerb der Rechenkompetenzen und der Rechenleistung von vorrangiger Bedeutung zu sein. Kognitive Strukturen und Prozesse, die dem Rechnen zu Grunde liegen, können aus der Perspektive von Theorien der Informationsverarbeitung betrachtet werden. Das im Folgenden vorgestellte Rahmenmodell findet sich hauptsächlich in der Gedächtnispsychologie und in der Gedächtnisentwicklung (Schneider und Pressley, 1997, zit. nach Grube, 2005), „lässt sich jedoch auch auf das Problemlösen beim Rechnen anwenden“ (Siegler, 2001; S. 422 zit. nach Grube, 2005; S. 106). Die Informationsverarbeitung wird in diesem Rahmenmodell auf der Basis a) des *Arbeitsgedächtnissystems*, b) der *Nutzung von Strategien*, c) der *Wissensbasis* und d) der *Metakognitionen* beschrieben.

Arbeitsgedächtnis

Das Arbeitsgedächtnis macht es den Menschen möglich, Informationen temporär zu speichern und zu verarbeiten. Wenn ein Kind eine Rechenaufgabe mündlich lösen soll, muss es diese Aufgabe zunächst im Arbeitsgedächtnis behalten, weiters muss es eine Lösungsprozedur planen und ausführen. Während dessen müssen notwendige Aspekte des Durchführungsplans als auch der aktuelle Stand der Verarbeitung präsent gehalten werden. Unter Umständen müssen auch Zwischenergebnisse, bis sie weiter verarbeitet werden, gespeichert werden. Die Belastung des Arbeitsgedächtnisses ist für einen Rechenanfänger enorm. Auftretende Fehler können dadurch erklärt werden, dass die Anforderungen an das Arbeitsgedächtnis beim Lösen einer Rechenaufgabe dessen Kapazität übersteigt (Grube, 2005).

Im Arbeitsgedächtnis werden nach Baddeley (1986, zit. nach Grube, 2006) drei Komponenten unterschieden, welche zusammen die Anforderungen der temporären Speicherung und Verarbeitung erfüllen: die *zentrale Exekutive* und deren Hilffsysteme, der *visuell-räumliche Skizzenblock* und die *phonologische Schleife*. Um die Frage zu beantworten, welche Komponenten des Arbeitsgedächtnisses beim Rechnen beteiligt sind, kam man auf der Basis verschiedener Studien zu der Schlussfolgerung, „dass – wenn erwachsene Probanden komplexere Rechenaufgaben im Kopf lösen sollen – die zentrale Exekutive die Gesamtkoordination der zur Lösung der Aufgabe notwendigen Prozesse unterstützt und die phonologische Schleife dem Behalten der Aufgabeninformation dient“ (De Rammelaere, 2002; Fürst & Hitsch, 2000; Seitz & Schumann-Hengsteler, 2000; zit. nach Grube, 2005; S. 109). In den Studien haben sich wenige empirische Belege gefunden, dass der visuell-räumliche Skizzenblock beim Lösen von komplexen Rechenaufgaben beteiligt ist (De Rammelaere, 2002; zit. nach Grube, 2005).

Nach Gathercole (1998, zit. nach Grube, 2005) nimmt die funktionale Kapazität des Arbeitsgedächtnisses im Grundschulalter zu. Dem zur Folge können ältere Kinder mehr Informationen temporär speichern als jüngere Kinder und Rechenaufgaben immer besser lösen, von einfachen Rechnungen angefangen bis hin zu immer komplexeren Aufgabenstellungen.

Strategienutzung

Während den Kindern die Grundrechenarten vermittelt werden, wird ihr konzeptuelles Verständnis für arithmetische Operationen geweckt und Strategien gelernt und gefestigt, die effizient zu einer Lösung von Rechenaufgaben führen. Erste Lösungsstrategien erwerben die Kinder wie eingangs erwähnt, bereits im Vorschulalter im informellen Kontext. Nach Carpenter und Moser (1983, zit. nach Grube, 2005) lassen sich zur Addition im Zahlenraum bis 20 folgende Strategiegruppen unterscheiden:

a) *Repräsentieren der vorgegebenen Zahlensymbole als Mengen*, das Kind repräsentiert, wenn es z.B. $2 + 5$ addieren soll, jede Zahl einzeln als Menge anhand von konkreten Objekten und zählt anschließend alle Elemente dieser Gesamtmenge.

b) *verbale Zählstrategien*, um dieselbe Rechenaufgabe mithilfe dieser Strategie zu lösen, kann das Kind die Zahlenreihe bis 2 aufsagen („1-2“) und im Anschluss fünf Schritte weiterzählen („3-4-5-6-7“). Verwendet das Kind bei dieser Strategie die Finger, so dient dies nur zur Kontrolle der Zähl Schritte und nicht zum Repräsentieren der Zahl als Menge. Kinder im ersten Schuljahr sind bereits in der Lage das weiterzählen bei einem der beiden Summanden zu beginnen, ohne bei Eins zu starten. Sie zählen dann „2/3-4-5-6-7“ (*counting-on from first*; Carpenter & Moser, 1984; zit. nach Grube, 2005). Oder sie verwenden eine weitere verbale Zählstrategie, die *min-Strategie* (*counting-on from larger*; Carpenter & Moser, 1984; zit. nach Grube, 2005). Das Kind erkennt hierbei den größeren der beiden Summanden und startet den Zählvorgang bei diesem („5/6-7“), wodurch die Lösung schneller gelingt und weniger fehleranfällig ist.

c) Der *Wissensabruf*, bei dieser Strategie ist die Lösung der Aufgabe gespeichert und bekannt und sie kann sofort abgerufen werden ohne weitere Strategien anzuwenden. „Nach Ashcraft und Fierman (1982, zit. nach Grube, 2005; S. 113) erfolgt der Übergang zur überwiegenden Nutzung des Wissensabrufs etwa im dritten Schuljahr“. d) Bei der *Zerlegungsstrategie* wird das Wissen über eine bereits bekannte Aufgabe genutzt. Soll ein Kind die Aufgabe $5 + 6$ lösen, so kann es durch das Zerlegen der Zahlen $5 + 6 = 5 + 5 + 1 = 10 + 1 = 11$ auf das Ergebnis kommen.

Wenn man nun die Strategiewahl aus der Entwicklungsperspektive betrachtet, so kann man die des Addierens durch Repräsentation an Objekten in einem früheren Alter erwarten als z.B. verbale Zählstrategien und die wiederum früher als Wissensabruf und Zerlegung, wobei zu erwähnen ist, dass der Entwicklungsverlauf nicht als Stufenfolge

beschreibbar ist. Bereits benutzte und gut beherrschte Strategien werden nicht abrupt aufgegeben, sobald eine neue Strategie zur Verfügung steht (Grube, 2005).

Die Strategien als auch „die Strategienentwicklung beim einfachen Subtrahieren entspricht in Grundzügen“ (Carpenter & Moser, 1984; zit. nach Grube, 2005; S. 115) derer der einfachen Addition. Auch beim einfachen Multiplizieren und einfachen Dividieren werden ähnliche „Strategien unterschieden, die sich in ihrer Gebundenheit an konkreten Objekten und an sprachlichen Werkzeugen unterscheiden“ (Carr & Hettinger, 2003; zit. nach Grube, 2005; S. 115)

Die zunehmende Nutzung der Wissensabrufstrategien, welche über die Grundschuljahre zunimmt, geht bei allen vier Grundrechenarten mit einer Entlastung des Arbeitsgedächtnisses einher und bringt erhebliche Vorteile beim Lösen komplexerer Rechenaufgaben (Kaye, 1986; Hecht, Torgesen, Wagner & Rashotte, 2001; zit. nach Grube, 2005).

Wissensbasis

Die soeben dargestellte Strategie des Wissensabrufs ist ohne ein grundsätzliches Wissen über Ergebnisse von Aufgaben undenkbar. So ist es wichtig ein Wissen über die Elemente und ihre Merkmale zu haben, wie z.B. das Wissen über Zahlen und deren Bedeutung als Ordnungs- oder Kardinalzahl und ihre Beziehungen zu anderen Zahlen. Weiters ist ein Wissen über die Einsatzmöglichkeiten, die Ziele und die präzise Durchführung von Strategien unabdingbar. Das Wissen über bisherig erfolgreiche Anwendungen betrachten Siegler und Shipley (1995; zit. nach Grube, 2005) als Determinante der Strategiewahl.

Metakognition

Der Sammelbegriff Metakognition wird für eine Reihe von Phänomenen, Erfahrungen und Aktivitäten verwendet, welche „mit dem Wissen und der Kontrolle über eigene kognitive Funktionen (z.B. Lernen, Gedächtnis, Verstehen, Denken) zu tun haben“ (Hasselhorn, 2001; S.541). Unter Metakognition versteht man, wenn eine Person über ihre kognitiven Zustände reflektiert. Sie übernehmen daher die Kommandofunktion der Kontrolle, Regulation und Steuerung während des Rechnens. Ein wesentliches Bestimmungstück der Metakognition ist demnach die „Bewusstheit“. Wenn eine Person in der Lage ist, über ihre eigenen Gedanken zu reflektieren, ist sie in der Lage planvoll und selbstreguliert zu lernen.

7.2. Definition der zu messenden Fähigkeit

Mit diesem Untertest soll die Fähigkeit erfasst werden, durch entsprechende Schlussfolgerungen die passenden Rechenoperationen anzuwenden.

Bei der Definition von einer bestimmten Fähigkeit sind nach Greeno, Moore und Smith (1993; zit. nach Böck, 2010) zwei Konzepte zu unterscheiden: *affordance* und *constraints*. Unter dem ersten versteht man Merkmale, die das zu messende Konstrukt umreißen und als Regeln und generative Elemente bei der Itemkonstruktion mit einfließen. Unter *constraints* versteht man kognitive Operationen, welche im zu messenden Konstrukt nicht berücksichtigt werden und somit bei der Itemkonstruktion ausgeschlossen werden sollten.

Affordance

In Bezug auf das zu messende Konstrukt ist das wesentliche Merkmal die Fähigkeit, aus einer in einem Text eingekleideten Rechenaufgabe heraus, durch entsprechende Schlussfolgerungen, die passenden Rechenoperationen anzuwenden.

Constraints

Ob die Testperson das Item lösen kann oder nicht soll weitgehend unabhängig sein vom grundlegenden Zahlen- und Mengenverständnis der zu testenden Person, ob sie die Rechenoperationen beherrscht sowie ob sie diese nach den mathematischen Regeln anwenden kann. Keinen Einfluss sollte des Weiteren die akustische Merkfähigkeit haben. Aus diesem Grund werden den jüngeren Kindern die Aufgaben in Zukunft in Bildform und den älteren Kindern in Textform vorgelegt, so dass sie die Aufgaben visualisieren bzw. mitlesen können. Aber auch hier sollte die Lösungswahrscheinlichkeit unabhängig von der Visualisierungsfähigkeit und vom Lese- und Textverständnis sein.

7.3. Testart und Aufgabentyp

Der Untertest „Angewandtes Rechnen“ im AID 3 ist ein mathematischer Leistungstest, welcher für die Einzeltestung vorgesehen ist. Ein Leistungstest zeichnet sich dadurch aus, dass von den Testpersonen die Lösung von Aufgaben verlangt wird. Eine wichtige Eigenschaft ist, dass sie von den getesteten Personen in nur einer Richtung, nämlich „nach unten“ verfälscht werden können. Was bedeutet, dass die Testperson sich absichtlich dümmer stellen oder sich keine Mühe geben könnte oder versucht die

Antwort zu erraten. Aber auf keinen Fall kann sie eine höhere Leistung erbringen als die, zu der sie fähig ist (Rost, 2004). Weiters ist der Untertest „Angewandtes Rechnen“ ein power-Test (nähere Erläuterungen, siehe nächster Abschnitt). Bei einigen Aufgaben ist eine Zeitbegrenzung von einer Minute vorgesehen, bei allen Anderen hat die Testperson 30 Sekunden Zeit, um das Item zu lösen (großzügig bemessen). Die Textaufgaben werden den Kindern bzw. Jugendlichen laut vorgelesen, den sechs und siebenjährigen Kindern werden zusätzlich die Aufgaben in Bildform vorgelegt und den älteren Kindern bzw. Jugendlichen wird die Möglichkeit geboten, die Textaufgaben in einem Heft mitzulesen. Die Testpersonen erhalten zudem Papier und Bleistift, so dass sie die Aufgaben nicht zwingend im Kopf lösen müssen. Wenn die Testperson eine Aufgabe mal nicht verstanden hat, kann sie ihr nochmals wörtlich vorgelesen werden, was aber nicht die Regel sein sollte.

Power-Tests

Power-Tests zählen zur Kategorie der Leistungstests, welche keine oder eine großzügig bemessene Zeitbegrenzung aufweisen (Häcker & Stapf, 2009). Bei ihnen zählt nur, ob die Aufgaben richtig oder falsch gelöst wurden, unabhängig von der dafür benötigten Zeit. Da jeder Test aber eine zeitliche Begrenzung haben muss, sind reine power-Tests aus technischen Gründen kaum durchführbar. Die Grenzen sollen aber so bemessen sein, dass der Testperson genügend Zeit bleibt die Aufgabe zu lösen.

7.4. Das Item

Nach Rost (2004) ist ein Item die kleinste Beobachtungseinheit, bei dem zwei Komponenten unterschieden werden, zum einen der Itemstamm und zum anderen das Antwortformat. Der Itemstamm ist beim beschriebenen Untertest eine Textaufgabe, die das Testverhalten der Person zeigt. Das Antwortformat, welches das Testverhalten registrieren soll, ist frei, das bedeutet, dass die Testperson nach freiem Ermessen die gestellte Aufgabe verbal zu beantworten hat, in diesem Fall die Lösung der Rechenaufgabe (Lienert & Raatz, 1998). Die Antwort wird als richtig verrechnet, wenn die Lösung korrekt und in der dafür vorgesehenen Zeit gegeben wurde. Die Verrechnung geschieht dichotom, das heißt, die Aufgabe wird letztendlich als „richtig“ oder „falsch“ verrechnet, selbst wenn die Aufgabe richtig gelöst, aber die dafür vorgesehene Zeit überschritten wurde, gilt die Aufgabe als falsch.

Beispielitem: Itemstamm: „Du sollst einen Gipsbrei im Verhältnis 2 Teile Gips zu 3 Teile Wasser anrühren. Du hast 2 kg Gips und willst alles verwenden. Wie viel Liter Wasser brauchst du?“

Antwortformat: 3 (Zeitbegrenzung: 1 Minute)

7.5. Zielgruppe

Der gesamte AID 3, so auch dieser Untertest, ist für deutschsprachige Kinder und Jugendliche im Alter von 6 Jahren (6;0) bis 15 Jahre und 11 Monate (15;11) vorgesehen.

7.6. Itemkonstruktion

Der Untertest „Angewandtes Rechnen“ ist ein UT, den es bereits im ursprünglichen AID sowie in seinen Revisionen gegeben hat. Der Wunsch nach Überarbeitung dieses Untertests hat seinen Ursprung, wie bereits berichtet, in den Rückmeldungen von praktizierenden Psychologen, die den Test anwenden sowie in der Aufgabenanalyse. Vorweg wird nochmals darauf hingewiesen, dass die Items zum Großteil aus der Version 2.2 übernommen wurden.

Probleme bereiteten, wie eingangs erwähnt, teilweise die Zusammenstellungen der Aufgabengruppen, weil in ihnen manchmal Items enthalten waren, die die Kinder aufgrund ihres Alters und dem somit noch nicht erlernten Schulstoff nicht lösen konnten wie z.B. ein zu hoher Zahlenraum sowie Multiplikations- und Divisionsaufgaben, als auch die Verwendung von Gewichts- und Längenmaßen für sechs- und siebenjährige Kinder. In einigen Aufgabengruppen waren mehrere Items in Bezug auf ihre Rechenoperationen nahezu gleich, die im AID 3 zu ersetzen sind und bei manchen Items war die Rechenaufgabe in einen zu komplizierten und zu langen Text eingekleidet, so dass mit diesen Aufgaben eher Textverständnis abgeprüft wird. Zudem fehlten leichte Items für jüngere Kinder. Für den eigentlichen Itemkonstruktionsprozess wurde ein Online-Diskussionsforum eingerichtet, in dem es für jeden Untertest eine Plattform gab, um die neu entwickelten Items öffentlich (d.h. für alle auf das Online-Forum zugriffsberechtigten DiplomandInnen, ProjektleiterInnen sowie PsychologInnen) einer Diskussion und Bewertung unterziehen zu können. Für die Itemkonstruktion gab es keine expliziten Regeln. Anhand der Analyse der Items aus dem AID 2, der recherchierten Literatur und angepasst an den für die entsprechenden Schulstufen

vorgegebenen Lehrplänen, wurden unter Zuhilfenahme von verschiedenen Mathematikbüchern neue Items entwickelt, auch Anregungen von LehrerInnen sind dabei berücksichtigt worden. Da der UT „Angewandtes Rechnen“ adaptiv nach dem *branched-testing* vorgegeben wird, ist es notwendig mehr Items zu entwickeln, als letztendlich benötigt werden, weil unter Umständen Items auszuschneiden sind, welche nicht dem Rasch-Modell entsprechen. In diesem UT der Rohversion des AID 3 werden rund 60 Items in den bekannten AID-Blöcken enthalten sein plus neun „Zusatzitems“, welche einen breiten Schwierigkeitsbereich abdecken. Die Aufgaben sollten über die Schwierigkeitsbereiche gleichmäßig verteilt sein.

Zunächst wurden, wie bereits erwähnt, unter Berücksichtigung der schulischen Lehrpläne, der Rückmeldungen und Veränderungsvorschlägen von praktizierenden PsychologInnen usw. von allen Zugriffsberechtigten viele Items mit den entsprechenden Lösungen konstruiert, einer der drei Schwierigkeitsstufen (einfach, mittel oder schwierig) zugeordnet und online gestellt. Anschließend wurden die Items von den Forumsbeteiligten ausführlich diskutiert, man erhielt zu den selbst entwickelten Items Feedback und Anregungen, so dass manche Items überarbeitet und erneuten Recherchen unterzogen wurden. Zur Veranschaulichung des komplizierten Konstruktionsprozesses sei exemplarisch ein Item, welches nicht in dem Itempool aufgenommen wurde, dargestellt. Unter der Kategorie *Buchstabierwettbewerb* und der Schwierigkeitsstufe *leicht* wurde folgendes Item gepostet: „Bei einem Buchstabierwettbewerb treten in der Endausscheidung 5 Kinder gegeneinander an. 2 Kinder können das nächste Wort nicht buchstabieren und scheiden aus. Wie viele Kinder sind noch im Rennen?“ Es gab hinsichtlich dieses Items Veränderungsvorschläge in Bezug auf einer eventuellen Umformulierung... 2 Kinder schaffen ein Wort nicht und scheiden aus. ...“Das Item wurde dem entsprechend umformuliert, aber letztendlich ist dieses Item ausgeschieden worden, da es für die leichte Rechenoperation „5-2“ in einem zu komplizierten Text eingekleidet gewesen und daher sprachlich sehr fordernd ist. Weiteres ist zu beachten, dass die leichten Items zukünftig im AID 3 zeichnerisch unterstützt werden, damit nicht auch Textverständnis und verbale Merkfähigkeit mitüberprüft werden. Jedoch kann diese Aufgabenstellung hinsichtlich des Zeichnens schlecht umgesetzt werden. So mussten einige Items von vornherein ausgeschieden werden, weil die Rechenoperationen in zu viel unnötigem Text eingekleidet waren oder weil viele ähnliche Aufgabenstellungen bei den Items vorkamen, die für ein und denselben Block

vorgesehen sind. Hier und da musste bei einigen neu entworfenen Items nachrecherchiert werden, ob es dieses vielleicht in einer ähnlichen Art und Weise schon gibt. Auch die Einschätzung des Schwierigkeitsgrades der Items wurde diskutiert, teilweise verändert und durch ein „Experten-Rating“ festgelegt. Hierzu ist zu erwähnen, dass die Schwierigkeiten der Items über die Rechenoperationen (Addition, Subtraktion, Multiplikation und Division), sowie deren Anzahl und Kombination in einer Aufgabenstellung und die zusätzliche Verwendung von Längen- und Gewichtsangaben als auch von Prozentrechnungen und der gleichen festgelegt wurde. Der gesamte Konstruktions- und Diskussionsprozess war kompliziert, sehr aufwändig und langwierig, da die einzelnen Items erst dann in den endgültigen Itempool aufgenommen wurden, wenn alle am Konstruktionsprozess beteiligten Personen ihre Zustimmung gaben.

In der Rohversion des AID 3 wurden letztendlich sechs Items aus dem AID 2 (Version 2.2) in ihrem Wortlaut verändert, sechs Items sind entfernt und fünf Items sind in ihrer Blockzuordnung verschoben worden. 16 von den neu konstruierten sind in die diesem Erhebungsinstrument aufgenommen worden, wovon neun „Zusatzitems“ sind. Letztendlich sind 70 Items in die Datenüberprüfung mit einbezogen worden.

Es liegt an den Blockzuordnungen der Items beim *branched-testing*, dass sich die Zahlen rein rechnerisch nicht ganz ausgehen, denn sieben Items sind neu dazugekommen und nur sechs alte sind entfernt worden. Items, welche vorher im Test an drei verschiedenen Stellen zu finden waren, sind nun in der Rohversion nur mehr einmal vorhanden oder es sind als Platzhalter jetzt zwei Positionen frei geworden, weil ein altes Item im AID 2 zweimal vorhanden gewesen ist und im neuen Test aber komplett entfernt wurde.

8. UT 9 Funktionen Abstrahieren

8.1. Theoretischer Hintergrund

Die Sprache hat eine entscheidende Bedeutung für die Entwicklung des *begrifflichen Denkens* (also das Denken in Sätzen und Urteilen), das aus unterschiedlichen Informationsquellen, wie unter anderem der Wahrnehmung, gespeist wird. Die Fähigkeit zur Kategorisierung ist eng mit der Entwicklung des *begrifflichen Denkens*

verbunden. Grundsätzlich liefert die Merkmalsstruktur der Umwelt Informationen darüber, ob Objekte zu einer bestimmten Kategorie gehören. Bei den Kategorien geht Goswami (2001) davon aus, dass sie hierarchisch strukturiert sind und sich zwischen untergeordneten, basalen und übergeordneten Ebenen unterscheiden lassen. Die äußere Struktur von Objekten kennzeichnet die hierarchische Ebene und hilft bei der Kategorisierung auf der basalen Ebene. Auf dieser Ebene haben wir einen sinnlichen Zugang zu den Objekten sowie die Fähigkeit Prototypen zu bilden, um die Dinge zu klassifizieren. So helfen verbale Benennungen schon Zweijährigen konzeptuelle Unterscheidungen vorzunehmen. Kinder lernen durch den Erwerb neuer Wörter immer mehr über die begrifflichen Relationen zwischen Objekten und Klassen von Objekten. Bezeichnungen geben Auskunft über Relationen zwischen Objekten und über ihre Kategorie, ob sie eher über- oder untergeordneter Natur sind. Z.B. der Erwerb des Wortes „Tier“ als Überbegriff für Katzen, Hunde, Hasen usw. macht es Kindern möglich, die Beispiele als Vertreter der gleichen übergeordneten Kategorie zu klassifizieren. Da Kinder noch nicht die Erfahrungen, den Wortschatz und das Wissen über die kulturellen Funktionen der Dinge wie Erwachsene haben, bilden sie Kategorien die breiter oder enger sein können als die Kategorien von Erwachsenen. Gleichwohl sie bei deren Bildung gleich vorgehen, nämlich die Objekt bzw. Dinge mit ähnlichen Formen, Funktionen, Eigenschaften usw. zu Gruppen zusammenfassen (Goswami, 2001).

Weiters kann die Sprache über die entsprechende Kategorie Auskunft geben, denn bestimmte Merkmale unserer Sprache sind verantwortlich für die Organisation der Begriffe. Kinder interpretieren z.B. Substantive als Hinweis auf eine übergeordnete Kategorie und Adjektive dagegen als Hinweise auf eine untergeordnete Kategorie (Goswami, 2001).

Auch die Ursachenerklärung von Alltagsphänomenen und Überzeugungen aus der Umwelt spielt bei Kindern eine Rolle bei der Bildung von Kategorien. Kinder bauen Begriffshierarchien innerhalb einer Kategorie auf, wie z.B. Tiere und Autos und gewinnen zudem auch Einsicht über grundlegende begriffliche Unterschiede, wie die zwischen Belebtem und Unbelebtem. In diesem Zusammenhang sind Gemeinsamkeiten und Unterschiede in ihren inneren und äußeren Merkmalen relevant, wobei die inneren Merkmale hier die entscheidenden sind. Im Zuge der Entwicklung von Kindern sind sie

mit der Zeit in der Lage auch Analogien bei der Bildung und Ausdifferenzierung von Begriffen heranzuziehen (Goswami, 2001).

Nach Neuhaus (1962) ist das Herausheben von Dingen aus dem Wahrnehmungsumfeld sowie das Ablösen und Bezeichnen von Formen, Farben und Qualitäten mit Eigennamen nicht als abstrahieren zu bezeichnen. Erst, wenn die Objekte miteinander verglichen werden, ihre Merkmale herausgearbeitet und in ihrer Wesenheit erkannt werden und sich ein unanschauliches Wissen über die Dinge und deren Eigenschaften entwickelt sowie die gegenständliche Welt kategorial geordnet wird, kann von abstrakt bzw. abstrahieren gesprochen werden.

Abstraktion

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

8.2. Definition der zu messenden Fähigkeit

Mit diesem Untertest soll die Fähigkeit erfasst werden, von zwei Begriffen die zentrale gemeinsame Funktion bzw. deren Verwendungszweck oder ihre zentrale gemeinsame Eigenschaft zu erkennen.

Affordance

Im Zusammenhang des zu messenden Konstrukts, ist das wesentliche Merkmal, die Fähigkeit die gemeinsame zentrale Funktion von zwei Begriffen zu erkennen und andere Gemeinsamkeiten, die diese nicht betrifft als auch die Funktion jedes einzelnen Begriffes auszublenden (Böck, 2010).

Constraints

Auf die Lösungswahrscheinlichkeit eines Items sollten der Wortschatz als auch besondere Kenntnisse in bestimmten Bereichen keinen Einfluss haben. Die erstgenannte Störvariable ist nicht vollkommen auszuschalten, da beim Lösen dieser Items ein gewisser Grundwortschatz vorausgesetzt wird. Die Letztgenannte Störvariable sollte jedoch weitgehend ausgeschaltet werden, indem auf Fremdwörter, selten verwendete oder veraltete Wörter wie auch auf Fremdsprachen und Fachjargons verzichtet wird (Böck, 2010).

8.3. Testart und Aufgabentyp

Der Untertest „Funktionen Abstrahieren“ im AID 3 ist ein sprachlicher Leistungstest, welcher für die Einzeltestung vorgesehen ist. Er ist ein reiner power-Test, da es für die Bearbeitung keine Zeitbegrenzung gibt. Sofern ein Kind bzw. Jugendlicher die Aufgabenstellung nicht versteht oder die Aufgabe falsch beantwortet sind insgesamt die ersten drei Wortpaare zu erklären. Wenn die Testperson bemüht ist Oberbegriffe zu finden, muss sie darauf aufmerksam gemacht werden, dass es darum geht das Gemeinsame zu finden.

8.4. Das Item

Der Itemstamm ist beim beschriebenen Untertest eine Frage, die das Testverhalten der Person zeigt. Das Antwortformat, welches, wie bereits erwähnt, das Testverhalten registrieren soll, ist frei, das bedeutet, dass die Testperson nach freiem Ermessen die gestellte Aufgabe verbal beantwortet (Lienert & Raatz, 1998). Die Antwort selbst kann mehrere richtige Antworten enthalten, wird aber dichotom verrechnet, d.h. wenn die Testperson eine von den als richtig eingestuften Antworten gibt, wird sie als richtig gezählt, wenn dies nicht der Fall sein sollte, wird sie als falsch gewertet. Wichtig ist, dass die Antwort eindeutig auf die Frage zurückzuführen ist, wenn das Testverhalten Rückschlüsse auf die Personeneigenschaft erlauben soll (Rost, 2004).

Beispielitem: Itemstamm: Was ist das Gemeinsame an „Korb und Tasche“

Antwortformat: mit beiden kann man etwas tragen; in beide kann man etwas hineinlegen

Das freie Antwortformat

Wenn es um die Erfassung spontaner Reaktionen, kreativer Leistungen oder um Assoziationen geht, ist ein freies (oder offenes) Antwortformat vorzuziehen. Bei Leistungstests sind sie ein Mittel, die Wahrscheinlichkeit einzuschränken, dass die richtige Antwort erraten wird. Das Alter des Probanden ist bei der Gestaltung eines Tests mit freiem oder gebundenem Antwortformat mitentscheidend. Es kann für Kinder unter Umständen durchaus schwierig sein sich bei gebundenen Antwortformaten frei zu entscheiden, jedoch um einiges einfacher eine freie Antwort selbst zu produzieren. Bei den freien Antwortformaten lassen sich drei Arten unterscheiden, wobei in dieser Arbeit nur auf die, für den AID 2 bzw. 3 relevante eingegangen wird. Diese Art des freien

Antwortformates ist die *formale Vorgabe* für die Produktion des Verhaltens, mit der eine gewisse Standardisierung erreicht werden kann (Rost, 2004). Ein Problem, welches sich aus diesem Antwortformat ergibt, ist die Auswertungsobjektivität, wenn es für eine Aufgabe mehrere richtige Antworten gibt und genau diese Antwort, die von dem Kind genannt wird nicht im Antwortkatalog enthalten sein sollte. So bewerten einige TestleiterInnen die von dem Kind gegebene, aber im Antwortkatalog nicht enthaltene, Antwort als richtig und andere TestleiterInnen als falsch, was letztendlich zu unterschiedlichen Ergebnissen führt.

8.5. Zielgruppe

Siehe unter 7.5. Da der AID 3 in Österreich, Deutschland und der Schweiz Anwendung findet, ist bei der Itemkonstruktion darauf zu achten, dass nur solche Begriffe verwendet werden, die bestimmte Personengruppen aufgrund ihrer Herkunft, ihrem soziokulturellen Status aber auch ihrer Geschlechtszugehörigkeit nicht systematisch benachteiligen (Kubinger, 2006).

8.6. Itemkonstruktion

Der Untertest „Funktionen Abstrahieren“ ist ein UT, den es ebenfalls im ursprünglichen AID sowie in seinen Revisionen gegeben hat. Grundsätzlich wurden die Items zum Großteil aus der Version 2.2 übernommen, es sollten aber Wörter, welche nicht mehr so aktuell im Gebrauch sind, wie z.B. „Schreibmaschine“ oder „Weinpresse“ ausgetauscht, bzw. durch neue ersetzt, manche davon ganz weggelassen werden, wie die Begriffe „Silo“ und „Heugabel“, weil sie Stadt und Landunterschiede wiedergeben oder der Begriff „Rauschgift“, weil dieser von Kindern im Alter von 8-9 Jahren nicht gekannt wird. Weiters war es ein Anliegen neue schwierigere Items zu konstruieren. Das Verzweigungsschema im AID 2 sieht 60 Items vor, wovon auf Grund der adaptiven Testvorgabe nach dem *branched-testing* einige Items doppelt oder dreifach vorhanden sind. Der allgemeine Konstruktionsprozess geschah in gleicher Weise wie beim Untertest „Angewandtes Rechnen“, in dem die neu konstruierten Items mit den dazugehörigen Antwortalternativen in dem bereits erwähnten Online-Diskussionsforum zur öffentlichen Diskussion gepostet wurden. Auch diese Items wurden zu den entsprechenden Schwierigkeitsstufen zugeordnet, wobei zu erwähnen ist, dass es schwer ist, weder anhand der Literaturrechereschwer noch anhand der Itemanalyse, verbindliche Regeln für die verschiedenen Schwierigkeitsstufen festzustellen. Dies

geschah fast ausschließlich durch ein „Experten-Rating“. Jedes einzelne Item wurde im Diskussionsforum sowie in den Workshops von den Forumbeteiligten dahingehend diskutiert und überprüft. Das gleiche gilt für die Antwortalternativen. Auch diese wurden nach ausführlichen Recherchen durch ein „Experten-Rating“ festgelegt. Weiteres war es auch bei diesem Untertest notwendig mehr Items zu entwickeln, da davon auszugehen ist, dass einige Items nicht dem Rasch-Modell entsprechen und dem zur Folge auszuschneiden sind. Bei der Erstellung eines Items ist der erste Schritt, sich ein Stimulus zu überlegen, dafür wurden mögliche Themenbereiche vorgeschlagen wie:

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

Diese könnten Gemeinsamkeiten haben hinsichtlich:

- Kommunikationstechnik
- Körper und Psyche
- Zeitlicher Stadien (z.B. beides jung)
- Grad der Realität (z.B. beides phantastisch)
- Bewegungsart
- Form
- Art der Aufbewahrung
- usw.

Angelehnt an die präsentierte Literatur, den Analysen der alten Items, den Rückmeldungen von praktizierenden Psychologen als auch an den Richtlinien zum Erstellen der Items, wurden dann in Bezug auf die Themenbereiche zwei Stimuli überlegt, die die gleiche grundlegende Funktion erfüllen, die gleichen Eigenschaften oder eine Gemeinsamkeit in Bezug auf die erwähnten Bereiche haben. Dies geschah, indem unter anderem im Familien- und Freundeskreis über mögliche Stimuli diskutiert wurde, ohne direkt auf die Items einzugehen aber auch hinsichtlich der

Antwortmöglichkeiten durch Recherchen in Wörterbüchern, Lexika, diversen Spielen und dergleichen. Selbst das Finden zwei eindeutig zu einander passender Stimuli gestaltete sich als Herausforderung, wie das Beispiel, „Notar – PIN-Code“, die beide der Bestätigung der Identität dienen sollen, was aber doch eher sehr weit hergeholt ist. Ebenfalls musste berücksichtigt werden, dass nicht zu viele Items aus ein und derselben Kategorie stammen oder die Antworten zu den neu konstruierten Items vielleicht schon bereits bei den Items im AID 2 vorkommen. Die Antwortmöglichkeiten mussten so gewählt werden, dass sie die möglichen Antworten der Kinder abdecken, damit die Objektivität gewährleistet ist und verschiedene Testleiter das gleiche als „richtig oder falsch“ bewerten und somit zum gleichen Ergebnis kommen. Denn es wäre fatal, wenn ein Kind eine gemeinsame Funktion, Eigenschaft oder Gemeinsamkeit nennt, die nicht im Antwortkatalog enthalten ist. Das Zusammenstellen des Antwortkataloges stellte sich als schwieriger heraus als gedacht, denn man musste stets die Instruktion, die die Kinder bekommen im Kopf behalten, die lautet: *...du sollst herausfinden, was sie gemeinsam haben oder was man mit beiden tun kann*. Oft hatte man selbst eine Gemeinsamkeit im Kopf, die aber nicht die zentrale war. Die Zusammenstellung des Antwortkataloges erfolgte letztendlich auf Grund von Rückmeldungen der Forumsbeteiligten und das Zusammentragen der Antwortmöglichkeiten in Diskussionsrunden in den einzelnen Workshops. Auch hier sei exemplarisch ein Item zur Veranschaulichung des schwierigen Konstruktionsprozesses dargestellt. Unter der Kategorie *Musik* und der Schwierigkeitsstufe *leicht* wurde folgendes Item gepostet: „Trompete – Tuba“ Antwort: „Es sind beides Blechblasinstrumente.“ Dieses Item ist von der Antwort her, unter Berücksichtigung der Instruktion überhaupt nicht eindeutig, es widerspricht ihr sogar, denn als Antwortmöglichkeiten müssen ebenfalls gelten: „man kann in beide hineinblasen, mit ihnen Musik machen, in einem Orchester spielen“ oder ganz allgemein, „sie sind beide Musikinstrumente“. Weiteres ist es für die Schwierigkeitsstufe *leicht* nicht so optimal, da viele Kinder förderbedingt „Tuba“ nicht kennen.

Sieben Items aus dem AID 2 (Version 2.2) wurden belassen aber sprachlich verändert und neun entfernt. 18 Items sind neu konstruiert worden, wovon neun Items als „Zusatzitems“ fungieren. In die Datenüberprüfung sind 69 Items mit einbezogen worden. Die Blockzuordnung der alten Items wurde beibehalten. Bis auf drei Items sind die Stimuli aller Items Substantive.

9. Methode

9.1. Hypothesen

Die Hypothesen ergeben sich aus den Fragestellungen: Messen beide Untertests des AID eindimensional und ist die Verrechnung der Testleistungen der Kinder und Jugendlichen zu Testwerten, die dann untereinander vergleichbar sein sollen in Bezug auf das Gütekriterium Skalierung fair?

Die Fragestellungen können beantwortet werden, in dem überprüft wird, ob die Untertests „Angewandtes Rechnen“ und „Funktionen Abstrahieren“ raschkonform sind.

Auf Grund der Teilungskriterien, die vom Verfasser durchgeführt wurden, können im Zuge der Modellprüfung auch Aussagen darüber gemacht werden, ob die Untertests Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit oder Muttersprache systematisch benachteiligen.

Für den Untertest „Angewandtes Rechnen“ ergeben sich hiermit eine Haupt- und drei Nebenhypothesen:

H₀-1: Die Items des Untertests „Angewandtes Rechnen“ entsprechen dem Rasch-Modell.

H₁-1: Die Items des Untertests „Angewandtes Rechnen“ entsprechen nicht dem Rasch-Modell.

H₀-2: Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit nicht.

H₁-2: Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit.

H₀-3: Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen keine Kinder bzw. Jugendlichen auf Grund ihrer Muttersprache.

H₁-3: Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen Kinder bzw. Jugendliche auf Grund ihrer Muttersprache.

H₀-4: Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen Kinder bzw. Jugendliche aufgrund ihres Alters nicht.

H₁-4: Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen Kinder bzw. Jugendliche aufgrund ihres Alters.

Für den Untertest „Funktionen Abstrahieren“ ergeben sich hiermit ebenfalls eine Haupt- und drei Nebenhypothesen:

H₀-1: Die Items des Untertest „Funktionen Abstrahieren“ sind raschkonform.

H₁-1: Die Items des Untertests „Funktionen Abstrahieren“ sind nicht raschkonform.

H₀-2: Die verbliebenen Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit nicht.

H₁-2: Die verbliebenen Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit.

H₀-3: Die verbleibenden Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendlichen auf Grund ihrer Muttersprache nicht.

H₁-3: Die verbleibenden Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendlichen auf Grund ihrer Muttersprache.

H₀-4: Die verbleibenden Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendliche aufgrund ihres Alters nicht.

H₁-4: Die verbleibenden Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendliche aufgrund ihres Alters.

9.2. Untersuchungsplan

Im Sommer 2009 haben die an dem Projekt mitarbeitenden DiplomandInnen ihr Thema und damit ihre zu bearbeitenden Untertests übernommen. Jeder hat sich für seinen Teil der Arbeit in die entsprechende Fachliteratur eingelesen und kurz darauf wurde mit der Itemkonstruktion begonnen. Hierfür wurde, wie unter 7.6. dargestellt, ein Diskussionsforum eingerichtet, bei dem jeder neue Items, aber auch Feedback und Anregungen nicht nur für seine Untertest, sondern bei allen zu bearbeitenden Untertests einfließen lassen konnte, so dass mit der Zeit ein großer Itempool entstanden ist. In gewissen Abständen fand im Institut ein Workshop statt, der von Frau Mag. Dr.

Holocher-Ertl organisiert wurde, an dem auch Herr Prof. Dr. Kubinger sowie alle DiplomandInnen teilnahmen. In diesem Workshop hat jede/r DiplomandIn zusätzliche Rückmeldungen zu seinen Items und zu seinem Untertest bekommen. Es wurde über Items diskutiert und manche nicht brauchbare sofort eliminiert. Ebenfalls wurde die weitere Vorgehensweise festgelegt. Die Workshops waren stets so gestaltet, dass man sich gut aufgehoben und im Itemkonstruktionsprozess gut betreut gefühlt hat. Im Oktober desselben Jahres gingen die DiplomandInnen auf die Suche nach geeigneten Schulen, die bereit waren, an dem Projekt teilzunehmen. Im gleichen Monat wurde beim Stadtschulrat ein Ansuchen auf Bewilligung einer wissenschaftlichen Erhebung an Schulen gestellt, wofür im November die Genehmigung vorgelegen ist. Unmittelbar darauf wurden an den dafür in Frage kommenden Schulen die LehrerInnen- und Eltern-Briefe ausgeteilt. An einem Wochenende im Dezember fand ein AID 2 Zertifizierungskurs statt, an welchem alle DiplomandInnen aber auch andere StudentInnen teilgenommen haben, um sich für die bevorstehenden Testungen noch weiter zu qualifizieren, damit Fehler bei der Kodierung als auch bei der Vorgabe der Untertests vermieden werden und somit Verrechnungssicherheit gewährleisten ist. Im Dezember und teilweise auch noch im Januar wurden die Rückmeldungen der Eltern wieder eingesammelt und es konnte, von den für die einzelnen Schulen zuständigen DiplomandInnen, die weitere Organisation durchgeführt werden. Es war wichtig einen Raumplan aufzustellen, damit es nicht zu Überschneidungen kommt, da jeder DiplomandIn zusätzlich KandidatInnen (welche sich für die dafür aufgewendete Zeit Hilfestunden anrechnen lassen konnten) zu Seite gestellt bekommen hat, welche ebenfalls am Zertifizierungskurs teilgenommen haben, um den zeitlichen Aufwand der Testungen (Einzeltestungen, pro Kind ca. 1,5 h) und Störung des Schulbetriebes in den Schulen so gering wie möglich zu halten. Im Januar fand nochmals ein Zusammentreffen aller wissenschaftlichen MitarbeiterInnen statt, welche letztendlich mit der Datenerhebung beginnen, um die Testzusammenstellungen noch einmal ausführlich durchzugehen und die letzten Unklarheiten zu beseitigen. Ich habe mit meiner Datenerhebung an der Kooperativen Mittelschule (kurz KMS) und der Ganztagsvolksschule (kurz GTVS) nach den Semesterferien im Februar 2010 begonnen. Die Testungen wurden an der KMS vormittags und an der GTVS vormittags als auch nachmittags durchgeführt. Hierfür wurde von allen DiplomandInnen die gesamte neu überarbeitete und neu zusammengestellte Rohversion des AID 3 vorgegeben, damit letztendlich beim Auswerten des Datenmaterials eine große Stichprobe zur Verfügung

steht. Es haben alle DiplomandInnen in den verschiedenen Schulen zur gleichen Zeit getestet. So konnte gewährleistet werden, dass jeder mit seiner Datenerhebung halbwegs zur gleich Zeit fertig wird, um im Anschluss daran sofort mit der Auswertung des Datenmaterials beginnen zu können. Die Testungen zogen sich bis in den Mai hinein. Im Anschluss folgte die Auswertung des gewonnenen Datenmaterials. Nach dieser Auswertung im Sommer 2010, wurden den Eltern die schriftlichen Ergebnisberichte über die erbrachten Leistungen ihres Kindes per Post zugesandt.

Anonymisierung

Um eine anonymisierte Verarbeitung und Speicherung der Daten der Testpersonen zu gewährleisten, erhielt jedes Kind bzw. Jugendlicher einen Probandencode, welcher sich aus den Initialen des Testleiters und einer fortlaufenden Nummer zusammensetzte.

9.3. Erhebungsinstrument

Das Erhebungsinstrument war die überarbeitete und aktualisierte Form des AID 2 (Kubinger, 2009) – die Rohform des AID 3. Um möglichst viele Items und Untertests einer ersten Überprüfung unterziehen zu können und dabei aber eine zumutbare Testdauer nicht zu überschreiten, gab es in diesem Erhebungsinstrument vier Testzusammenstellungen. Die Untertests jeder dieser Zusammenstellung waren statt wie üblich mit Zahlen nun mit Buchstaben gekennzeichnet. In allen vier Testzusammenstellungen sind 9 Untertests (A bis I) enthalten, welche zwingend jeder Testperson vorgegeben wurden. UT A „Alltagswissen“, UT B „Visuelle Merkfähigkeit“*, UT C „Antonyme“*, UT D „Realitätssicherheit“, UT E „Angewandtes Rechnen“, UT F „Synonyme Finden“, UT G „Logische Blöcke“*, UT H „Funktionen Abstrahieren“ und UT I „Soziales Erfassen und sachliches Reflektieren“. Alle mit einem Stern gekennzeichneten Untertests sind, von an dieser Untersuchung mitarbeitenden DiplomandInnen, völlig neu konstruiert worden. Die restlichen sechs Untertests sind wie eingangs erwähnt überarbeitete und aktualisierte Untertests des bereits bekannten AID 2. Bei einigen Untertests mussten wegen der großen Itemanzahl, Untergruppen von Personen jeweils andere Untermengen von Aufgaben vorgegeben werden.

Ebenfalls war pro Testzusammenstellung je ein weiterer Untertest (aus dem AID 2 bereits bekannt), welcher aber gleichfalls überarbeitet und mit neuen Items versehen wurde, vorzugeben. Testzusammenstellung 1: „Soziale und Sachliche Folgerichtigkeit“,

Testzusammenstellung 2: „Antizipieren und Kombinieren-figural“, Testzusammenstellung 3: „Analysieren und Synthetisieren-abstrakt“ und Testzusammenstellung 4 mit dem Untertest: „Unmittelbares Reproduzieren-numerisch“ und „Kodieren und Assoziieren“. Die optionalen Zusatztests („Unmittelbares Reproduzieren-figural/abstrakt“, „Merken und Einprägen“ und „Strukturieren-visumotorisch“), welche beim AID 2 vorgesehen sind, wurden in diesem Erhebungsinstrument nicht vorgegeben (der Protokollbogen ist dem Anhang beigelegt).

9.4. Stichprobe

Die Stichprobe musste so gewählt werden, dass man SchülerInnen findet, die dem Altersbereich von 6 (6;0) – 16 (15;11) Jahren entsprechen und wenn möglich gleichmäßig verteilt sind. Das Finden der Stichprobe gestaltete sich etwas schwierig. Die DiplomandInnen gingen zu Beginn auf die Suche nach Schulen, die das Kriterium des Altersbereichs erfüllten. Hierfür kamen Volksschulen, Hauptschulen, Mittelschulen sowie Gymnasien in Frage. Beim Nachfragen bei den infrage kommenden Schulen, ob sie bereit wären, einer Erhebung an SchülerInnen mit dem AID 3 zuzustimmen, erhielten die DiplomandInnen überall die gleiche Auskunft: Dass die Schulen einer Erhebung nur mit Genehmigung des Stadtschulrates zustimmen würden. Das Paradoxe daran ist jedoch, dass wenn man um eine Genehmigung beim Stadtschulrat ansucht, man schon Schulen vorweisen muss, die bereit wären an dem Projekt/Erhebung teilzunehmen. So wurden die Schulen abermals kontaktiert, meist zuerst telefonisch, dann per Mail, um von ihnen eine vorbehaltliche Zusage zu dem Projekt zu erhalten, bis die Genehmigung vom Stadtschulrat vorläge. Insgesamt erklärten sich acht Schulen aus Wien und zwei aus Niederösterreich bereit an dem Projekt teilzunehmen (fünf Volksschulen (VS), darunter eine Ganztagsvolksschule (GTVS), eine Kooperative Mittelschule (KMS), eine Hauptschule (HS), zwei Gymnasien (BRG) und eine Berufsbildende Höhere Schule (BHS). Nachdem der Universität diese Genehmigung vorlag, welche im Anhang zu finden ist, konnte eine detaillierte Planung beginnen. Es wurden die Schulen davon in Kenntnis gesetzt, dass die Genehmigung vorliege und mit ihnen abgesprochen, welche Klassenstufen für die Testungen in Frage kämen. In meinen Aufgabenbereich fielen wie bereits erwähnt zwei Schulen: 1. GTVS Pastinakweg 3, 1220 Wien 2. KMS Stromstraße 40, 1200 Wien. Für die Datenerhebung waren aus der GTVS SchülerInnen der dritten und vierten Klassen und aus der KMS 1. - 4. Schulstufe erforderlich. In diesen Schulstufen wurden LehrerInnen-Briefe ausgeteilt

(ebenfalls im Anhang), mit der Bitte, die beiliegenden Elternbriefe (dem Anhang zugefügt) an ihre SchülerInnen auszuteilen und die angehängten schriftlichen Rückmeldungen der Eltern gesammelt in der Direktion abzugeben. Den Eltern wurde bei einer eventuellen Teilnahme ihres Kindes an dem Projekt ein kurzer schriftlicher Ergebnisbericht, über das Abschneiden ihres Kindes in dem Test, in Aussicht gestellt. Da sich alle statistischen Analysen auf die Gesamtstichprobe beziehen, wird auf eine ausführliche Vorstellung der Teilstichprobe (GTVS und KMS) verzichtet. Die Gesamtstichprobe umfasst insgesamt 711 Kinder bzw. Jugendliche, wobei letztendlich beim Untertest „Angewandtes Rechnen“ nur von einem Datensatz von 707 Kinder bzw. Jugendlichen und beim Untertest „Funktionen Abstrahieren“ nur vom Datensatz von 691 Kindern bzw. Jugendlichen die statistischen Analysen durchgeführt wurden. Auf die Ursachen wird in einem späteren Abschnitt eingegangen. Für die statistischen Analysen wurde das Programm SPSS Software für MS Windows verwendet. Am geeignetsten wäre es, den Datensatz hinsichtlich bekannter Teilungskriterien zu teilen, wie: *Schulform*, *Alter*, *Geschlecht* und *Nationalität*. Da es aber in Österreich, vor allem in Wien, viele Kinder und Jugendliche gibt, welche sich in ihrer Nationalität nicht unterscheiden, sehr wohl aber in ihrer Muttersprache (der AID 2 aber für den deutschen Sprachraum konzipiert wurde) und da es Schulen gibt, in denen hauptsächlich Kinder bzw. Jugendliche mit nicht deutscher Muttersprache unterrichtet werden, wurden die Daten hinsichtlich *Schulform*, *Alter*, *Geschlecht* und *Muttersprache* analysiert.

9.4.1. Schulform

Tabelle 2 gibt die Häufigkeit und den Prozentsatz der Gesamtstichprobe, aller getesteten Schüler und Schülerinnen, in Bezug auf die *Schulform* wieder. Hieraus wird ersichtlich, dass die meisten Kinder bzw. Jugendlichen in den Volksschulen und in Gymnasien getestet wurden mit einem Prozentanteil von 80%, wobei die Volksschulkinder den größeren Anteil ausmachen.

Tabelle 2: Deskriptive Statistik der Variable Schulform

Schulform	Häufigkeit	Prozent %
Volksschule	300	42.2
Hauptschule	56	7.9

Kooperative Mittelschule	68	9.6
Gymnasium	270	38.0
Berufsbildende höhere Schule	17	2.4
Gesamt	711	100

9.4.2. Muttersprache

Bei der *Muttersprache* sind über 10 verschiedene Sprachen zu unterscheiden gewesen. Der Vollständigkeit halber werden alle Sprachen mit ihren Häufigkeiten und Prozentanteilen in Tabelle 3 dargestellt. Den größten Anteil machen die Kinder bzw. Jugendlichen aus, die Deutsch als Muttersprache haben. Es sind immerhin 67,1 % der Gesamtstichprobe. Der zweitgrößte Anteil entfällt auf die Kinder bzw. Jugendlichen die Kroatisch, Serbisch oder Slowenisch als Muttersprache haben, mit 10,5 % und den drittgrößten Anteil machen die Kinder bzw. Jugendlichen mit türkischer Muttersprache aus, welche zu 7,2 % vertreten sind. Die anderen Sprachen machen insgesamt 14,8 % aus. Daraus lässt sich ableiten, dass die Gesamtstichprobe zu 33 % aus Kindern bzw. Jugendlichen besteht, welche deutsch nicht als Muttersprache haben.

Tabelle 3: Deskriptive Statistik der Variable Muttersprache

Muttersprache	Häufigkeit	Prozent %
Deutsch	477	67.1
Türkisch	51	7.2
Kroatisch/Serbisch/Slowenisch	75	10.5
Slowakisch/Tschechisch	8	1.1
Polnisch	15	2.1
Russisch	6	0.8
Englisch	1	0.1
Französisch/Italienisch/Spanisch	8	1.1

Arabisch	22	3.1
Sonstige	42	6.1
Fehlend	3	0.4
Gesamt	711	100

9.4.3. Alter und Geschlecht

Bei der Verteilung der Testpersonen in Bezug auf das *Geschlecht*, wurde über die Jahre hinweg ein relativ ausgewogenes Verhältnis hergestellt. So ergibt sich für die Gesamtstichprobe, dass die Kinder und Jugendlichen in Bezug auf das Geschlecht annähernd gleich verteilt sind. Es wurden insgesamt 328 männliche Personen und 383 weibliche Personen getestet. Wenn man die Anzahl der SchülerInnen in Bezug auf die Altersstufen betrachtet, wird ersichtlich, dass in der Altersstufe der 10-jährigen am meisten Kinder und in der Altersstufe der 6-jährigen am wenigsten Kinder getestet wurden.

Tabelle 4: Deskriptive Statistik der Variable Alter und Geschlecht

Alter in Jahren	6	7	8	9	10	11	12	13	14	≥15	Gesamt
Männlich	13	35	36	34	53	43	39	29	30	16	328
Weiblich	23	33	27	41	43	51	51	36	49	29	383
Gesamt	36	68	63	75	96	94	90	65	79	45	711

9.5. Durchführung der Untersuchung

Die Testungen wurden wie bereits erwähnt in der KMS vormittags und in der GTVS sowohl vormittags als auch nachmittags durchgeführt, wobei stets darauf geachtet wurde, dass die Kinder bzw. Jugendlichen ihre Pausen einhielten, d.h. dass während der

Frühstückspause und dem Mittagstisch nicht getestet wurde. Den LehrerInnen war es des Weiteren wichtig, dass die Kinder bzw. Jugendlichen nicht während der Sportstunden getestet wurden, so dass für jede Klasse ein Testplan zusammengestellt wurde, wann das Testen der Kinder bzw. Jugendlichen in den verschiedenen Klassen möglich war.

Die Testungen selbst waren Einzeltestungen. Der zeitliche Aufwand für jede Testung schwankte zwischen 70 und 120 Minuten.

Für jede Schule wurde vom Verfasser für die Testungen zusätzlich zum Testplan ein Raumplan und „Verhaltensregeln“ (wie Patschenpflicht und dergleichen) zur Verfügung gestellt, so dass die Testungen, auch von den uns zur Seite gestellten TestleiterInnen, organisiert durchgeführt werden konnten. Von den Schulen wurden als Testräume unter anderem Klassenräume, Arztzimmer, Bibliothek und Vorbereitungszimmer zur Verfügung gestellt.

Die LehrerInnen zeigten sich an den Testungen interessiert und waren zu jeder Zeit entgegenkommend.

10. Ergebnisse

Statistische Berechnungen und Deskriptive Statistik

Die statistischen Analysen der Daten, um die Hypothesen zu prüfen erfolgte für beide Untertests mittels dem Programm R Version 2.15.2 (A Language and Environment for Statistical Computing) (R Core Team, 2012) und dem Paket eRm Version 0.15-1 (Extended Rasch Modeling) (Mair, Hatzinger & Maier, 2012).

Der ursprüngliche Datensatz enthielt die Daten von 711 Kindern bzw. Jugendlichen. Wie bereits erwähnt, konnten die statistischen Analysen für den Untertest „Angewandtes Rechnen“ nur von einem Datensatz von 707 und beim Untertest „Funktionen Abstrahieren“ nur von einem Datensatz von 691 Personen durchgeführt wurden. Bei den ausgeschlossenen Personen handelt es sich um Kinder, bei denen am Protokollbogen die Muttersprache nicht angegeben wurde (welche auch nicht nachrecherchiert werden konnte), aber auch um Personen, denen der entsprechende Untertest aus Gründen des Sprachverständnisses, der Überforderung oder Unkonzentration nicht vorgegeben bzw. frühzeitig beendet wurde.

10.1. Überprüfung der Hypothesen des Untertests Angewandtes Rechnen

Die Rohdaten des Untertests (UT) wurden in Bezug auf vier Teilungskriterien (TK) inferenzstatistisch mit Hilfe des Likelihood-Ratio-Tests (LR-Test) von Anderson analysiert:

Internes Teilungskriterium:

- Rohscore (niedriger Score vs. hoher Score)

Externes Teilungskriterium:

- Geschlecht (männlich vs. weiblich)
- Muttersprache (deutsch vs. nicht deutsch)
- Alter (≤ 9 vs. ≥ 10)

Wenn nun bei der Überprüfung der Raschkonformität der Modelltest (LR-Test) hinsichtlich eines dieser Teilungskriterien signifikant ($\alpha = .01$) ausfällt, werden weitere Tests, wie der graphische Modelltest und der Wald-Test herangezogen, um die nicht modellkonformen Items zu identifizieren und sukzessive auszuschließen. Beim graphischen Modelltest werden jene Items als auffällig bewertet und als nicht

modellkonform angenommen und dem zur Folge ausgeschlossen, die deutlich außerhalb des Toleranzbereiches liegen. Der Toleranzbereich ist in den folgenden Abbildungen mit Linien, die parallel zur 45° -Geraden liegen, gekennzeichnet. Auch die Konfidenzintervall-Ellipsen der einzelnen Items (die Abbildungen hierzu finden sich im Anhang) sind bei der Entscheidung, welches Item als nächstes auszuschneiden ist, mit herangezogen worden. Wenn z.B. ein Item außerhalb der parallelen Linien zu der 45° -Geraden liegt, aber ihr Konfidenzintervall die 45° -Gerade eindeutig schneidet, wurde dieses Item in diesem bestimmten Durchgang nicht ausgeschieden. Wenn aber die Konfidenzintervall-Ellipse die 45° -Gerade nicht schneidet und das Item zudem außerhalb der parallelen Linien zu dieser liegt, ist das ein eindeutiger Hinweis darauf, dass das entsprechende Item nicht raschkonform und dem zur Folge auszuschneiden ist. Unter anderem sind beim Ausschluss der Items auch die Erfahrungen, welche die anderen Testleiter während der Testungen gemacht haben, hinsichtlich inhaltlich kritischer Items berücksichtigt worden. Das Ausscheiden der auffälligen Items geschieht so lange, bis in Bezug auf die erwähnten Teilungskriterien keine signifikante Modellabweichung mehr festzustellen ist.

10.1.1. Erster Berechnungsdurchgang

Um die statistischen Analysen mit dem Programm R Version 2.15.2 (A Language and Environment for Statistical Computing) (R Core Team, 2012) durchführen zu können, musste ein dafür passendes Skript geschrieben werden, was durch einen Kollegen erfolgte. Die Ergebnisse, entfernten Items und Kommentare der einzelnen Berechnungsdurchgänge wurden zur Übersicht und weiteren Verwendung in einem „Rauswurfdokument“ gespeichert.

Teilungskriterium Rohscore

In Bezug auf den *Rohscore* zeigte der LR-Test für den Untertest „Angewandtes Rechnen“ ein signifikantes Ergebnis. Daher ist die Hypothese H_{0-1} : „Die Items des Untertests „Angewandtes Rechnen“ entsprechen dem Rasch-Modell“ zunächst zu verwerfen. In Tabelle 5 werden die asymptotisch χ^2 -verteilten Testgrößen des LR-Tests bezüglich dieses Teilungskriteriums angezeigt. Unter anderem werden die Anzahl der dazu berücksichtigten Aufgaben ($df+1$), die Wahrscheinlichkeit, dass die H_{0-1} gilt (p-Wert) und der kritische Wert der χ^2 -Verteilung ($\alpha = .01$) angegeben.

Tabelle 5: UT „Angewandtes Rechnen“ LR-Test für das TK Rohscore, erster Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
198.12	58	0.001	85.95

In der Abbildung 4 wird der graphische Modelltest für das Teilungskriterium *Rohscore* für alle in die Berechnungen einbezogene Items angezeigt.

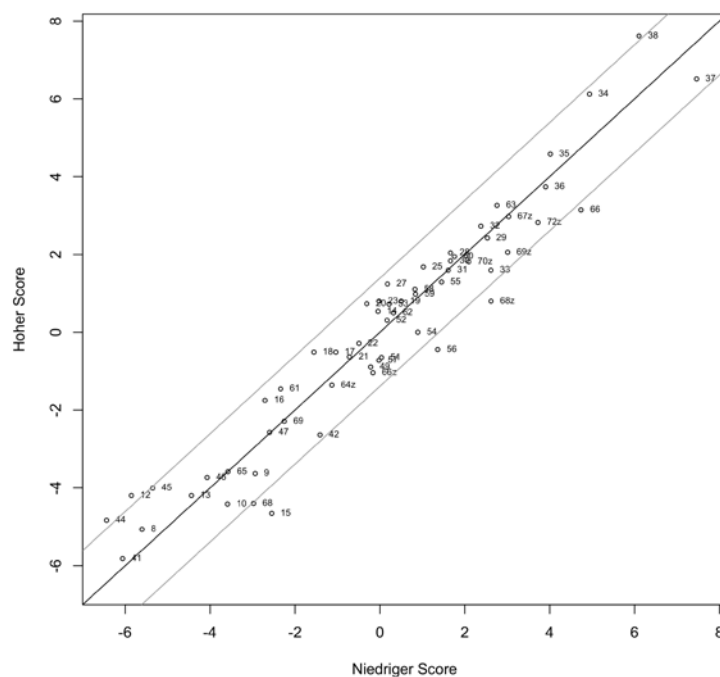


Abbildung 4: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Rohscore

Tabelle 6 zeigt die Personen und Items, welche aufgrund der Nichtpassung des Modells, bei diesem Berechnungsdurchgang von den Analysen ausgeschlossen wurden sowie die nicht modellkonformen Items. Einige von ihnen mussten ausgeschieden werden, da sie von allen Teilnehmern gelöst wurden (immer gelöst) und andere wiederum, weil sie ungünstige Antwortmuster gezeigt haben (inappropriate response patterns). Die inappropriate response patterns wurden nicht grundsätzlich vom Itempool entfernt, da sie bei den verbleibenden Teilungskriterien sehr wohl geschätzt werden konnten. Beim

ersten Berechnungsdurchgang dieses Teilungskriteriums gab es 11 Items, die vom Berechnungsdurchgang von vornherein ausgeschlossen wurden und sieben nicht modellkonforme Items. Im Folgenden wird ersichtlich, dass bei den Berechnungen pro Teilungskriterium hinsichtlich der inappropriate response patterns, unterschiedliche Items von den Analysen ausgeschlossen wurden.

Tabelle 6: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Rohscore, erster Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh213, bh41, i11 ng110, ng122		i64, i5, i7, i43, i67	i1, i2, i3, i4, i6	i15, i56, i68z, i66, i44, i12, i38

Teilungskriterium Geschlecht

Bei der Rasch-Modellprüfung in Bezug auf das Teilungskriterium *Geschlecht* erbrachte der LR-Test ebenfalls ein signifikantes Ergebnis. Daher ist auch hier die Hypothese H_0 -2: „Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit nicht.“ zunächst zu verwerfen. Die Tabelle 7 zeigt die asymptotisch χ^2 -verteilten Testgrößen des LR-Tests vom ersten Berechnungsdurchgang des Teilungskriteriums *Geschlecht*.

Tabelle 7: UT „Angewandtes Rechnen“ LR-Test für das TK Geschlecht, erster Berechnungsdurchgang

χ^2-Wert empirisch	df	p-Wert	χ^2-Wert kritisch
102.10	60	0.001	88.38

Die Abbildung 5 zeigt den dazugehörigen graphischen Modelltest für das Teilungskriterium *Geschlecht*. Fast alle Items liegen nahe der 45°-Geraden.

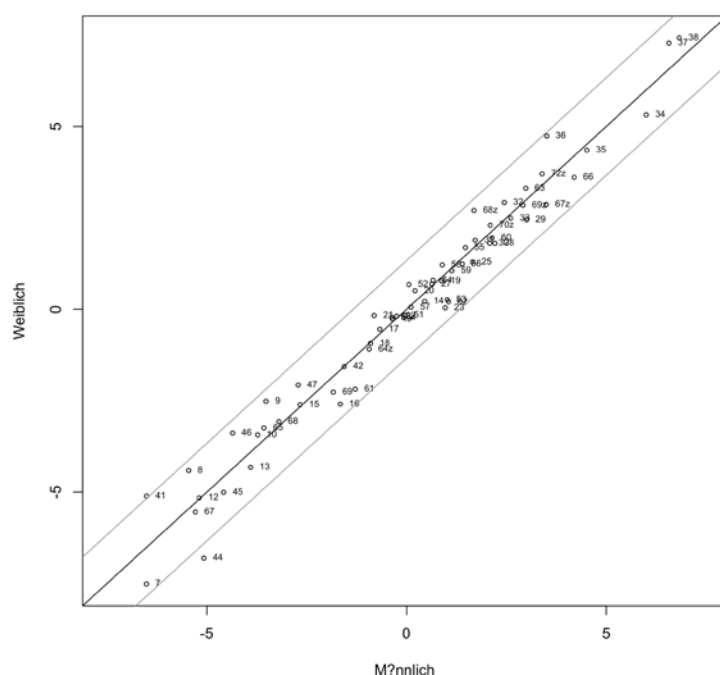


Abbildung 5: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Geschlecht

Die Tabelle 8 zeigt die „Statistik“ der Personen und Items. Siehe dazu Abschnitt 10.1.1. Beim ersten Berechnungsdurchgang dieses Teilungskriteriums wurden insgesamt 9 Items aufgrund ungünstiger Antwortmuster von den Analysen ausgeschlossen und es gab nur ein nicht modellkonformes Item.

Tabelle 8: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Geschlecht, erster Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh213, bh41, ng110, ng122	i11	i5, i43, i64	i1, i2, i3, i4, i6	i44

Teilungskriterium Muttersprache

Der LR-Test wird beim Teilungskriterium *Muttersprache* ebenfalls signifikant, aus diesem Grund muss auch diese Hypothese H_0 -3: „Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen keine Kinder bzw. Jugendlichen auf Grund ihrer Muttersprache“ erst einmal verworfen werden (siehe Tabelle 9).

Tabelle 9: UT „Angewandtes Rechnen“ LR-Test für das TK *Muttersprache*, erster Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
118.01	58	0.001	85.95

Abbildung 6 zeigt den graphischen Modelltest für das Teilungskriterium *Muttersprache*.

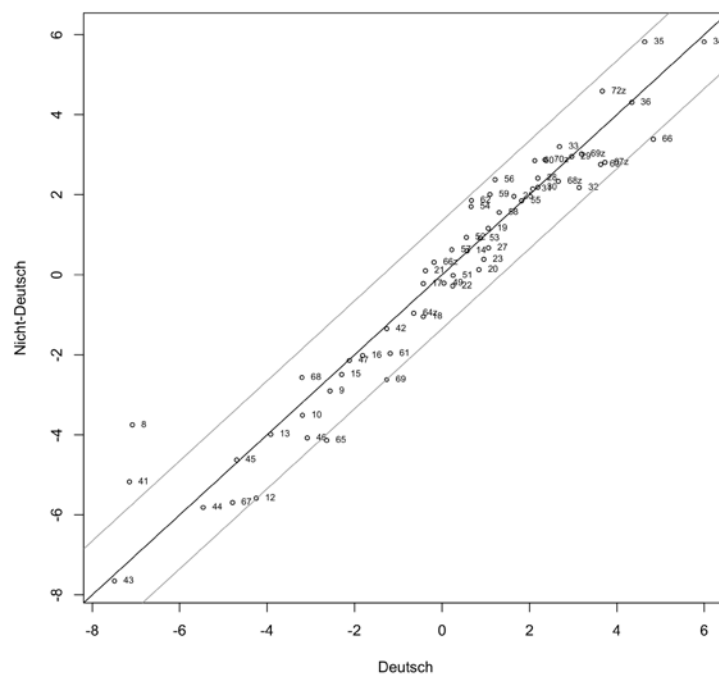


Abbildung 6: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK *Muttersprache*

In der Tabelle 10 wird die „Statistik“ der Personen und Items, welche vom Berechnungsdurchgang von vornherein ausgeschlossen wurden angeführt. Siehe dazu

Abschnitt 10.1.1. Bei diesem Berechnungsdurchgang wurden 11 Items von den Analysen ausgeschlossen und drei Items sind als nicht modellkonform zu bewerten.

Tabelle 10: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Muttersprache, erster Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modellkonforme Items
bh213, bh41, ng110, ng122	i11	i5, i7, i37, i38, i64	i1, i2, i3, i4, i6	i8, i41, i65

Teilungskriterium Alter

In Bezug auf das Teilungskriterium *Alter* fällt der LR-Test bei der Rasch-Modell-Prüfung gleichfalls signifikant aus. Auch hier muss die Hypothese H_0-4 : „Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen Kinder bzw. Jugendliche aufgrund ihres Alters nicht“ zunächst verworfen werden (siehe dazu Tabelle 11).

Tabelle 11: UT „Angewandtes Rechnen“ LR-Test für das TK Alter, erster Berechnungsdurchgang

χ^2-Wert empirisch	df	p-Wert	χ^2-Wert kritisch
71.91	33	0.001	54.78

In der Abbildung 7 ist der graphische Modelltest für dieses Teilungskriterium angegeben.

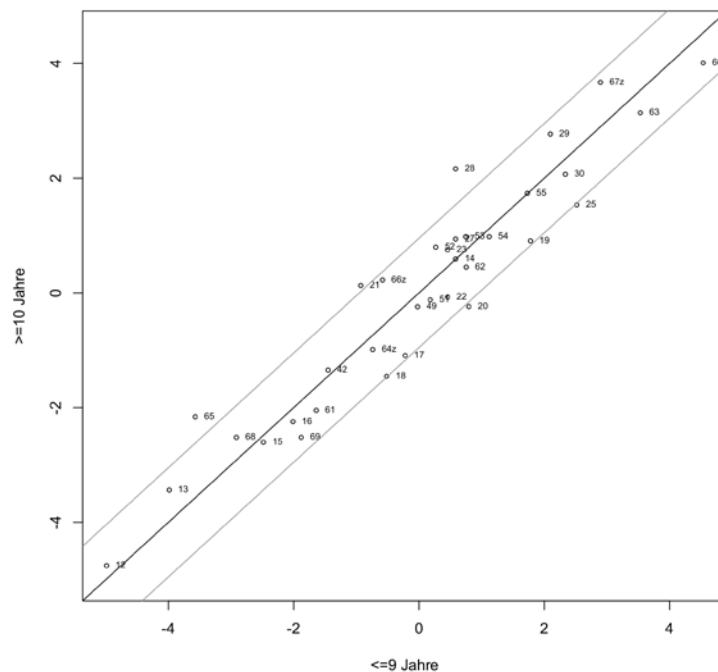


Abbildung 7: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Alter

Die Tabelle 12 zeigt die Statistik der Personen und Items, welche für dieses Teilungskriterium von den Analysen ausgeschlossen wurden. Siehe dazu Abschnitt 10.1.1. 35 Items sind bei diesem Durchgang ausgeschlossen worden und vier Items entsprachen nicht dem Rasch-Modell.

Tabelle 12: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Alter, erster Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh213, ng110, ng122	bh41, i11	i31, i32, i33, i34, i35, i36, i37, i38, i56, i57, i58, i59, i60, i64, i70z, i69z, i72z, i5, i7, i8, i10, i41,	i1, i2, i3, i4, i6	i20, i21, i28, i65

Ausschluss nicht modellkonformer Items

Die nicht modellkonformen Items wurden nun, wie bereits angedeutet sukzessive ausgeschieden, was bedeutet, dass nach jedem Ausschluss ein neuerlicher Berechnungsdurchgang durchgeführt wurde. Für die Beurteilung, welches Item auszuschneiden ist, wurden neben inhaltlichen Analysen der graphische Modelltest und der Wald-Test herangezogen. Weiters wurde, wie bereits erwähnt, auf die Konfidenzintervalle der einzelnen Items geachtet und jene Items ausgeschieden, welche mit ihren Konfidenzintervallen die 45°-Gerade nicht schneiden. Das Hauptaugenmerk beim Ausschluss der Items wurde auf das Kriterium *Rohscore* gelegt. Wenn es Items gab, welche in zwei Teilungskriterien wie z.B. sowohl beim Kriterium *Sprache* als auch beim Kriterium *Alter* im graphischen Modelltest und im Wald-Test auffällig waren, wurden ohne Rücksicht auf den Vorrang des Kriteriums *Rohscore*, diese Items ausgeschlossen. Eine Übersicht über die in den 12 Berechnungsdurchgängen ausgeschlossenen 11 Items findet sich im Anhang. Sieben Items wurden ausgeschieden, weil sie im Kriterium *Rohscore* auffällig waren, ein Item war sowohl im *Rohscore* als auch im Kriterium *Geschlecht* auffällig, ein weiteres gleichzeitig in den Kriterien *Sprache* und *Alter*, ein anderes in den Kriterien *Geschlecht* und *Sprache* und eines nur im Kriterium *Alter*. Wenn ein Item im Kriterium *Rohscore* auffällig geworden ist und sich in dem Dreieck zwischen 45°-Geraden und der x-Achse befindet, bedeuten das, dass dieses Item für Kinder mit niedrigen Score unverhältnismäßig schwieriger ist, als für Kinder mit hohem Score. Wenn ein Item im Kriterium *Geschlecht* auffällig wurde, kann das bedeuten, dass dieses Item entweder für Jungen oder für Mädchen zu schwierig war. Wenn sich z.B. ein Item zwischen der 45°-Geraden und der x-Achse befindet, dann fällt dieses Item männlichen Testpersonen unverhältnismäßig schwerer als weiblichen Testpersonen. Dieses Item misst dann nicht mehr eindimensional, sondern in gewisser Weise auch das Geschlecht und dies ist ein Umstand, der in diesem Test vermieden werden soll. Wurde ein Item im Kriterium *Sprache* auffällig, so bedeutet das, entsprechend der Lage des Items im Koordinatensystem, dass dieses Item entweder Kindern mit deutscher Muttersprache oder Kindern mit nicht deutscher Muttersprache zu schwer gefallen ist und dieses Item zusätzlich zu dem, was es messen

soll auch die Sprache mit erfasst. Beim Kriterium *Alter* ist die Lage der Items im Koordinatensystem ebenso zu interpretieren. Liegt ein Item zwischen der 45°-Geraden und der x-Achse, so ist dieses Item für Kinder unter neun Jahren unverhältnismäßig schwieriger als für Kindern über neun Jahren. Wenn die Items entlang der 45°-Geraden gleichmäßig verteilt sind, ist das ein Hinweis darauf, dass die Items über die verschiedenen Schwierigkeitsstufen hinweg streuen. Aus inhaltlicher Sicht soll exemplarisch auf das ausgeschlossenes Item 37 eingegangen werden. Item 37: „Es soll eine quadratische Metallplatte von 20 cm Seitenlänge hergestellt werden. Eine Nachprüfung ergibt, dass die eine 20,06 cm und die andere 20,14 cm lang ist. Um wie viel Prozent weicht die Fläche der Platte von der gewünschten Fläche ab?“ Dieses Item wurde im Kriterium *Rohscore* auffällig, es liegt in der Nähe der 45°-Geraden im rechten oberen Drittel, was bedeutet, dass es ein eher schwierigeres Item ist. Zudem liegt es unterhalb dieser Geraden, d.h. zwischen 45°-Gerade und x-Achse, was bedeutet, dass dieses Item für Testpersonen mit niedrigem Score unverhältnismäßig schwieriger ist, als für Testpersonen mit hohem Score und dem zur Folge diese benachteiligen würde. Dies kann natürlich an den vielen zum Einsatz kommenden Rechenoperationen liegen, denn die Testperson muss als erstes die Fläche der ursprünglichen Platte berechnen, nämlich $20\text{ cm} \times 20\text{ cm} = 400\text{ cm}^2$. Im Anschluss muss die Fläche der tatsächlichen Platte berechnet werden, und zwar: $20,06\text{ cm} \times 20,14\text{ cm} = 404,0084\text{ cm}^2$ und nun kommt eine Prozentrechnung zum Einsatz, bei der die Testperson wissen muss, in welches Verhältnis sie die beiden Flächen zueinander zu setzen hat. Die ursprüngliche Platte sind dabei die 100%, so macht die tatsächliche Platte 101,0021% aus, wovon nun noch, weil ja nach dem Unterschied gefragt wird, die 100% abzuziehen sind. Als Ergebnis würde exakt 1,0021% herauskommen. An Rechenoperationen kommen zwei Multiplikationen, eine Prozentrechnung und eine Subtraktion zum Einsatz. Erschwert wird das Ganze noch, indem als Operanten Dezimalzahlen verwendet wurden. Obwohl Items im AID 3 vorhanden sein müssen, die die höheren Fähigkeitsbereiche abdecken und nicht von jeder Testperson zu lösen sein sollten, dürfen sie aber nicht bestimmte Testpersonen benachteiligen. Aus diesem Grund wurde das Item 37 entfernt. In ähnlicher Weise und inhaltlicher Analyse wurde mit den anderen entfernten Items verfahren.

10.1.2. Letzter Berechnungsdurchgang

Teilungskriterium Rohscore

Nach 12 Berechnungsdurchgängen und dem Ausschluss von 11 Items war der LR-Test in Bezug auf das Teilungskriterium *Rohscore* immer noch signifikant. Dem zur Folge liegt der empirische χ^2 -Wert immer noch über dem kritischen χ^2 -Wert ($\alpha = .01$). Warum es zu diesem relativ großen Unterschied zwischen dem empirischen und kritischen χ^2 -Wert kommt, obwohl die Items bis auf das Item i34 (siehe Tabelle 14) im graphischen Modelltest relativ eng um die 45°-Gerade streuen (siehe Abbildung 8) und in wie weit die Hypothese H_0 -1: „Die Items des Untertests „Angewandtes Rechnen“ entsprechen dem Rasch-Modell“ gilt, wird im Abschnitt unter Diskussion und Ausblick näher behandelt. Der Wald-Test identifiziert noch 3 signifikante Items (siehe Tabelle 38 im Anhang). Das Item, welches nicht modellkonform ist und mit seiner Ellipse die 45°-Gerade nicht schneidet und im folgenden Durchgang noch hätte ausgeschieden werden können, war jedoch in den vorangegangenen Durchgängen modellkonform, so dass der Ausschluss weiterer Items hier abgebrochen und die Signifikanz des LR-Tests erst einmal hingenommen wurde.

Tabelle 13: UT „Angewandtes Rechnen“ LR-Test für das TK Rohscore, letzter Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
113.08	45	0.001	69.95

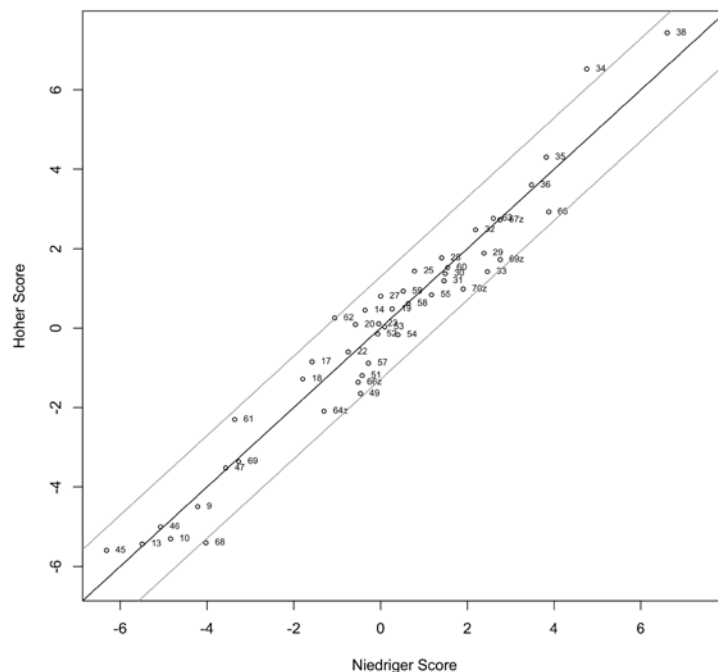


Abbildung 8: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Rohscore, letzter Berechnungsdurchgang

Tabelle 14: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Rohscore, letzter Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh213, bh41, ng110, ng122	i7, i11, i41	i5, i15, i43, i64, i67	i1, i2, i3, i4, i6	i34

Teilungskriterium Geschlecht

Nach Ausschluss der nicht modellkonformen Items ergibt der LR-Test für das Teilungskriterium *Geschlecht* ein nicht signifikantes Ergebnis. Der empirische χ^2 -Wert liegt unter dem kritischen χ^2 -Wert, was in Tabelle 15 ersichtlich ist. In der Abbildung 9

sieht man sehr beispielhaft, dass die Items dem Rasch-Modell entsprechen, denn alle Items streuen eng um die 45°-Gerade. Auch in der Tabelle 16 wird ersichtlich, dass es keine Items gibt, welche nicht raschkonform sind. Das heißt, für die Items in Bezug auf das Teilungskriterium *Geschlecht* gilt das Rasch-Modell. Demnach kann die Hypothese H_0-2 : „Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit nicht“ angenommen werden.

Tabelle 15: UT „Angewandtes Rechnen“ LR-Test für das TK Geschlecht, letzter Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
70.89	47	0.001	72.44

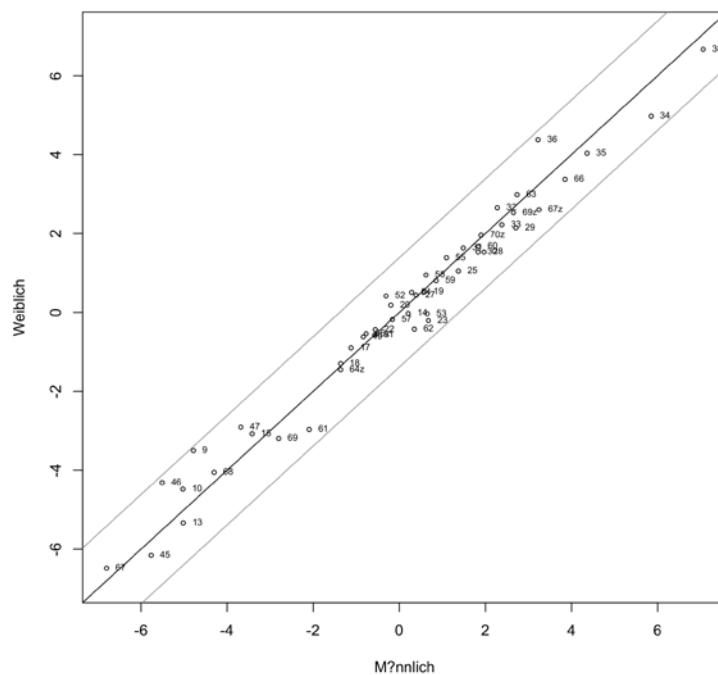


Abbildung 9: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Geschlecht, letzter Berechnungsdurchgang

Tabelle 16: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Geschlecht, letzter Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh213, bh41, ng110, ng122	i7, i11, i41	i5, i43, i64	i1, i2, i3, i4, i6	-

Teilungskriterium Muttersprache

Der LR-Test für das Teilungskriterium *Muttersprache* ergibt ein signifikantes Ergebnis. Dem zur Folge liegt der empirische χ^2 -Wert auch hier immer noch über dem kritischen χ^2 -Wert ($\alpha = .01$). Die Hypothese H_0 -3: „Die verbleibenden Items des Untertests „Angewandtes Rechnen“ benachteiligen keine Kinder bzw. Jugendlichen auf Grund ihrer Muttersprache“ muss deshalb nicht zwingend verworfen werden. Eine Diskussion darüber findet sich im Abschnitt 12. Wenn man die Abbildung 10 und die Tabelle 18 heranzieht, kommt es nämlich zu einem Widerspruch, da bei der Betrachtung des graphischen Modelltests ersichtlich wird, dass die Daten sehr wohl eng um die 45°-Gerade streuen und es keine Items in Bezug auf dieses Teilungskriterium mehr gibt, welche nicht rasch-konform sind. Dies lässt sich vielleicht dadurch erklären, dass der Wald-Test immer noch zwei Items als signifikant identifiziert, aber beim graphischen Modelltest nicht erkannt werden.

Tabelle 17: UT „Angewandtes Rechnen“ LR-Test für das TK Muttersprache, letzter Berechnungsdurchgang

χ^2-Wert empirisch	df	p-Wert	χ^2-Wert kritisch
86.18	47	0.001	72.44

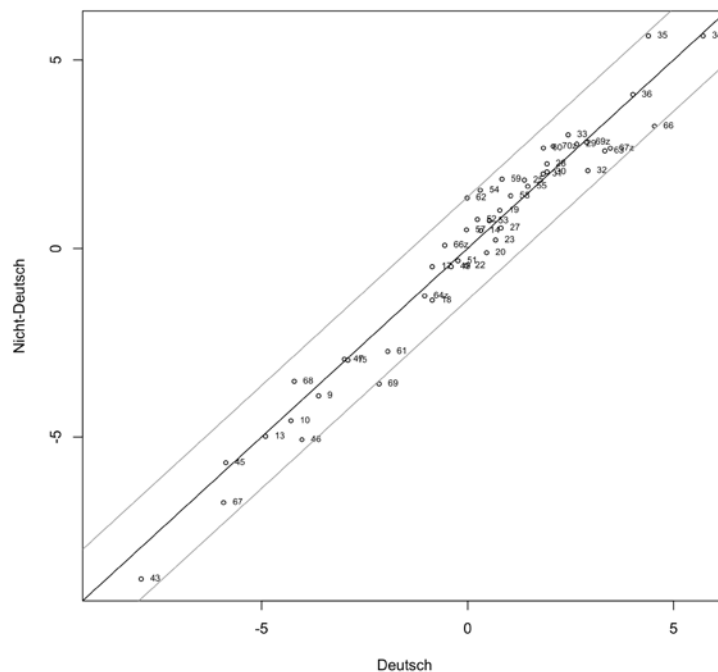


Abbildung 10: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Muttersprache, letzter Berechnungsdurchgang

Tabelle 18: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Muttersprache, letzter Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh213, bh41, ng110, ng122	i7, i11, i41	i5, i38, i64	i1, i2, i3, i4, i6	-

Teilungskriterium Alter

In Bezug auf das Teilungskriterium *Alter* wurde der LR-Test nach Ausschluss der oben erwähnten 11 Items nicht signifikant. Der empirische χ^2 -Wert liegt knapp unter dem kritischen χ^2 -Wert, was in Tabelle 19 zu sehen ist. Hier kann man das umgekehrte Phänomen betrachten. Der LR-Test ist nicht signifikant, obwohl beim graphischen

Modelltest (siehe Abbildung 11) einige wenige Items etwas entfernter von der 45°-Geraden liegen und immer noch 3 Items (siehe Tabelle 20) nicht raschkonform sind. Eine Diskussion darüber findet sich im Abschnitt Diskussion und Ausblick.

Tabelle 19: UT „Angewandtes Rechnen“ LR-Test für das TK Alter, letzter Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
47.75	28	0.001	48.28

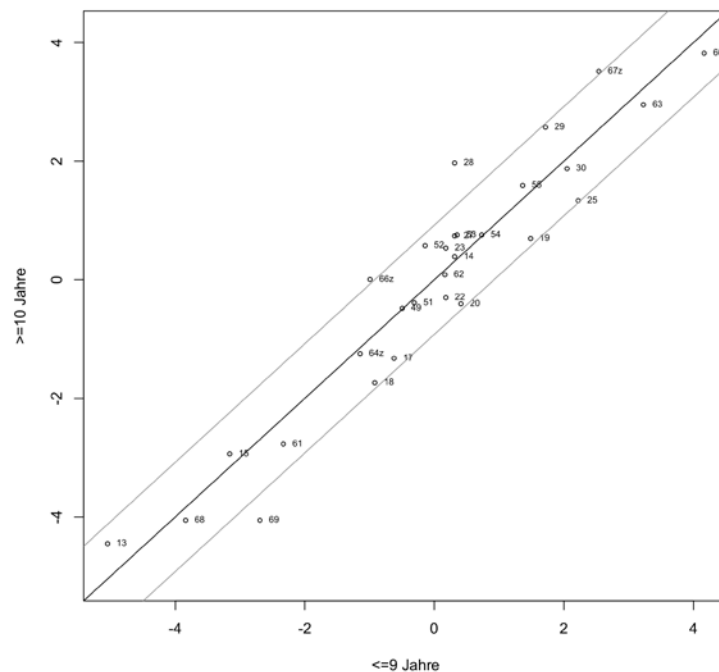


Abbildung 11: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Alter, letzter Berechnungsdurchgang

Tabelle 20: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Alter, letzter Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh213, bh41, ng110, ng122	i7, i11, i41	i31, i32, i33, i34, i35, i36, i38, i57, i58, i59, i60, i64, i70z, i69z, i5, i10, i47, i9, i43, i45, i46, i67	i1, i2, i3, i4, i6	i28, i69, i66z

10.2. Überprüfung der Hypothesen des Untertests Funktionen Abstrahieren

Auch die Rohdaten dieses Untertests wurden in Bezug auf die gleichen vier Teilungskriterien inferenzstatistisch mit Hilfe des Likelihood-Ratio-Tests (LR-Test) von Anderson analysiert:

Internes Teilungskriterium:

- Rohscore (niedriger Score vs. hoher Score)

Externes Teilungskriterium:

- Geschlecht (männlich vs. weiblich)
- Muttersprache (deutsch vs. nicht deutsch)
- Alter (≤ 9 vs. ≥ 10)

Wenn auch hier bei der Überprüfung der Raschkonformität der Modelltest hinsichtlich eines dieser Teilungskriterien signifikant ($\alpha = .01$) ausfällt, werden weitere Tests, wie der graphische Modelltest und der Wald-Test herangezogen, um die nicht modellkonformen Items zu identifizieren und sukzessive auszuschließen. Dies geschieht, wie bereits erwähnt, so lange, bis in Bezug auf die erwähnten Teilungskriterien keine signifikante Modellabweichung mehr vorliegt.

10.2.1. Erster Berechnungsdurchgang

Teilungskriterium Rohscore

In Bezug auf den *Rohscore* zeigte der LR-Test für den Untertest „Funktionen Abstrahieren“ ein signifikantes Ergebnis. Daher ist die Hypothese H_0-1 : „Die Items des Untertest „Funktionen Abstrahieren“ sind raschkonform“ zu verwerfen. In Tabelle 21 werden wie beim vorangegangenen Untertest die asymptotisch χ^2 -verteilten Testgrößen des LR-Tests bezüglich dieses Teilungskriteriums angezeigt. Unter anderem werden die Anzahl der dazu berücksichtigten Aufgaben (df+1), die Wahrscheinlichkeit, dass die H_0-1 gilt (p-Wert) und der kritische Wert der χ^2 -Verteilung ($\alpha = .01$) angegeben.

Tabelle 21: UT „Funktionen Abstrahieren“ LR-Test für das TK Rohscore, erster Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
173.60	66	0.001	95.63

In der Abbildung 12 wird der graphische Modelltest für das Teilungskriterium *Rohscore* für alle in die Berechnungen einbezogene Items angezeigt.

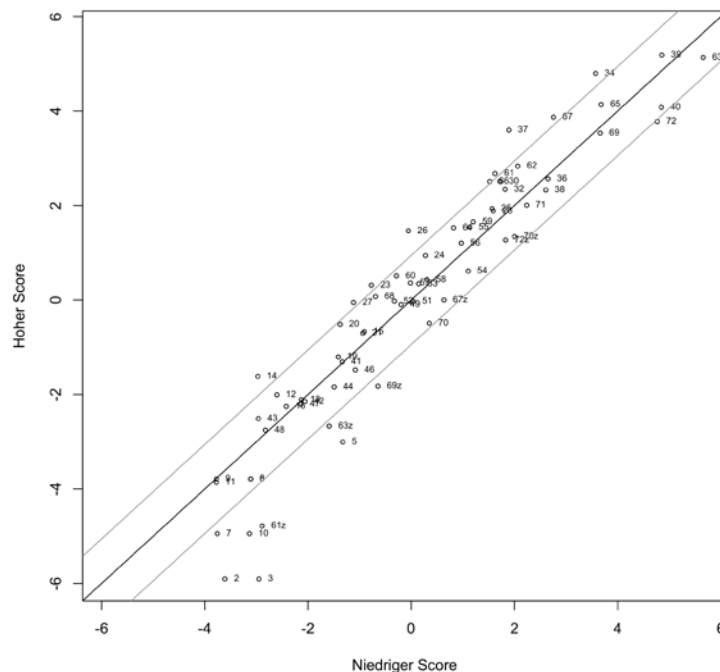


Abbildung 12: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Rohscore

Tabelle 22 zeigt die Personen und Items, welche bei diesem Berechnungsdurchgang aus berechnungstechnischen Gründen von den Analysen ausgeschlossen wurden. Beim ersten Berechnungsdurchgang dieses Teilungskriteriums betraf das 1 Item und 8 Personen.

Tabelle 22: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Rohscore, erster Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh41, bh176, bh179, bh190, bh200, bh214, ng47,	bh45, i14, bh178, bh188, bh191, bh205, bh222, ng49,	-	-	i2, i3, i7, i10, i61z, i5, i63z, i69z, i34, i57, i37, i61, i26, i23, i27, i14

ng64, ng110,
ng122, mh45,
mh92, mh115

Teilungskriterium Geschlecht

Bei der Rasch-Modellprüfung in Bezug auf das Teilungskriterium *Geschlecht* erbrachte der LR-Test gleichfalls ein signifikantes Ergebnis. Daher ist auch hier die Hypothese H_0-2 : „Die verbliebenen Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit nicht“, zunächst zu verwerfen. Die Tabelle 23 zeigt die asymptotisch χ^2 -verteilten Testgrößen des LR-Tests vom ersten Berechnungsdurchgang des Teilungskriteriums *Geschlecht*.

Tabelle 23: UT „Funktionen Abstrahieren“ LR-Test für das TK Geschlecht, erster Berechnungsdurchgang

χ^2-Wert empirisch	df	p-Wert	χ^2-Wert kritisch
110.76	66	0.001	95.63

Die Abbildung 13 zeigt den dazugehörigen graphischen Modelltest für das Teilungskriterium *Geschlecht*.

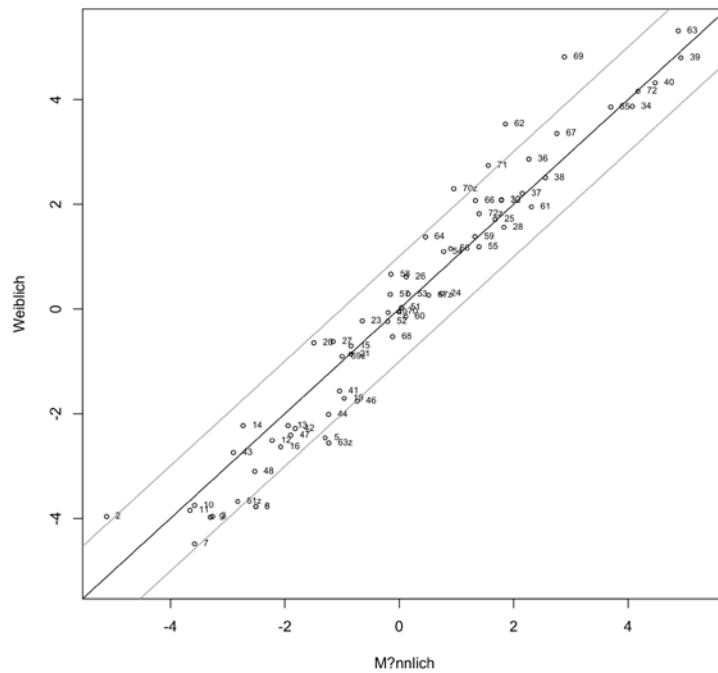


Abbildung 13: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Geschlecht

Die Tabelle 24 zeigt die „Statistik“ der Personen und Items. Siehe dazu Abschnitt „Teilungskriterium Rohscore“ Kapitel 10.1.1. Beim ersten Berechnungsdurchgang dieses Teilungskriteriums wurde 1 Item von den Analysen ausgeschlossen.

Tabelle 24: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Geschlecht, erster Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh41, bh176, bh179, bh190, bh200,	bh45, i14, bh178, bh188, bh191, bh205,	-	-	i8, i5, i63z, i2, i70z, i71, i62, i69

bh214, bh222,
 ng47, ng49,
 ng64, ng110,
 ng122, mh45,
 mh92, mh115

Teilungskriterium Muttersprache

Der LR-Test wird beim Teilungskriterium *Muttersprache* ebenfalls signifikant, aus diesem Grund muss auch diese Hypothese H_0-3 : „Die verbleibenden Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendlichen auf Grund ihrer Muttersprache nicht“ verworfen werden (Testgrößen siehe Tabelle 25).

Tabelle 25: UT „Funktionen Abstrahieren“ LR-Test für das TK Muttersprache, erster Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
155.37	62	0.001	90.80

Abbildung 14 zeigt den graphischen Modelltest für das Teilungskriterium *Muttersprache*.

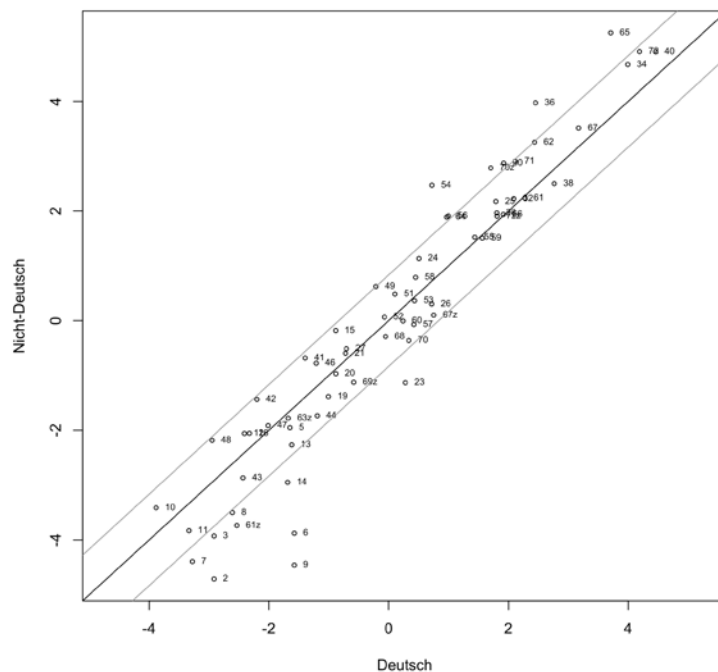


Abbildung 14: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Muttersprache

In der Tabelle 26 wird die „Statistik“ der Personen und Items, welche vom Berechnungsdurchgang von vornherein ausgeschlossen wurden angeführt. Siehe dazu Abschnitt „Teilungskriterium Rohscore“ Kapitel 10.1.1. Bei diesem Berechnungsdurchgang wurden 5 Items von den Analysen ausgeschlossen.

Tabelle 26: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Muttersprache, erster Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell- konforme Items
bh41, bh45, i14		i63, i69, i37, -		i2, i7, i3, i61z,
bh176, bh178,		i39		i9, i6, i14, i23,
bh179, bh188,				i54, i36, i65
bh190, bh191,				

bh200, bh205,
 bh214, bh222,
 ng47, ng49,
 ng64, ng110,
 ng122, mh45,
 mh92, mh115

Teilungskriterium Alter

In Bezug auf das Teilungskriterium *Alter* fällt der LR-Test bei der Rasch-Modell-Prüfung nicht signifikant aus. Hier kann die Hypothese H_0 -4: „Die verbleibenden Items des Untertests Funktionen Abstrahieren benachteiligen Kinder bzw. Jugendliche aufgrund ihres Alters nicht“ von Beginn an angenommen werden. Die Tabelle 27 zeigt ganz deutlich, dass der empirische χ^2 -Wert unter dem kritischen χ^2 -Wert liegt.

Tabelle 27: UT „Funktionen Abstrahieren“ LR-Test für das TK Alter, erster Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
35.87	29	0.001	49.59

In der Abbildung 15 ist der graphische Modelltest für dieses Teilungskriterium angegeben.

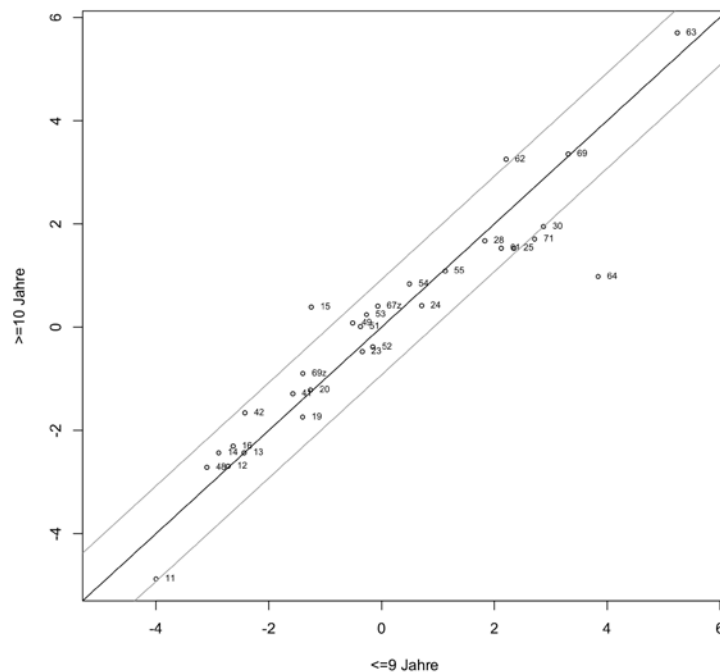


Abbildung 15: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Alter

Die Tabelle 28 zeigt die Statistik der Personen und Items, welche für dieses Teilungskriterium von den Analysen ausgeschlossen wurden. Siehe dazu Abschnitt „Teilungskriterium Rohscore“ im Kapitel 10.1.1. 38 Items sind bei diesem Durchgang ausgeschlossen worden und drei Items entsprachen nach dem Berechnungsdurchgang nicht dem Rasch-Modell.

Tabelle 28: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Alter, erster Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell- konforme Items
bh41, bh45,	i14	i32, i34, i36, -		i15, i62, i64
bh176, bh178,		i37, i38, i39,		
bh179, bh188,		i40, i56, i57,		
bh190, bh191,		i58, i59, i60,		

bh200, bh205,	i65, i66, i67,
bh214, bh222,	i70z, 72, i72z,
ng47, ng49,	i21, i26, i27, i2,
ng64, ng110,	i3, i5, i6, i7, i8,
ng122, mh45,	i9, i10, i43, i44,
mh92, mh115	i46, i47, i68,
	i63z, i70, i61z

Ausschluss nicht modellkonformer Items

Bei diesem Untertest geschah das Ausscheiden der nicht modellkonformen Items nach der gleichen Vorgehensweise wie beim vorhergehenden Untertest „Angewandtes Rechnen“. Eine Übersicht über die in den 16 Berechnungsdurchgängen ausgeschiedenen 15 Items findet sich im Anhang. Sieben Items sind ausgeschieden worden, weil sie im Kriterium *Rohscore* auffällig waren und fünf Items weil sie im Kriterium *Sprache* auffällig wurden. Eines wurde auffällig im Kriterium *Geschlecht* und zwei Items sind ausgeschieden worden, weil sie in jeweils zwei Teilungskriterien auffällig wurden, das eine in den Kriterien *Rohscore* und *Sprache* und das andere in den Kriterien *Geschlecht* und *Sprache*. Auch hier soll exemplarisch nach den Analysen der nicht modellkonformen Items inhaltlich auf einige wenige eingegangen werden. So wurde das Item 6: „Was ist das Gemeinsame an *Fisch* und *Schiff*“ im Kriterium *Sprache* auffällig. Es ist grundsätzlich eher ein leichtes Item, fällt aber unverständlicher Weise deutschsprachigen Kinder unverhältnismäßig schwerer als nicht deutschsprachigen Kindern und misst dem zur Folge nicht eindimensional. Ein Grund dafür könnte sein, dass deutschsprachige Kinder die sehr ähnlich klingenden Wörter eher verwechseln und dadurch eine Art Verwirrung entsteht und sie unfähig sind diese beiden Begriffe zu abstrahieren und nicht deutschsprachige Kinder durch das Erlernen der Vokabeln einen anderen Zugang zu diesen beiden Wörtern besitzen und ihnen das Abstrahieren dieser leichter fällt. Aber eher wahrscheinlich sind die Umstände, dass einerseits dieses Item nicht deutschsprachigen Kindern leichter fällt als deutschsprachigen Kindern, die bei diesem Item nach einer komplizierten Lösung suchen, als nur die relativ anspruchslose Antwort zu geben: „schwimmen; im Wasser“. Andererseits ist dieses Item eine Aufgabe aus dem zweiten Block und nur sechs- bis siebenjährigen Kindern, die nur 0 oder 1 Item aus dem dritten Block gelöst haben, werden diese Items vorgelegt und somit wurde

dieses Item sehr selten vorgegeben und die Lösungshäufigkeit ist dem zur Folge sehr gering. Wenn dieses Item z.B. insgesamt nur von drei Kindern richtig gelöst worden ist, wovon ein Kind deutschsprachig ist und zwei Kinder nicht deutschsprachig, dann ergibt sich eine doppelt so große Lösungshäufigkeit für nicht deutschsprachige Kinder und dies verzerrt letztendlich die Sicht auf das Item. Ein weiteres Beispiel für ein nicht modellkonformes Item ist das Item 71: „Was ist das Gemeinsame an *Diesel* und *Kerosin*“. Dieses Item wurde im Kriterium *Geschlecht* auffällig und verstößt wie das vorhergehende gegen die Eindimensionalität der Messung. Dieses Item fällt weiblichen Testpersonen unverhältnismäßig schwerer als männlichen Testpersonen, damit wird neben der Fähigkeit Funktionen zu abstrahieren und zu einer Begriffsbildung zu kommen auch das Geschlecht mit erfasst, ein Umstand, der in diesem Test nicht vorgesehen ist. Im Antwortkatalog sind als mögliche Antworten: „Kraft-/Treibstoff; Erdölerzeugnis“ angeführt. Die Begriffe können eher dem technischen Bereich zugeordnet werden, mit welchem sich Jungen häufiger beschäftigen als Mädchen. Ein weiteres nicht modellkonformes Item ist das Item 72: „Was ist das Gemeinsame an *Viren* und *Läuse*“. Dieses Item wurde im Teilungskriterium *Rohscore* auffällig. Die Antwort: „Parasiten; ernähren sich aus anderen Organismen“ fällt demnach Testpersonen mit niedrigem Score unverhältnismäßig schwerer als Testpersonen mit einem hohen Score, was bedeutet, dass diese Testleistung nicht zu einem numerischen Testwert verrechnet werden sollte und dem zur Folge auszuschneiden ist.

10.2.2. Letzter Berechnungsdurchgang

Teilungskriterium Rohscore

Nach 16 Berechnungsdurchgängen und dem Ausschluss von 15 Items wurde der LR-Test in Bezug auf das Teilungskriterium *Rohscore* nicht signifikant. Dem zur Folge liegt der empirische χ^2 -Wert unter dem kritischen χ^2 -Wert ($\alpha = .01$). Die Hypothese H_0 -1: „Die Items des Untertest „Funktionen Abstrahieren“ sind raschkonform“ kann zu diesem Zeitpunkt angenommen werden. In der Abbildung 16 ist der graphische Modelltest für dieses Teilungskriterium angezeigt. Einige wenige Items (i5, i69z, i62) sind von der 45°-Geraden etwas weiter entfernt, die Mehrzahl der Items streuen jedoch um die 45°-Gerade. In der Tabelle 30 ist die Statistik der Personen und Items zu sehen, welche von vornherein ausgeschlossen wurden.

Tabelle 29: UT „Funktionen Abstrahieren“ LR-Test für das TK Rohscore, letzter Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
69.69	50	0.001	76.15

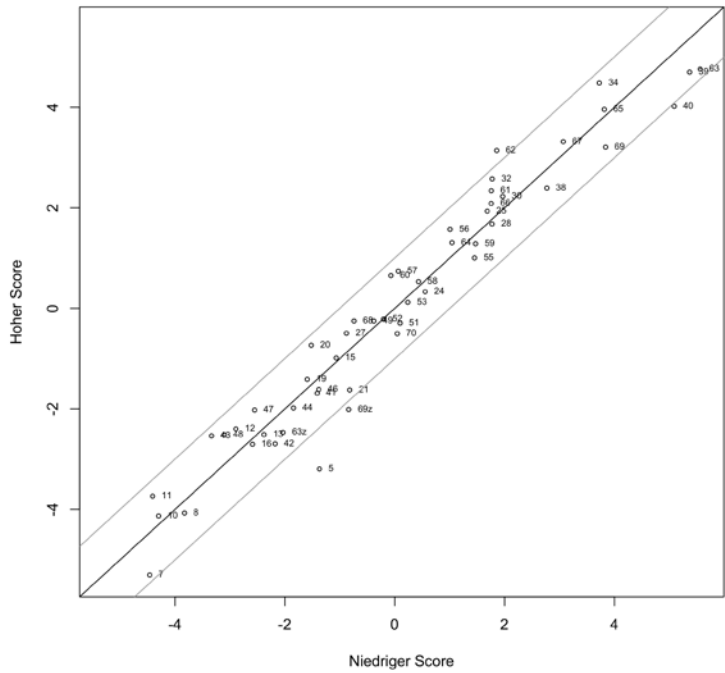


Abbildung 16: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Rohscore, letzter Berechnungsdurchgang

Tabelle 30: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Rohscore, letzter Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh41, bh176, bh179,	bh45, i14, bh178, bh188,	-	-	i5, i62, i69z

bh190, bh191,
 bh200, bh205,
 bh214, bh222,
 ng47, ng49,
 ng64, ng110,
 ng122, mh45,
 mh92, mh115

Teilungskriterium Geschlecht

Nach Ausschluss der nicht modellkonformen Items ergibt der LR-Test für das Teilungskriterium *Geschlecht* ein nicht signifikantes Ergebnis. Der empirische χ^2 -Wert liegt unter dem kritischen χ^2 -Wert, was in Tabelle 31 ersichtlich ist. Das bedeutet, dass für die Items in Bezug auf das Teilungskriterium *Geschlecht* das Rasch-Modell gilt. Demnach kann die Hypothese H_0 -2: „Die verbliebenen Items des Untertests Funktionen Abstrahieren benachteiligen Kinder bzw. Jugendliche in Bezug auf ihre Geschlechtszugehörigkeit nicht“ angenommen werden. In der Abbildung 17 wird der graphische Modelltest angezeigt und in der Tabelle 32 die Statistik der Items.

Tabelle 31: UT „Funktionen Abstrahieren“ LR-Test für das TK Geschlecht, letzter Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
73.40	51	0.001	77.39

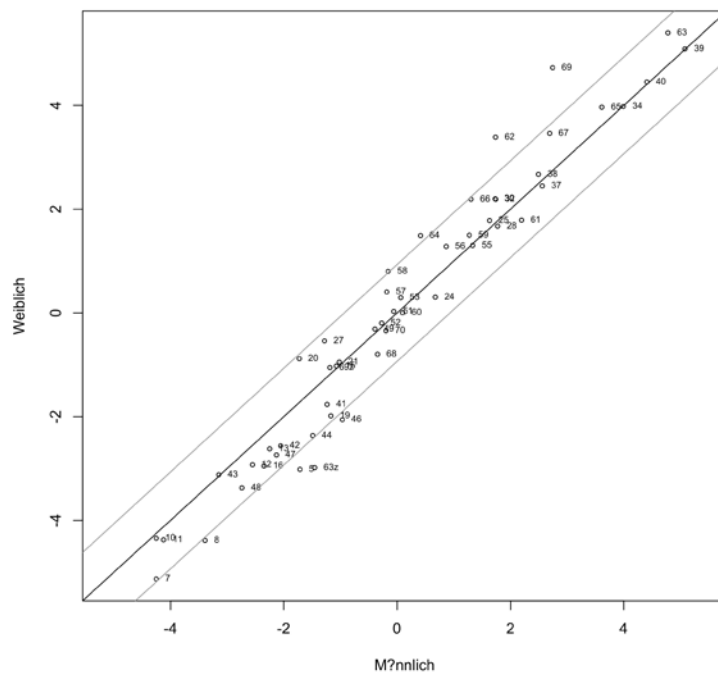


Abbildung 17: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Geschlecht, letzter Berechnungsdurchgang

Tabelle 32: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Geschlecht, letzter Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh41, bh176, bh179, bh190, bh200, bh214, ng47, ng64, ng122, mh92, mh115	bh45, i14, bh178, bh188, bh191, bh205, bh222, ng49, ng110, mh45, mh115	-	-	i5, i63z, i46, i69, i62, i64

Teilungskriterium Muttersprache

Der LR-Test für das Teilungskriterium *Muttersprache* ergibt ebenfalls ein nicht signifikantes Ergebnis. Dem zur Folge liegt der empirische χ^2 -Wert auch hier unter dem kritischen χ^2 -Wert ($\alpha = .01$). Die Hypothese H_0 -3: „Die verbleibenden Items des Untertests „Funktionen Abstrahieren“ benachteiligen Kinder bzw. Jugendlichen auf Grund ihrer Muttersprache nicht“ kann angenommen werden. Wenn man die Abbildung 18 und die Tabelle 34 heranzieht erkennt man, dass nur mehr fünf Items etwas weiter entfernt liegen von der 45°-Geraden. Jedoch identifiziert der Wald-Test keine signifikanten Items mehr.

Tabelle 33: UT „Funktionen Abstrahieren“ LR-Test für das TK *Muttersprache*, letzter Berechnungsdurchgang

χ^2 -Wert empirisch	df	p-Wert	χ^2 -Wert kritisch
68.60	47	0.001	72.44

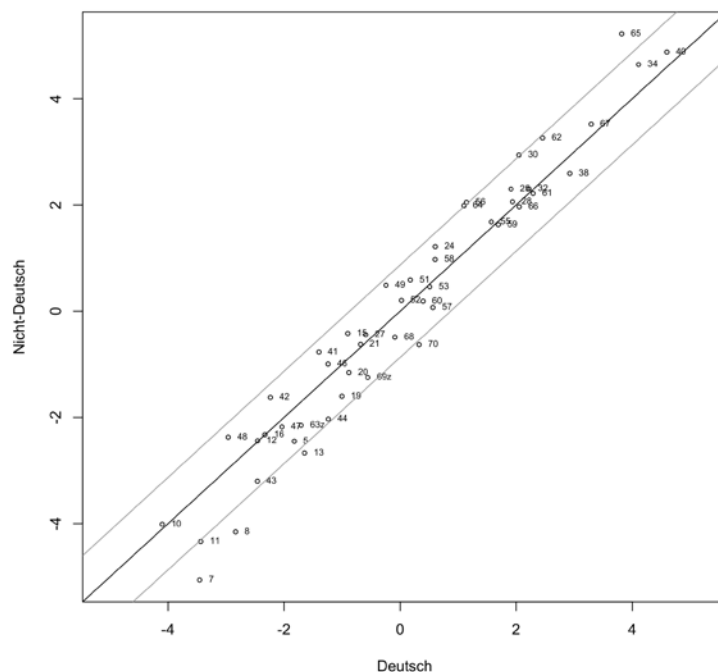


Abbildung 18: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK *Muttersprache*, letzter Berechnungsdurchgang

Tabelle 34: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Muttersprache, letzter Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht modell-konforme Items
bh41, bh45, i14 bh176, bh178, bh179, bh188, bh190, bh191, bh200, bh205, bh214, bh222, ng47, ng49, ng64, ng110, ng122, mh45, mh92, mh115		i63, i69, i37, - i39		i7, i8, i13, i70, i65

Teilungskriterium Alter

In Bezug auf das Teilungskriterium *Alter* blieb der LR-Test nach Ausschluss der oben erwähnten 14 Items nicht signifikant. Der empirische χ^2 -Wert liegt wie erwähnt unter dem kritischen χ^2 -Wert, was in Tabelle 35 zu sehen ist. Der graphische Modelltest sowie die näheren Informationen über die Items sind in der Abbildung 19 und in der Tabelle 36 ersichtlich.

Tabelle 35: UT „Funktionen Abstrahieren“ LR-Test für das TK Alter, letzter Berechnungsdurchgang

χ^2-Wert empirisch	df	p-Wert	χ^2-Wert kritisch
34.73	24	0.001	42.98

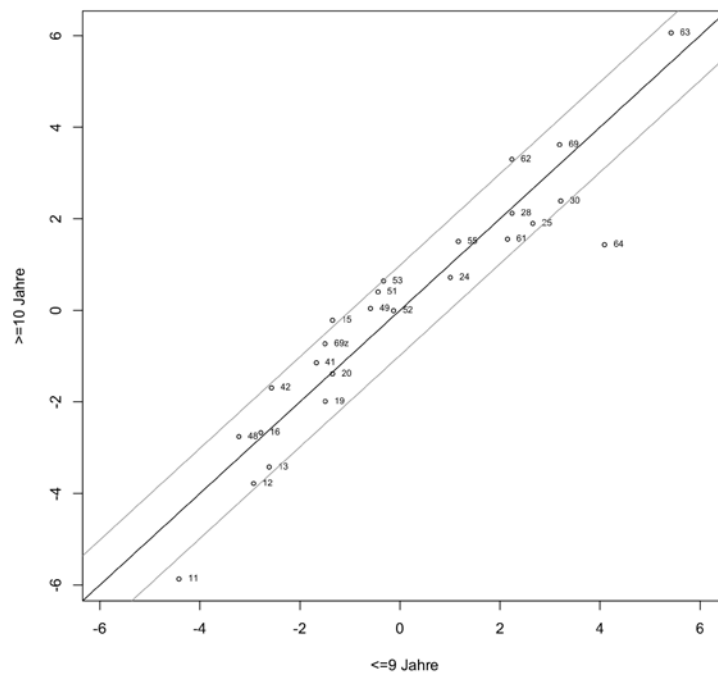


Abbildung 19: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Alter, letzter Berechnungsdurchgang

Tabelle 36: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Alter, letzter Berechnungsdurchgang

ausgeschlossene Personen	ill-conditioned Items	inappropriate response patterns	immer gelöst	nicht konforme Items
bh41, bh45, i14		i32, i34, i36, -		i11, i15, i64
bh176, bh178,		i37, i38, i39,		
bh179, bh188,		i40, i56, i57,		
bh190, bh191,		i58, i59, i60,		
bh200, bh205,		i65, i66, i67,		
bh214, bh222,		i72z, i21, i27,		
ng47, ng49,		i3, i5, i7, i8,		
ng64, ng110,		i10, i43, i44,		
ng122, mh45,		i46, i47, i68,		
mh92, mh115		i63z, i70		

11. Diskussion und Ausblick

Die Ergebnisse der Modelltests für die Untertests „Angewandtes Rechnen“ und „Funktionen Abstrahieren“ werden in diesem Kapitel zusammenfassend berichtet, einschließlich einer Diskussion über diese mit einem kurzen Ausblick auf kommende Untersuchungen.

Weil der AID eine Testbatterie ist, die sich zum adaptiven Testen eignet, musste bei den Untertests von einem großen Itempool ausgegangen werden. Viele Items wurden dabei vom AID 2 übernommen. Die Items, welche für den AID 3 überarbeitet und teilweise neu entwickelt wurden, haben von den 70 zur Verfügung stehenden Items des Untertests „Angewandtes Rechnen“ und den 69 zur Verfügung stehenden Items des Untertests „Funktionen Abstrahieren“, einen eher relativ geringen Prozentsatz ausgemacht. Für den Untertest „Angewandtes Rechnen“ wurden 16 Items neu konstruiert, wovon 7 Items den bekannten Blöcken zugeordnet und 9 Items auf drei zusätzliche Blöcke aufgeteilt wurden. Für den Untertest „Funktionen Abstrahieren“ wurden 18 Items neu konstruiert, von denen wurden 9 Items den bekannten Blöcken zugeordnet und die anderen 9 sind auf drei Zusatzblöcke verteilt worden. Auf Grund des adaptiven Testdesigns, dem *branched-testing*, wurden Untergruppen von Personen jeweils andere Untermengen von Aufgaben vorgegeben (nähere Ausführungen siehe Abschnitt 4.2.).

Für die Analysen nach dem Rasch-Modell, wurden anhand der gewonnenen Daten für die zwei Untertests unter Heranziehung des Likelihood-Quotienten-Tests von Anderson sukzessive solange Items ausgeschlossen, bis sich hinsichtlich aller Teilungskriterien (*Rohscore*, *Geschlecht*, *Muttersprache* und *Alter*) keine signifikanten Modellabweichungen mehr feststellen lassen konnte, wovon aber zwei Fälle auszunehmen sind, bei denen der Likelihood-Quotienten-Test immer noch signifikant blieb.

Im ersten Berechnungsdurchgang erbrachten die Likelihood-Quotienten-Tests durchgehend bis auf den Untertest „Funktionen Abstrahieren“ für das Teilungskriterium *Alter* hohe, folglich signifikante Ergebnisse. Die Analysen für das Teilungskriterium *Alter* laut Kubinger (2009) sollten für beide Untertests aber eher nicht durchgeführt werden, da bei ihnen von vornherein zu viele Items, wegen der altersgemäßen Testeinstiege und der daraus resultierenden Teilstichproben, ausgeschieden wurden. Der Vollständigkeit halber werden sie und die errechneten Testgrößen jedoch angegeben.

Im letzten Berechnungsdurchgang erbrachte der Likelihood-Quotienten-Test im Untertest „Angewandtes Rechnen“ nach Ausschluss von 11 Items zwei signifikante (TK *Rohscore* und TK *Sprache*) und zwei nicht signifikante (TK *Geschlecht* und TK *Alter*) Testgrößen und der Untertest „Funktionen Abstrahieren“, nach dem Ausschluss von 15 Items keine signifikanten Testgrößen über alle Teilungskriterien (TK *Rohscore*, TK *Geschlecht*, TK *Sprache* und TK *Alter*). In Bezug auf den Untertest „Angewandtes Rechnen“, bei dem beim Teilungskriterium *Rohscore* eine extrem hohe Abweichung zwischen Testgröße und kritischem Wert gegeben ist und zu interpretieren wäre, dass der Summenscore kein faires Maß ist und global keine Modellgültigkeit angenommen werden dürfte, zeigt jedoch der graphischen Modelltests relativ gute Modellgeltung. Als Erklärung hierzu ist anzuführen, dass beliebig kleine Unterschiede bei der Modellprüfung, egal welches Teilungskriterium betreffend, für hinreichend große Stichproben signifikant werden, ohne dass diese Unterschiede praktisch, das heißt inhaltlich relevant sein müssen. Dem zur Folge laufen Signifikanztests zum Rasch-Modell Gefahr, an der inhaltlichen Relevanz vorbeizugehen. „Bei Modelltest zum Rasch-Modell ist also der Stichprobenumfang regelmäßig so groß“ (Kubinger, 2009; S.44), dass die Wahrscheinlichkeit für einen Fehler 2. Art (die Null-Hypothese irrtümlich beizubehalten) nicht erkannt wird. Für die üblichen statistischen Prüfverfahren gibt es vielfältige Möglichkeiten den notwendigen Stichprobenumfang vorweg zu bestimmen, jedoch scheitern diese Möglichkeiten an der statistischen Prüfgröße beim Likelihood-Quotienten-Test, da sie Freiheitsgrade besitzt, welche unabhängig vom Stichprobenumfang sind. So muss beim Likelihood-Quotienten-Test in Bezug auf den Stichprobenumfang und die Abweichungen der Parameterschätzung auf Erfahrungswerte zurückgegriffen werden (Kubinger, 2009).

Da schon bei der Erstellung der Items auf Aktualität, wörtliche Formulierung, der altersgemäßen Passung und dergleichen geachtet wurde, sind vor allem solche Items ausgeschieden worden, welche im Wald-Test und dem graphischen Modelltest extreme Abweichungen zeigten. Natürlich kam aber auch a posteriori eine inhaltliche Analyse der Items nicht zu kurz. In den Modelltests der beiden beschriebenen Untertests weichen nach dem Ausschluss der Items nur mehr sehr wenige Aufgaben von der 45°-Geraden ab. Die selbst nach dem letzten Berechnungsdurchgang immer noch signifikanten Testgrößen sind beim Teilungskriterium *Rohscore* des Untertestes „Angewandtes Rechnen“ aus bereits erwähnten Gründen auf praktisch nicht so

relevante Abweichungen zurückzuführen, zumal ab dem achten Berechnungsdurchgang Items auffällig geworden sind, die es vorher nicht waren.

Zu erwähnen ist weiters, dass der Likelihood-Quotienten-Test des Untertests „Angewandtes Rechnen“ für das Teilungskriterium *Muttersprache* signifikant bleibt und erst im vorletzten Berechnungsdurchgang beim Untertest „Funktionen Abstrahieren“ nicht signifikant geworden ist. Gleichwohl der graphische Modelltest eine gute Modellpassung zeigt und die nunmehrigen Abweichungen auf inhaltlich nicht so relevante Abweichungen zurückzuführen sein sollten, soll auf mögliche Gründe dafür eingegangen werden. Auf Grund dessen, dass die Stichprobe doch zu gut einem Drittel aus Kindern und Jugendlichen zusammengesetzt war, deren Muttersprache eine andere als Deutsch ist, war es für diese Testpersonen sicherlich schwieriger die Instruktion beim Untertest „Funktionen Abstrahieren“ bzw. die Aufgaben des Untertests „Angewandtes Rechnen“ zu verstehen und umzusetzen. Zum Großteil wurden dennoch von den Testleitern beide Untertests vorgegeben und zu Testwerten verrechnet, wobei die Testwerte dann kein faires Maß darstellen, da mit ihnen neben der zu erfassenden intendierten Fähigkeit die Muttersprache der jeweiligen Testperson mit erfasst wird. Dieser Umstand könnte im unterschiedlichen Spracherwerb liegen, so könnte es sein, dass nicht deutschsprachige Kinder über einen qualitativ anderen Wortschatz verfügen als deutschsprachige Kinder. Es sollte dabei auch berücksichtigt werden, wann ein Kind eine zweite Sprache lernt, also ob das Kind von Anfang an zweisprachig aufwächst oder es die zweite Sprache erst nach dem Erwerb der Muttersprache lernt.

Im Vorhinein war die Vermutung über die Testergebnisse eine andere. Es war zu vermuten, dass der Likelihood-Quotienten-Test in Bezug auf das Teilungskriterium *Rohscore* beim Untertest „Angewandtes Rechnen“ eher nicht signifikant ausfallen wird, da im Antwortkatalog, im Gegensatz zum Untertest „Funktionen Abstrahieren“ eindeutige Antworten für die Items, nämlich die Lösungen der Rechenaufgaben definiert sind. Eingetroffen ist vielmehr, mit den daraus resultierenden Konsequenzen, wie bereits berichtet, dass der LR-Test beim Teilungskriterium *Rohscore* des Untertests „Angewandtes Rechnen“ selbst nach dem Ausschluss von den Items signifikant blieb. Obwohl der Untertest „Funktionen Abstrahieren“ in diesem Teilungskriterium nicht signifikant ausgefallen ist, soll trotz alledem kurz auf den Antwortkatalog dieses Untertests eingegangen werden. Obwohl bei der Itemkonstruktion sehr sorgfältig gearbeitet wurde, ist nach der Testanwendung und dem Zusammentragen der

Rückmeldungen der TestleiterInnen zu erwähnen, dass es bei dem freien Antwortformat dieses Untertestes manchmal für die Testleiter sehr schwierig war, sich an die Instruktion, nur die Antworten als richtig zu werten, die auch im Antwortkatalog aufgeführt sind, zu halten. Denn die Kinder produzierten manchmal sehr einfache und naheliegende und manchmal auch weiter hergeholte Antworten, welche ebenfalls die gemeinsame Funktion der genannten Begriffe beschreibt. Wie bei dem Item: „Was ist das Gemeinsame an *Scheidung* und *Tod*“; im Antwortkatalog enthalten sind die Antworten: „Trennung (von Menschen); weggehen; etwas endet“. Als Antworten von den Kindern kamen unter anderem: „traurig; verlassen werden; man ist allein“. Da von manchen Testleitern die nicht im Antwortkatalog enthaltenen Antworten als richtig gewertet wurden und von anderen Testleitern nicht, ist dieser Test nicht verrechnungssicher und der Antwortkatalog sollte einer eingehenden Überprüfung unterzogen werden.

Zu vermuten war weiters, dass beide Untertests in Bezug auf das Teilungskriterium Muttersprache signifikant ausfallen, da diese einen hohen sprachlichen Anteil haben. Dies wurde aber nur beim Untertest „Angewandtes Rechnen“ bestätigt, was sicherlich unter anderem auch daran liegt, dass dieser viel sprachlastiger ist als der Untertest „Funktionen Abstrahieren“. Interessant wäre, ob ein Test zum Rechnen (hier müsste der Umstand des mathematischen Schlussfolgerns außen vor gelassen werden), der vom Inhalt die gleichen Aufgaben enthält aber nur aus reinen Rechenoperationen besteht, die gleichen Ergebnisse liefern würde.

Unter Berücksichtigung der resultierenden Ergebnisse aus dieser Rohfassung zum AID 3 sollte bei weiterführenden Untersuchungen darauf geachtet werden, dass bei der Stichprobenzusammenstellung noch mehr deutschsprachige Kinder und Jugendliche teilnehmen und diese von der Verteilung her in Bezug auf das Alter noch homogener sind. Weiters sollte der Antwortkatalog des Untertestes „Funktionen Abstrahieren“ einer Prüfung unterzogen werden, um ihn noch verrechnungssicherer zu machen. Auch könnte die Testanweisung dieses Untertests dahingehend überarbeitet werden, dass der/die TestleiterIn eine eindeutige Anweisung bekommt, wie sie in Bezug auf die Verrechnung mit den Wortpaaren zu verfahren hat, die der Testperson erklärt werden, sofern sie die Items nicht oder falsch beantwortet. Weiters könnten die wörtlichen Formulierungen der Rechenaufgaben des Untertests „Angewandtes Rechnen“ noch weiter vereinfacht oder in rein arithmetische Rechnungen umgewandelt werden.

Grundsätzlich lässt sich aber sagen, dass die Aktualisierung und die Revision der Aufgaben zu den Untertests „Angewandtes Rechnen“ und „Funktionen Abstrahieren“ gelungen sind. Die Items beider Untertests decken einen breiten Fähigkeitsbereich ab und weisen dennoch eine hohe testtheoretische Güte in Bezug auf die verwendeten Teilungskriterien auf. Sowohl der Untertest „Angewandtes Rechnen“ als auch der Untertest „Funktionen Abstrahieren“ finden bei den Kindern und Jugendlichen, die getestet werden, hohe Akzeptanz. Besonders gut gelungen beim Untertest „Angewandtes Rechnen“ die visuelle Unterstützung der Items durch Bilder (für jüngere Kinder).

12. Zusammenfassung

Die Zielsetzung dieser Arbeit bestand in der Revision und Aktualisierung der Untertests „Angewandtes Rechnen“ und „Funktionen Abstrahieren“ des AID 2 (Version 2.2) in Bezug auf ihre Items. Anlass dafür waren einerseits Rückmeldungen von praktisch tätigen PsychologInnen, dass teilweise die Zusammenstellung der Items in den Aufgabengruppen nicht optimal war und andererseits ging es um eine inhaltliche Überarbeitung der einzelnen Aufgaben. Weiteres fordert die DIN 33430 spätestens alle acht Jahre hinsichtlich der Auswahl, Zusammenstellung, Durchführung, Auswertung und Interpretation eine Überprüfung der Gültigkeit der Normwerte von Verfahren.

Viele Items wurden aus dem AID 2 übernommen, es mussten lediglich Aufgaben aktualisiert werden, so dass sie jene Fähigkeiten und jenes Wissen überprüfen, welches in der heutigen Zeit tatsächlich relevant ist. Beim Untertest „Angewandtes Rechnen“ war die Zusammenstellung der Aufgabengruppen problematisch, weil in ihnen manchmal Items enthalten waren, welche die Kinder aufgrund ihres Alters und dem damit noch nicht erlernten Schulstoff nicht lösen konnten. In einigen Aufgabenblöcken waren mehrere Items in Bezug auf ihre Rechenoperationen nahezu gleich, die zu ersetzen waren. Bei anderen Items wiederum war die Rechenaufgabe in einem zu komplizierten und zu langen Text eingekleidet. Weiters fehlten leichte Items für jüngere Kinder. Beim Untertest „Funktionen Abstrahieren“ sind veraltete Wörter wie z.B. „Schreibmaschine“ oder „Weinpresse“ ausgetauscht und durch neue ersetzt worden, manche Items wurden ganz entfernt. Es wurden daraufhin, aufbauend auf die recherchierte Literatur, das in der heutigen Zeit relevante Wissen, sowie bezogen auf den Untertest „Angewandtes Rechnen“ angepasst an den schulischen Lehrplänen, neue Items konstruiert, welche in die Aufgabenblöcke eingefügt wurden. Im Zuge des Konstruktionsprozesses wurden für den Untertest „Angewandtes Rechnen“ 16 neue Items und für den Untertest „Funktionen Abstrahieren“ 18 neue Items entwickelt. Die resultierenden 69 bzw. 70 Items pro Untertest wurden den 18 Blöcken und zusätzlichen 3 Blöcken mit je drei Items zugeordnet. Die gesamte Rohversion des AID 3 bestand insgesamt aus 10 Untertests, welche von mehreren DiplomandInnen einer Stichprobe von 711 Kindern und Jugendlichen im Alter von 6 bis 16 Jahren vorgegeben wurde. Das Geschlechterverhältnis der Stichprobe war relativ ausgeglichen und das Alter war normalverteilt. Den größten Anteil der Stichprobe machten Volksschüler aus, gefolgt

von Gymnasiasten. Ein Drittel der Kinder und Jugendlichen hatte Deutsch nicht als Muttersprache.

Beide Untertests wurden hinsichtlich der Teilungskriterien *Rohscore*, *Geschlecht*, *Muttersprache* und *Alter* auf die Geltung des Rasch-Modells überprüft, was zwingend notwendig war, da sowohl der Untertest „Angewandtes Rechnen“ als auch der Untertest „Funktionen Abstrahieren“ im AID 3, wie auch in seinen vorangegangenen Revisionen, adaptiv nach dem *branched-testing* vorgegeben werden. Nach dem Ausschluss von 11 Items beim Untertest „Angewandtes Rechnen“ konnte für diesen a posteriori die Geltung des Rasch-Modells in Bezug auf die Teilungskriterien *Geschlecht* und *Alter* angenommen werden, aber nicht in Bezug auf die Teilungskriterien *Rohscore* und *Muttersprache*. In Bezug auf diese Teilungskriterien ist anzumerken, dass der graphische Modelltest eine relativ gute Modellpassung zeigt und dass beliebig kleine Unterschiede bei der Modellprüfung für hinreichend große Stichproben signifikant werden, ohne dass diese Unterschiede inhaltlich relevant sein müssen. Beim Untertest „Funktionen Abstrahieren“ konnte nach dem Ausschluss von 14 Items in Bezug auf alle Teilungskriterien *Rohscore*, *Geschlecht*, *Muttersprache* und *Alter* a posteriori die Geltung des Rasch-Modells festgestellt werden. Was bedeutet, dass das Verrechnen der Testleistungen zu Testwerten empirisch gerechtfertigt ist und Kinder bzw. Jugendliche hinsichtlich ihres Geschlechts, ihrer Muttersprache und ihres Alters nicht benachteiligt werden. Offen bleibt trotzdem die Frage inwieweit die Items Kinder bzw. Jugendliche mit nicht deutscher Muttersprache benachteiligen.

Literaturverzeichnis

- Akhurst, B.A. (1972) *Die Untersuchung intellektueller Fähigkeiten*. Weinheim und Basel: Beltz.
- Arendasy, M., Hornke, L.F., Sommer, M., Wagner-Menghin, M., Gittler, G., Heidinger, C., Herle, M., Körtner, T. (2012) *Manual Intelligenz-Struktur-Batterie (INSBAT)*. (Version 27). Mödling: Schuhfried.
- Böck, J. (2010) *AID 2 als Gruppentestung? Eine Machbarkeitsstudie*. Dipl. Arbeit, Universität, Wien.
- Brandstätter, E. (2001). *Faktorenanalyse oder Rasch-Modell? Eine Kreuzvalidierung am Beispiel des Leistungs-Motivations-Tests*. Frankfurt am Main: Peter Lang GmbH Europäischer Verlag der Wissenschaft.
- Cattell, R. B. (1987). *Intelligence: its structure, growth, and action*. The importance of intelligence. Amsterdam: Elsevier.
- Daseking, M., Janke, N. & Petermann, F. (2006). *Monatsschrift Kinderheilkunde* (4.2006). Bremen: Springer Medizin Verlag.
- Fischer, G.H. (1989). Spezifische Objektivität: Eine wissenschaftstheoretische Grundlage des Rasch-Modells. In K.D. Kubinger (Hrsg.) *Moderne Testtheorie*. (S.87-111). Weinheim: Beltz.
- Frey, A. (2006). *Validitätssteigerung durch adaptives Testen*. Frankfurt am Main: Peter Lang.
- Goswami, U. (2001). *So denken Kinder – Einführung in die Psychologie der kognitiven Entwicklung*. Bern: Huber.
- Grube, D. (2005) Entwicklung des Rechnens im Grundschulalter. In M. Hasselhorn, H. Marx, W. Schneider (Hrsg.) *Diagnostik von Mathematikleistungen. Test und Trends. N.F. Band 4 Jahrbuch der pädagogisch-psychologischen Diagnostik*. Göttingen: Hogrefe.
- Häcker, H.O. & Stapf, K.H. (Hrsg.). (2009). *Dorsch Psychologisches Wörterbuch*. (15., vollst. überarb. und erw. Aufl.). Bern: Hans Huber.

- Hagenmüller, B. (2011). *Entwicklung des Untertests „Formale Folgerichtigkeit“ zur Erfassung von Reasoning in der Intelligenz-Testbatterie AID 3*. Dipl. Arbeit, Universität, Wien.
- Hasselhorn, M. (2001). Metakognition. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 466-471). Weinheim: Beltz.
- Hellebart, M. (2013). *Revision der Untertests „Alltagswissen“ und „Soziales Erfassen und Sachliches Reflektieren“ für die Intelligenztestbatterie AID 3*. Dipl. Arbeit, Universität, Wien.
- Jacobs, C. & Petermann, F. (2007) *Rechenstörungen*. Leitfaden Kinder- und Jugendpsychologie. Göttingen: Hogrefe.
- Kubinger, K.D. (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. (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. (2006). *Psychologische Diagnostik – Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kubinger, K.D. (2009). *Adaptives Intelligenz Diagnostikum 2* (Version 2.2). (2., neu geeichte und überarb. Aufl.). Göttingen: Beltz.
- Kubinger, K.D. & Jäger, R.S. (Hrsg.). (2003). *Schlüsselbegriffe der psychologischen Diagnostik*. Weinheim: Beltz.
- Kubinger, K.D. & Wurst, E. (1985). *Adaptives Intelligenz Diagnostikum (AID)*. Weinheim: Beltz.
- Kubinger, K.D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum 2 (AID 2)*. (2., überarb. Aufl.). Göttingen: Beltz.
- Melchers, P. & Preuß, U. (2009). *Kaufman – Assessment Battery for Children – deutschsprachige Fassung*. (8., unveränd. Aufl.). Frankfurt am Main: Pearson.
- Neuhaus, W. (1962). *Der Aufbau der geistigen Welt des Kindes*. (2., überarb. u. erw. Auflage). München: Reinhardt.

- Oerter, R. & Dreher, M. (2002) Entwicklung des Problemlösens. In R. Oerter & L. Montada (Hrsg.) *Entwicklungspsychologie* (5. vollst. überarb. Auflage, S. 469 – 494). Weinheim. Beltz.
- Patrick Mair, Reinhold Hatzinger and Marco J. Maier (2012). eRm: Extended Rasch Modeling. R package Version 0.15-1. <http://CRAN.R-project.org/package=eRm>.
- Petermann, F. & Petermann, U. (2007). HAWIK-IV. *Hamburger-Wechsler-Intelligenztest für Kinder – IV*. Bern: Huber.
- Preusche, I. & Leiss, U. (2003). Intelligenztests für Kinder. Hawik-III, AID 2 und K-ABC im Vergleich. *Report Psychologie*, 28, (1), 12-26.
- R Core Team (2012). R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Rost, J. (2004). *Lehrbuch Testtheorie- Testkonstruktion*. (Zweite, vollständig überarbeitete und erweiterte Auflage). Bern: Huber.
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik*. Berlin: Springer.
- Süß, H.M. (2003). Intelligenztheorien. In K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 217-224). Weinheim: Beltz.
- Weber, B. (2011). *Konstruktion des sprachlichen Untertests „Antonyme finden“ für die Intelligenztestbatterie AID 3*. Dipl. Arbeit, Universität, Wien.
- Wechsler, D. (1944). *The measurement of adult intelligence* (3rded.) Baltimore: Williams & Wilkins

Tabellenverzeichnis

<i>Tabelle 1: Vollständige Übersicht der Subtests des AID 2 (Kubinger, 2009).....</i>	<i>15</i>
<i>Tabelle 2: Deskriptive Statistik der Variable Schulform.....</i>	<i>48</i>
<i>Tabelle 3: Deskriptive Statistik der Variable Muttersprache.....</i>	<i>49</i>
<i>Tabelle 4: Deskriptive Statistik der Variable Alter und Geschlecht</i>	<i>50</i>
<i>Tabelle 5: UT „Angewandtes Rechnen“ LR-Test für das TK Rohscore, erster Berechnungsdurchgang</i>	<i>54</i>
<i>Tabelle 6: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Rohscore, erster Berechnungsdurchgang.....</i>	<i>55</i>
<i>Tabelle 7: UT „Angewandtes Rechnen“ LR-Test für das TK Geschlecht, erster Berechnungsdurchgang</i>	<i>55</i>
<i>Tabelle 8: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Geschlecht, erster Berechnungsdurchgang</i>	<i>56</i>
<i>Tabelle 9: UT „Angewandtes Rechnen“ LR-Test für das TK Muttersprache, erster Berechnungsdurchgang</i>	<i>57</i>
<i>Tabelle 10: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Muttersprache, erster Berechnungsdurchgang</i>	<i>58</i>
<i>Tabelle 11: UT „Angewandtes Rechnen“ LR-Test für das TK Alter, erster Berechnungsdurchgang</i>	<i>58</i>
<i>Tabelle 12: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Alter, erster Berechnungsdurchgang</i>	<i>59</i>
<i>Tabelle 13: UT „Angewandtes Rechnen“ LR-Test für das TK Rohscore, letzter Berechnungsdurchgang</i>	<i>62</i>
<i>Tabelle 14: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Rohscore, letzter Berechnungsdurchgang</i>	<i>63</i>
<i>Tabelle 15: UT „Angewandtes Rechnen“ LR-Test für das TK Geschlecht, letzter Berechnungsdurchgang</i>	<i>64</i>
<i>Tabelle 16: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Geschlecht, letzter Berechnungsdurchgang.....</i>	<i>65</i>

Tabelle 17: UT „Angewandtes Rechnen“ LR-Test für das TK Muttersprache, letzter Berechnungsdurchgang	65
Tabelle 18: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Muttersprache, letzter Berechnungsdurchgang.....	66
Tabelle 19: UT „Angewandtes Rechnen“ LR-Test für das TK Alter, letzter Berechnungsdurchgang	67
Tabelle 20: UT „Angewandtes Rechnen“ Statistik über Personen und Items für das Teilungskriterium Alter, letzter Berechnungsdurchgang	68
Tabelle 21: UT „Funktionen Abstrahieren“ LR-Test für das TK Rohscore, erster Berechnungsdurchgang	69
Tabelle 22: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Rohscore, erster Berechnungsdurchgang.....	70
Tabelle 23: UT „Funktionen Abstrahieren“ LR-Test für das TK Geschlecht, erster Berechnungsdurchgang	71
Tabelle 24: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Geschlecht, erster Berechnungsdurchgang	72
Tabelle 25: UT „Funktionen Abstrahieren“ LR-Test für das TK Muttersprache, erster Berechnungsdurchgang	73
Tabelle 26: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Muttersprache, erster Berechnungsdurchgang	74
Tabelle 27: UT „Funktionen Abstrahieren“ LR-Test für das TK Alter, erster Berechnungsdurchgang	75
Tabelle 28: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Alter, erster Berechnungsdurchgang	76
Tabelle 29: UT „Funktionen Abstrahieren“ LR-Test für das TK Rohscore, letzter Berechnungsdurchgang	79
Tabelle 30: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Rohscore, letzter Berechnungsdurchgang	79
Tabelle 31: UT „Funktionen Abstrahieren“ LR-Test für das TK Geschlecht, letzter Berechnungsdurchgang	80

<i>Tabelle 32: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Geschlecht, letzter Berechnungsdurchgang.....</i>	<i>81</i>
<i>Tabelle 33: UT „Funktionen Abstrahieren“ LR-Test für das TK Muttersprache, letzter Berechnungsdurchgang</i>	<i>82</i>
<i>Tabelle 34: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Muttersprache, letzter Berechnungsdurchgang.....</i>	<i>83</i>
<i>Tabelle 35: UT „Funktionen Abstrahieren“ LR-Test für das TK Alter, letzter Berechnungsdurchgang</i>	<i>83</i>
<i>Tabelle 36: UT „Funktionen Abstrahieren“ Statistik über Personen und Items für das Teilungskriterium Alter, letzter Berechnungsdurchgang</i>	<i>84</i>

Abbildungsverzeichnis

Abbildung 1: Struktur des CHC-Modells - adaptiert nach Arendasy et al. (2012).....	11
Abbildung 2: Prinzip der Testvorgabe nach dem branched-testing-Design - AID2 Manual (Kubinger, 2009; S. 37).....	18
Abbildung 3: Itemcharakteristikkurve (ICC) nach dem dichotom logistischen Modell von Rasch für drei verschiedene Itemschwierigkeiten (Kubinger, 1989).....	25
Abbildung 4: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Rohscore	54
Abbildung 5: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Geschlecht	56
Abbildung 6: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Muttersprache.....	57
Abbildung 7: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Alter ..	59
Abbildung 8: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Rohscore, letzter Berechnungsdurchgang	63
Abbildung 9: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Geschlecht, letzter Berechnungsdurchgang	64
Abbildung 10: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Muttersprache, letzter Berechnungsdurchgang	66
Abbildung 11: UT „Angewandtes Rechnen“ graphischer Modelltest für das TK Alter, letzter Berechnungsdurchgang	67
Abbildung 12: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Rohscore	70
Abbildung 13: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Geschlecht	72
Abbildung 14: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Muttersprache.....	74
Abbildung 15: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Alter	76

<i>Abbildung 16: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Rohscore, letzter Berechnungsdurchgang</i>	<i>79</i>
<i>Abbildung 17: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Geschlecht, letzter Berechnungsdurchgang</i>	<i>81</i>
<i>Abbildung 18: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Muttersprache, letzter Berechnungsdurchgang</i>	<i>82</i>
<i>Abbildung 19: UT „Funktionen Abstrahieren“ graphischer Modelltest für das TK Alter, letzter Berechnungsdurchgang</i>	<i>84</i>

ANHANG

BRIEF AN DEN STADTSCHULRAT.....	- 3 -
ANTWORTSCHREIBEN DES STADTSCHULRATS	- 11 -
LEHRERINNENBRIEF	- 13 -
ELTERNBRIEF	- 15 -
EINVERSTÄNDNISERKLÄRUNG DER ELTERN.....	- 17 -
PROTOKOLLBOGEN 1.....	- 19 -
PROTOKOLLBOGEN 2.....	- 21 -
DANKESBRIEF AN DIE ELTERN.....	- 23 -
SCHRIFTLICHER ERGEBNISBERICHT.....	- 25 -
P-WERT-TABELLEN	- 29 -
TABELLEN AUSGESCHLOSSENER ITEMS.....	- 37 -
KONFIDENZINTERVALL-ELLIPSEN UT "ANGEWANDTES RECHNEN" 1...	- 39 -
KI-ELLIPSEN UT "ANGEWANDTES RECHNEN" 2	- 45 -
KI-ELLIPSEN UT "FUNKTIONEN ABSTRAHIEREN" 1	- 51 -
KI-ELLIPSEN UT "FUNKTIONEN ABSTRAHIEREN" 2	- 57 -

BRIEF AN DEN STADTSCHULRAT

Fakultät für Psychologie

Institut für Entwicklungspsychologie
und Psychologische Diagnostik

Liebiggasse 5

A-1010 Wien

T +43 (1) 4277-478 51

F +43 (1) 4277-479 05

stefana.holocher-ertl@univie.ac.at

Wien, am 19.10.2009

An Mag. Klaus Rick
Präsidium des Stadtschulrates für Wien
Wipplingerstrasse 28
1010 Wien

Betrifft: Schüler(innen)-Testung mit dem AID 3 – einer Revision der Intelligenztestbatterie AID 2 zur Erfassung von Intelligenz bei Kindern und Jugendlichen

Bei der Intelligenz-Testbatterie AID 2 (*Adaptives Intelligenz Diagnostikum: in der aktuellen Fassung AID 2, Version 2.2, Kubinger, 2009*) handelt es sich um ein im deutschen Sprachraum bereits seit vielen Jahren sehr etabliertes psychologisches Verfahren zur Erfassung der intellektuellen Fähigkeiten von Kindern und Jugendlichen im Alter von 6 bis 15 Jahren. Insbesondere in der Schulpsychologie und Bildungsberatung wird es sehr häufig eingesetzt. Es dient vor allem der Beratung hinsichtlich schulischer Leistungsschwierigkeiten oder der Schullaufbahnberatung und ist insofern eine wesentliche Entscheidungshilfe für Schüler(innen) und deren Eltern.

Um mit dem AID tatsächlich jenes Wissen zu überprüfen, welches in heutiger Zeit von Relevanz ist, wurden die Testaufgaben nun aktualisiert sowie um einige inhaltliche Dimensionen erweitert. Eine erste revidierte Version liegt nun fertig vor und bedarf einer Erprobung an Schülern(innen) im Alter von 6 bis 15 Jahren.

Wir ersuchen daher Schulen in Wien, Niederösterreich und Burgenland um deren Unterstützung bzw. Mitwirkung bei unserem Projekt.

Die Schüler(innen)-Testung wird von Wien aus koordiniert (Projektleitung: Univ. Prof. Dr. Mag. Klaus D. Kubinger, Projektkoordination/Ansprechpartnerin: Dr. Stefana Holocher-Ertl), es stehen darüber hinaus stets Ansprechpartner(innen) vor Ort zur Verfügung. Wir bemühen uns, den Aufwand für die jeweilige Schule so gering wie möglich zu halten.

Die teilnehmenden Schulen tragen ganz wesentlich zur Etablierung einer fundierten Leistungsdiagnostik bei. Damit sollen die Schüler(innen) in Zukunft optimal hinsichtlich verschiedener leistungsbezogener Aspekte beraten werden können, was wiederum den Schulen zugute kommt.

Es besteht die Möglichkeit – sollte das gewünscht sein –, den Schüler(innen) bzw. deren Eltern als Dankeschön einen kurzen schriftlichen Ergebnisbericht über kognitive Stärken und Schwächen des Kindes auszuhändigen.

Mit der Bitte um Unterstützung und freundlichen Grüßen,

Dr. Stefana Holocher-Ertl

Klinische- und Gesundheitspsychologin

Univ.-Ass. an der Fakultät für Psychologie

Anlagen:

Ansprechpartnerin

Informationen zum AID 3

Untersuchungsziel und –ablauf

Stichprobenzusammensetzung

Elternbrief

Lehrer(innen)brief

Ansprechpartnerin:

Projektleitung:

Projektleiter ist der Leiter des Instituts für Entwicklungspsychologie und Psychologische Diagnostik der Universität Wien, Univ. Prof. Dr. Mag. Klaus D. Kubinger.

Projektkoordination:

Als Ansprechpartnerin steht Ihnen die Projektkoordinatorin, Dr. Stefana Holocher-Ertl, zur Verfügung. Sie beantwortet alle Fragen bezüglich Inhalt und Ablauf der Schüler(innen)-Erhebung. Bitte entnehmen Sie nachfolgend die Kontaktdaten:

Dr. Stefana Holocher-Ertl

Institut für Entwicklungspsychologie und Psychologische Diagnostik
Fakultät für Psychologie, Universität Wien
1010 Wien, Liebiggasse 5

Tel 1: 01 - 4277 47851

Tel 2: 0650 - 8512188

Fax: 01 – 4277 47905

Email: stefana.holocher-ertl@univie.ac.at

Homepage: <http://psychologie.univie.ac.at/diagnostik/home/>

Informationen zum AID 3:

Das *Adaptive Intelligenz Diagnostikum* AID 2 (Kubinger, 2009) stellt seit seiner erstmaligen Veröffentlichung im Jahr 1985 (damals noch als AID) eine Intelligenz-Testbatterie für Kinder und Jugendliche (im Alter von 6-16 Jahren) dar, die aufgrund ihrer hohen Güte in der psychologisch-diagnostischen Praxis (Schulpsychologie, Diagnostik bei Lern- und Verhaltensauffälligkeiten) häufige Anwendung findet. Bearbeitet ein Kind/Jugendlicher die Aufgaben des AID 2, so können daraus wesentliche Informationen abgeleitet werden: Die verbalen, rechnerischen und visuell-gestalthaften intellektuelle Fähigkeiten des Kindes/Jugendlichen, seine akustische und visuelle Merkfähigkeit, seine Informationsverarbeitungsgeschwindigkeit (Auffassungsgabe) sowie sein Verständnis in sozialen Situationen. Des weiteren ermöglicht die Vorgabe des AID 2 die Erfassung wesentlicher Teilleistungsfähigkeiten; dabei handelt es sich um basale neurologische Funktionen, die unabhängig vom Intelligenzniveau Störungen komplexer kognitiver Vorgänge (Erlernen der Kulturtechniken) und des kindlichen Verhaltens nach sich ziehen können. Anleitungen zur Verhaltensbeobachtung während der Testung erlauben zudem Rückschlüsse über den Arbeitsstil des getesteten Kindes/Jugendlichen sowie über seine Konzentration und Aufmerksamkeit. Der Einsatz des AID 2 erfolgte bislang bei einer Vielzahl schulpsychologischer Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Verhaltensproblemen im schulischen Kontext, Abklären einer Aufmerksamkeitsdefizitstörung, Abklärung einer möglichen Legasthenie/Dyskalkulie, Schullaufbahnberatungen.

Eine Revision steht nun aus folgenden Gründen an:

1. Die zur Messung der intellektuellen Fähigkeiten eines Kindes vorgegebenen Aufgaben müssen *aktualisiert* werden, sodass jene *Fähigkeiten und jenes Wissen überprüft* wird, welches *in der heutigen Zeit tatsächlich von Relevanz* ist. So werden z.B. in demjenigen Untertest, welcher das bisher angeeignete Wissen eines Kindes überprüft, neue Wissensgebiete (wie z.B. Informationstechnologien) aufgenommen und historische und geografische Entwicklungen der letzten 25 Jahren in den Fragen berücksichtigt. Aufgaben, die den Wortschatz und das Sprachverständnis der Kinder überprüfen, werden an den heutigen Sprachgebrauch angeglichen.

2. Die Aufgaben werden je gemessene Fähigkeit dahingehend erweitert, dass sowohl *genaue und differenzierte Diagnosen* über intellektuelle Stärken und Schwächen bei *leistungsschwachen* als auch bei sehr *leistungsstarken* Kindern und Jugendlichen möglich sind.

3. Das Konzept der *Teilleistungsstörungsdiagnostik* wird durch die Entwicklung zusätzlicher Aufgaben ausgebaut, sodass mögliche Teilleistungsstörungen noch besser erkannt werden und hilfreiche Interventionsmaßnahmen geplant werden können.

4. Neben der Erfassung der *förderungsabhängigen Intelligenz* (also jener Fähigkeiten, die in ihrer Entwicklung stark von entsprechender familiärer und institutioneller Förderung abhängig sind) wird auch die Erfassung der so genannten „Grundintelligenz“, darunter versteht man die relativ *förderungsunabhängige Fähigkeit* des formal-logischen Denkens, durch die Entwicklung eines neuen Untertests möglich sein. Damit kann zwischen grundsätzlich gut begabten, aber wenig geförderten, und gut begabten und entsprechend gut geförderten Kindern unterschieden werden.

5. Zur möglichst *objektiven und gültigen Einschätzung des Arbeitsverhaltens* eines Kindes (also z.B. seiner Leistungsmotivation, Frustrationstoleranz, Ausdauer und Sorgfalt) wird die Testbatterie um Beobachtungsrichtlinien und -anleitungen für die Testleiter(innen) ergänzt.

Eine revidierte Version liegt nun entsprechend vor.

Untersuchungsablauf:

Da es sich um Einzeltestungen handelt, wird jedes Kind / jeder Jugendliche von einem Testleiter bzw. einer Testleiterin gesondert getestet. Dazu werden die Kinder/Jugendlichen von den Testleiter(innen) einzeln von der Klasse abgeholt und in einen von der Schule zur Verfügung gestellten Raum (freier Klassenraum, Bibliothek, Arzttraum etc.) gebracht, in dem die Testungen stattfinden. Der Raum muss abgesehen von zwei Stühlen und einem Tisch über keine weitere Ausstattung verfügen.

Die Testung findet während der Unterrichtszeit des Kindes statt, es wird nicht in der Freizeit (Mittagspause, nach dem Unterricht etc.) des Kindes getestet. Es kommen sowohl Vormittags- als auch Nachmittagstermine in Frage. Die Dauer der Testung entspricht ca. 2 Schulstunden, wobei diese angegebene Zeitspanne ein Durchschnittswert ist und je nach individueller Arbeitsweise der Kinder schwanken

kann. Während dieser Zeit ist es den Kindern selbstverständlich möglich – je nach individuellem Bedarf – kurze Pausen einzulegen. Nach Abschluss der Testung werden die Kinder wieder in ihre Klasse zurückgebracht.

Die teilnehmenden Schüler(innen) sollten im Alter von 6 bis 15 Jahre alt sein. (siehe die genauere Stichprobenzusammensetzung weiter unten). Natürlich werden nur Kinder getestet, deren Eltern ihr Einverständnis gegeben haben. Der Elternbrief ist in Deutsch abgefasst und enthält wichtige Informationen an die Erziehungsberechtigten.

Wenn klar ist, welche Kinder der jeweiligen Schule das Einverständnis der Eltern haben, werden Termine fixiert, an denen Testleiter(innen) zum Testen kommen dürfen. Bei der Terminvereinbarung legen wir großen Wert darauf, den Ablauf des regulären Unterrichts so wenig wie möglich zu unterbrechen. Wir werden versuchen – im Rahmen der Möglichkeiten – Rücksicht auf die Wünsche der einzelnen Schulen zu nehmen.

Testleiter(innen) sind mehrere Diplomanden(innen) des Arbeitsbereichs Psychologische Diagnostik der Fakultät für Psychologie, die selbst an der Revision des AID 3 mitgearbeitet haben und sowohl theoretisch als auch aufgrund ihrer praktischen Erfahrung besondere Eignung für die Testung der Kinder und Jugendlichen aufweisen.

Natürlich werden die gewonnenen Daten im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke verwendet, die Testleiter(innen) sind angewiesen, die Ergebnisse der Kinder zu anonymisieren. Wünschen die Eltern als Dankeschön einen kurzen schriftlichen Ergebnisbericht über Stärken und Schwächen der getesteten Schüler(innen), erfolgt die Anonymisierung erst nach der Erstellung dieses Befundes, ansonsten gleich nach Beendigung der Testung.

Stichprobenzusammensetzung:

<i>Alter in Jahren</i>	<i>Schulform</i>	<i>Schulstufe</i>	<i>Anzahl</i>	<i>aus Bundesland</i>
6-7 Jahre	VKS	1, 2	40	Wien und/oder NÖ
8-9 Jahre	VKS	3, 4	40	Wien und/oder NÖ
10 Jahre	VKS	4	20	Bgld und/oder NÖ
10-11 Jahre	HS oder neue MS	5	30	Wien und Bgld oder NÖ
10-11 Jahre	Gymnasium	5	30	Wien und/oder NÖ
12-14 Jahre	HS (oder neue MS)	6, 7, 8	45	Wien und Bgld oder NÖ
12-14 Jahre	Gymnasium	6, 7, 8	45	Wien und/oder NÖ
15 Jahre	Polytechnikum	9	15	Wien und Bgld oder NÖ
15 Jahre	Gymnasium	9	15	Wien oder NÖ
INSGESAMT			280	

ANTWORTSCHREIBEN DES STADTSCHULRATS



Frau

Dr. Stefana Holocher-Ertl

Fakultät f. Psychologie
Liebiggasse 5
1010 Wien

Ihr Zeichen, Unser Zeichen/GZ
Ihre Nachricht vom 000.0029/60/2009

Bearbeiter/In
OR Mag. Klaus Rick
office@ssrwien.gv.at

Tel. 525 25
DW: 77033
Fax DW/9977033

Datum
18.11.2009

Bewilligung einer wissenschaftlichen Erhebung

Sehr geehrte Frau Dr. Holocher-Ertl!

Auf Grund Ihres Ansuchens erteilt Ihnen der Stadtschulrat für Wien gemäß § 46 Absatz 2 Schulunterrichtsgesetz die Genehmigung eine wissenschaftliche Erhebung zur Thematik „Intelligenztestbatterie AID 3“ in den angeführten Schulen durchzuführen.

Es wird darauf hingewiesen, dass das Einvernehmen mit den zuständigen Direktionen herzustellen ist und hinsichtlich der Mitarbeit der LehrerInnen der unbedingte Nachweis der Freiwilligkeit vorliegen muss.

Es ist daher besonders darauf zu achten, dass die Fragebögen sowohl den einzelnen Personen als auch den einzelnen Schulen bei der Auswertung **nicht** zuordenbar sind (keine Abfrage von Namen, Schule und Adresse). Um das Prinzip der Anonymität zu wahren, ist daher auf jegliche namentliche Zuordnung auf dem Fragebogen zu verzichten.

Die vorgelegten Unterlagen sind mit der Einschränkung, dass die Befragung der SchülerInnen höchstens eine Unterrichtsstunde in Anspruch nimmt, verbindlich. Diese Bedingung ist in Hinblick auf die Vielzahl der Erhebungen und für die lehrplanmäßige Abhaltung des Unterrichts unerlässlich.

Alle Erhebungsergebnisse unterliegen der strikten Anonymität und dürfen nur für die Studie Verwendung finden.

Der Ablauf der Befragung (Verteilung und Einsammlung der Zustimmungserklärung der Erziehungsberechtigten, inhaltliche und formale Einführung, Verteilung und Einsammlung der Fragebögen) ist von der Antragstellerin (oder ihren MitarbeiterInnen) durchzuführen.

Alle Erhebungsergebnisse unterliegen der strikten Anonymität und dürfen nur für die Studie Verwendung finden.

Der Stadtschulrat für Wien ersucht um Zusendung eines Exemplars Ihrer Studie nach Fertigstellung zu obiger Geschäftszahl.

Mit freundlichen Grüßen

Für die Amtsführende Präsidentin:

OR Mag. Klaus Rick eh.

LEHRERINNENBRIEF

Univ.-Ass. Dr. Stefana Holocher-Ertl
Klinische- und Gesundheitspsychologin
Fakultät f. Psychologie der Universität Wien
Inst. f. Entwicklungspsychologie und Psychologische Diagnostik
Liebiggasse 5
A-1010 Wien
email: stefana.holocher-ertl@univie.ac.at
Tel: 01 – 4277 47851



universität
wien



Test- und
Beratungsstelle

Wien, Oktober 2009

Sehr geehrte Lehrerinnen und Lehrer!

Viele Kinder werden im Laufe ihrer schulischen Karriere aus den unterschiedlichsten Gründen mit psychologischen Tests untersucht. Bei schulpyschologischen Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Verhaltensprobleme im schulischen Kontext, Aufmerksamkeitsprobleme, Abklärung einer möglichen Legasthenie/Dyskalkulie, Schullaufbahnberatungen etc. kommt dabei meist ein Intelligenztest zum Einsatz, um die intellektuellen Stärken und Schwächen des Kindes abschätzen zu können.

Das *Adaptive Intelligenz Diagnostikum AID* – aktuelle Version 2.2 (AID 2.2, Kubinger, 2009) – ist ein im deutschen Sprachraum sehr etabliertes Verfahren zur Erfassung der intellektuellen Fähigkeiten von Kindern und Jugendlichen zwischen 6 und 15 Jahren. Diese Intelligenz-Testbatterie wurde nun um neue Wissensgebiete und historische und geografische sowie sprachliche Entwicklungen aktualisiert.

Im Zuge eines Forschungsprojekts der Universität Wien (Leitung: Univ. Prof. Dr. Mag. Klaus D. Kubinger) findet nun eine Schüler(innen)-Erhebung mit der aktualisierten Form AID 3 statt.

Wir wenden uns daher mit der Bitte an die Eltern Ihrer Schüler(innen), diese an der Untersuchung teilnehmen zu lassen, vorausgesetzt natürlich, dass das Kind damit einverstanden ist. Die Untersuchungen finden während der Schulzeit einzeln statt und dauern ca. zwei Schulstunden. Durchgeführt werden die Testungen von speziell dafür ausgebildeten Testleitern(innen). Erfahrungsgemäß macht den Kindern die Mitarbeit an den Aufgaben viel Spaß. Möchte das Kind jedoch einmal eine Pause einlegen oder die Untersuchung aus irgendeinem Grund frühzeitig abbrechen, ist das natürlich jederzeit möglich.

Auf Wunsch werden wir den Eltern als kleines Dankeschön einen kurzen schriftlichen Ergebnisbericht über die intellektuellen Stärken und Schwächen des Kindes zuschicken.

Die gewonnenen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke genutzt. Sämtliche Ergebnisse der Schüler(innen) werden (noch während der Testung) bzw. nach Erstellung des schriftlichen Ergebnisberichts von den Testleitern(innen) anonymisiert. Es können keine Ergebnisse an Sie oder die Schuldirektion weitergegeben werden.

Wir möchten Sie bitten, die Elternbriefe an Ihre Schüler(innen) zu übergeben und die Rückmeldungen der Eltern gesammelt in der Direktion Ihrer Schule abzugeben. Wenn sich Eltern Ihrer Schüler(innen) mit der Teilnahme an der Untersuchung einverstanden erklären, werden wir bei der Terminvereinbarung großen Wert darauf legen, den Ablauf Ihres Unterrichts so wenig wie möglich zu unterbrechen.

Für eventuelle Rückfragen stehen als Ansprechpersonen Frau Dr. Stefana Holocher-Ertl, Projektkoordinatorin (Tel: +43-1-4277 47851, email: stefana.holocher-ertl@univie.ac.at), und Frau Nicole Gömer, Projektassistentin (email: nicole.goerner@gmx.at), jederzeit gerne zur Verfügung.

Mit der Bitte um Ihre Unterstützung, freundlichen Grüßen und herzlichen Dank im Voraus!

Dr. Stefana Holocher-Ertl (im Auftrag der Projektleitung)

ELTERNBRIEF

Univ.-Ass. Dr. Stefana Holocher-Ertl
Klinische- und Gesundheitspsychologin
Fakultät f. Psychologie der Universität Wien
Inst. f. Entwicklungspsychologie und Psychologische Diagnostik
Liebiggasse 5
A-1010 Wien
email: stefana.holocher-ertl@univie.ac.at
Tel: 01 – 4277 47851



universität
wien



Test- und
Beratungsstelle

Wien, Oktober 2009

Liebe Eltern!

Viele Kinder werden im Laufe ihrer schulischen Karriere aus den unterschiedlichsten Gründen mit psychologischen Tests untersucht. Bei schulpyschologischen Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Verhaltensprobleme im schulischen Kontext, Aufmerksamkeitsprobleme, Abklärung einer möglichen Legasthenie/Dyskalkulie, Schullaufbahnberatungen etc. kommt dabei meist ein Intelligenztest zum Einsatz, um die intellektuellen Stärken und Schwächen des Kindes abschätzen zu können.

Das *Adaptive Intelligenz Diagnostikum AID* – aktuelle Version 2.2 (AID 2.2, Kubinger, 2009) – ist ein im deutschen Sprachraum sehr etabliertes Verfahren zur Erfassung der intellektuellen Fähigkeiten von Kindern und Jugendlichen zwischen 6 und 15 Jahren. Diese Intelligenz-Testbatterie wurde nun um neue Wissensgebiete und historische und geografische sowie sprachliche Entwicklungen aktualisiert.

Im Zuge eines Forschungsprojekts der Universität Wien (Leitung: Univ. Prof. Dr. Mag. Klaus D. Kubinger) findet nun eine Schüler(innen)-Erhebung mit der aktualisierten Form AID 3 statt.

Wir würden es sehr begrüßen, in dieses Forschungsprojekt auch Ihr Kind einbeziehen zu können.

Wir wenden uns daher mit der Bitte an Sie, Ihr Kind/Ihre Kinder an dieser Untersuchung teilnehmen zu lassen, vorausgesetzt natürlich, Ihr Kind ist einverstanden. Die Untersuchungen finden während der Schulzeit einzeln statt und dauern ca. zwei Schulstunden. Durchgeführt werden die Testungen von speziell dafür ausgebildeten Testleitern(innen). Erfahrungsgemäß macht den Kindern die Mitarbeit an den Aufgaben viel Spaß. Möchte Ihr Kind jedoch einmal eine Pause einlegen oder die Untersuchung aus irgendeinem Grund frühzeitig abbrechen, ist das natürlich jederzeit möglich.

Auf Wunsch ist es auch möglich, Ihnen als kleines Dankeschön einen kurzen schriftlichen Ergebnisbericht über die intellektuellen Stärken und Schwächen Ihres Kindes/Ihrer Kinder zuzuschicken.

Für eventuelle Rückfragen stehen als Ansprechpersonen Frau Dr. Stefana Holocher-Ertl, Projektkoordinatorin (Tel: +43-1-4277 47851, email: stefana.holocher-ertl@univie.ac.at), und Frau Nicole Görner, Projektassistentin (email: nicole.goerner@gmx.at), jederzeit gerne zur Verfügung.

Die gewonnenen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke genutzt. Sämtliche Ergebnisse der Schüler(innen) werden (noch während der Testung) bzw. nach Erstellung des schriftlichen Ergebnisberichts von den Testleitern(innen) anonymisiert. Direktion bzw. Lehrer(innen) der Schule werde selbstverständlich nicht über die Ergebnisse informiert.

Wir bitten Sie, mit Ihrer Unterschrift auf dem beiliegenden Formular, Ihr Einverständnis zur Teilnahme Ihres Kindes an der oben beschriebenen Untersuchung zu erteilen.

Mit freundlichen Grüßen und herzlichen Dank im Voraus!

Dr. Stefana Holocher-Ertl (im Auftrag der Projektleitung)

EINVERSTÄNDNISERKLÄRUNG DER ELTERN

Ich erkläre mich mit der Teilnahme meiner Tochter/meines Sohnes

_____, geboren am _____, an der
Name des Kindes

Schüler(innen)-Erhebung zum AID 3 einverstanden.

☐ Ich bitte um die Zusendung eines kurzen schriftlichen Ergebnisberichts an die Adresse:

☐ Ich wünsche keinen Ergebnisbericht.

Datum

Unterschrift des/der Erziehungsberechtigten

PROTOKOLLBOGEN 1

AID 3 Protokollbogen

Vorname: _____ Nachname: _____

Geburtsdatum: _____ Geschlecht: männlich ☐ weiblich ☐

Schriftlicher Befund gewünscht: ja ☐ nein ☐

wenn ja: Wohnadresse: _____

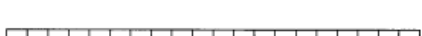
Schule: _____ Klasse: _____

Muttersprache: _____

Testdatum: _____ Testleiter(in): _____

Testzusammenstellung: 1 ☐ 2 ☐ 3 ☐ 4 ☐

PROTOKOLLBOGEN 2

Untertest	Block	Aufgaben	RW	Block	Aufgaben	RW	Block	Aufgaben	RW	Block	Aufgaben	RW	PS	FP	T-Wert
A															
B	Lern tafel: 	Lernzeit 1:  min  sek	Prüfdurchgang 1: 	Lern tafel: 	Lernzeit 2:  min  sek	Prüfdurchgang 2: 	Lern tafel: 	Lernzeit 3:  min  sek	Prüfdurchgang 3: 	Lern tafel: 	Lernzeit 4:  min  sek	Prüfdurchgang 4: 			
	Prüfdurchgang nach 20 min: 														
C	Testform: 														
D															
E															
F															
G	Testform: 														
H															
I															
Test-zusammenstellung 1															
Test-zusammenstellung 2															
Test-zusammenstellung 3															
Test-zusammenstellung 4	„vorwärts“			Versuche											
	„rückwärts“			Versuche											
	Kodiermenge 1 Minute		Differenz:		Kodiermenge 2 Minuten										
	Assoziationen														

DANKESBRIEF AN DIE ELTERN

Univ.-Ass. Dr. Stefana Holocher-Ertl
Klinische- und Gesundheitspsychologin
Fakultät f. Psychologie der Universität Wien
Inst. f. Entwicklungspsychologie und Psychologische Diagnostik
Liebiggasse 5
A-1010 Wien
email: stefana.holocher-ertl@univie.ac.at
Tel: 01 – 4277 47851



universität
wien



Test- und
Beratungsstelle

Wien, Jänner bis März 2010

Liebe Eltern!

Vielen herzlichen Dank, dass Ihr Kind an der Schüler(innen)-Testung mit dem Intelligenztest AID 3 teilnehmen konnte!

Den von Ihnen gewünschten Ergebnisbericht über die intellektuellen Stärken und relativen Schwächen Ihres Kindes werden wir Ihnen Ende Juni 2010 an die von Ihnen angegebene Adresse zuschicken. Sollten sich bis dahin Änderungen der Adresse ergeben, bitten wir Sie, uns rechtzeitig darüber zu informieren. Als Ansprechperson steht Ihnen dafür Frau Nicole Görner, Projektassistentin (email: nicole.goerner@gmx.at), jederzeit gerne zur Verfügung. Nochmals möchten wir darauf hinweisen, dass nach Verschicken des Ergebnisberichtes die Daten anonymisiert werden.

Mit freundlichen Grüßen und herzlichem Dank!

Dr. Stefana Holocher-Ertl (im Auftrag der Projektleitung)

SCHRIFTLICHER ERGEBNISBERICHT

Univ.-Ass. Dr. Stefana Holocher-Ertl
Klinische- und Gesundheitspsychologin
Fakultät f. Psychologie der Universität Wien
Inst. f. Entwicklungspsychologie und Psychologische
Diagnostik
Liebiggasse 5
A-1010 Wien
email: stefana.holocher-ertl@univie.ac.at
Tel: 01 – 4277 47851



universität
wien



Test- und
Beratungsstelle

Wien, im Juni 2010

Liebe Eltern!

Vielen Dank für Ihr Einverständnis zur Teilnahme Ihres Kindes an der Schüler(innen)erhebung zur Intelligenz-Testbatterie AID 3 im Rahmen eines Forschungsprojekts der Universität Wien (Leitung: Univ. Prof. Dr. Mag. Klaus D. Kubinger). Das Adaptive Intelligenz Diagnostikum AID – aktuelle Version 2.2 (AID 2.2, Kubinger, 2009) - ist ein im deutschen Sprachraum sehr etabliertes Verfahren zur Erfassung der intellektuellen Fähigkeiten von Kindern und Jugendlichen zwischen 6 und 15 Jahren. Diese Intelligenz-Testbatterie wurde in der Version AID 3 um neue Wissensgebiete erweitert und aktualisiert und nun erstmals an Schüler/innen in Wien und Niederösterreich erprobt.

Wir wollen Ihnen nun über die Testergebnisse Ihres Kindes berichten. Für eine anonymisierte Verarbeitung und Speicherung der Daten haben wir Ihrem Kind folgenden Probandencode zugeteilt:

bh1

Die Testung Ihres Kindes mit dem AID 3 fand innerhalb der Schulzeit statt und umfasste die Dauer von ungefähr einer Stunde. Durchgeführt wurde diese von einer/einem speziell dafür ausgebildeten Testleiter/in.

Testergebnisse:

Die Leistungen in den einzelnen Untertests wurden jeweils mit einer altersspezifischen Stichprobe aus Wien und Niederösterreich verglichen. Die Testergebnisse werden in Prozenträngen (PR) angegeben, wobei ein Prozentrangwert (PR) von 25 bis 75 als durchschnittlich (alterstypisch) gilt. Der PR gibt an, wie viel Prozent der Gleichaltrigen in der Vergleichsstichprobe eine gleich gute oder niedrigere Leistung erbringen.

Untertest	Interpretation
Alltagswissen PR = 76	Es wird die Fähigkeit gemessen, sich Sachkenntnisse über Inhalte anzueignen, die in der heutigen Gesellschaft alltäglich sind (Wissen zu den Themen: Geschichte, Erdkunde, Sport, Kunst, Biologie). Die Leistungen des/der Schülers/in liegen hier über dem Altersdurchschnitt.
Antonyme PR = 66	Gemessen wird die Fähigkeit, die Gegensätzlichkeit von Begriffen zu erkennen und die Größe des Wortschatzes, der solche Gegensätze auszudrücken vermag. Der/die Schüler/in erbrachte hier eine durchschnittlich gute Leistung.
Realitätssicherheit PR = 79	Es wird die Fähigkeit gemessen, wesentliche Merkmale von Dingen des Alltags zu erkennen, wenn diese auf Bildern fehlen. Weiters zeigt es auch die Ausprägung der visuellen Differenzierungsfähigkeit. Der/die Schüler/in erreicht hier ein überdurchschnittliches Ergebnis.
Angewandtes Rechnen PR = 82	Dieser Untertest zeigt die rechnerische Fähigkeit, unabhängig von schulischen Rechenfertigkeiten. Es zeigt die Fähigkeit, Problemstellungen des Alltags durch Anwendung passender Rechenoperationen lösen zu können. Die Leistung des/der Schülers/in liegen hier über dem Durchschnitt.
Synonyme finden PR = 70	Es wird die Fähigkeit gemessen, die Bedeutung sprachgebundener Begriffe zu erkennen, und die Größe des Wortschatzes, der solche Begriffe durch andere Worte auszudrücken vermag. Der/die Schüler/in erbrachte hier eine durchschnittliche Leistung.
Logische Blöcke PR = 40	Es zeigt sich die Fähigkeit, des Erkennens und Zuordnens der Zugehörigkeit von Figuren zu einer vorgegebenen Figurenreihe. Es handelt sich hierbei um die Fähigkeit zum logisch-schlussfolgernden Denken bei visuellem Aufgabenmaterial. Die Leistungen des/der Schülers/in liegen dabei im durchschnittlichen Bereich.

Funktionen abstrahieren PR = 45	Geprüft wird die Fähigkeit, durch schlussfolgerndes Denken im sprachlichen Bereich Funktionen zu abstrahieren und diese sprachlich ausdrücken zu können. Der/die Schüler/in erreichte ein durchschnittliches Ergebnis.
Soziales Erfassen und sachliches Reflektieren PR = 76	Gemessen wird das Verständnis, über Sachzusammenhänge der gesellschaftlichen Umwelt und über soziale angepasste Verhaltensweisen und gesellschaftliche Bedingungen bescheid zu wissen. Die Leistung des/der Schülers/in liegt dabei über dem Durchschnitt.

Ergänzende Bemerkungen: falls es die gibt ... Der/die Schüler/in arbeitete konzentriert und sehr bemüht. Er/sie zeigte allerdings Schwächen im Sprachverständnis....

Die Leistungen Ihres Kindes wurden im Rahmen eines Forschungsprojektes gewonnen und sind daher nur eingeschränkt aussagekräftig. Sollten Sie genauere Informationen zu der Leistungsfähigkeit Ihres Kindes wünschen, so raten wir Ihnen zu einer nochmaligen Testung mit ausführlicher Beratung bei einem/einer niedergelassenen Kinderpsychologen/in (www.psychologie.at).

Nochmals vielen Dank für Ihre Teilnahme und alles Gute für Sie und für die Zukunft Ihres Kindes,

Dr. Stefana Holocher-Ertl

P-WERT-TABELLEN

Tabelle 1: p-Werte des UT „Angewandtes Rechnen“ für alle Teilungskriterien, erster Berechnungsdurchgang

ItemNr	p.Median	p.Geschlecht	p.Sprache	p.Alter
it1				
it10	0,458	0,670	0,552	
it11	0,068	0,261	0,207	0,714
it12	0,002	0,900	0,013	0,967
it13	0,465	0,383	0,889	0,495
it14	0,127	0,520	0,754	0,976
it15	0,045	0,877	0,835	0,837
it16	0,014	0,017	0,766	0,628
it17	0,147	0,728	0,436	0,033
it18	0,004	0,931	0,155	0,022
it19	0,358	0,762	0,571	0,054
it2				
it20	0,009	0,455	0,124	0,016
it21	0,740	0,026	0,058	0,000
it22	0,539	0,844	0,216	0,292
it23	0,013	0,005	0,156	0,499
it25	0,054	0,287	0,272	0,040
it27	0,002	0,877	0,405	0,722
it28	0,218	0,168	0,350	0,134
it29	0,757	0,078	0,822	0,115
it3				
it30	0,570	0,358	0,801	0,731
it31	0,996	0,654	0,698	
it32	0,440	0,281	0,116	
it33	0,017	0,754	0,151	
it34	0,047	0,259	0,937	
it35	0,234	0,738	0,229	
it36	0,738	0,008	0,919	
it37	0,418	0,429		
it38	0,187	0,624		
it4				
it41	0,841	0,289	0,503	

it42	0,012	0,993	0,995	0,766
it43			0,541	
it44	0,026	0,043	0,683	
it45	0,012	0,441	0,840	
it46	0,359	0,044	0,041	
it47	0,977	0,116	0,888	
it49	0,071	0,800	0,657	0,589
it5				
it51	0,083	0,763	0,634	0,465
it52	0,645	0,049	0,155	0,173
it53	0,126	0,026	0,723	0,473
it54	0,013	0,694	0,002	0,761
it55	0,758	0,659	0,834	0,955
it56	0,019	0,721	0,004	
it57	0,228	0,900	0,231	
it58	0,477	0,389	0,388	
it59	0,675	0,817	0,006	
it6				
it60	0,561	0,568	0,031	
it61	0,031	0,029	0,085	0,513
it62	0,807	0,246	0,244	0,797
it63	0,110	0,285	0,021	0,621
it64				
it64z	0,511	0,649	0,523	0,539
it65	0,815	0,499	0,005	0,003
it66	0,151	0,508	0,124	0,650
it66z	0,009	0,610	0,085	0,011
it67	0,228	0,751	0,154	
it67z	0,916	0,157	0,076	0,092
it68	0,020	0,794	0,211	0,789
it68z	0,000	0,004	0,519	
it69	0,714	0,415	0,013	0,661
it69z	0,009	0,854	0,820	
it7		0,615		
it70z	0,620	0,648	0,271	
it72z	0,018	0,385	0,047	
it8	0,739	0,379	0,230	

it9	0,062	0,150	0,549	
-----	-------	-------	-------	--

Tabelle 2: p-Werte des UT „Angewandtes Rechnen“ für alle Teilungskriterien, letzter Berechnungsdurchgang

ItemNr	p.Median	p.Geschlecht	p.Sprache	p.Alter
it1				
it10	0,568	0,467	0,705	
it11				
it13	0,926	0,561	0,890	0,523
it14	0,029	0,522	0,691	0,944
it15		0,488	0,905	0,682
it17	0,055	0,557	0,352	0,094
it18	0,176	0,859	0,183	0,054
it19	0,526	0,863	0,515	0,073
it2				
it20	0,099	0,350	0,171	0,060
it22	0,681	0,724	0,245	0,320
it23	0,659	0,007	0,184	0,452
it25	0,067	0,330	0,239	0,058
it27	0,021	0,888	0,460	0,690
it28	0,252	0,135	0,332	0,119
it29	0,112	0,068	0,728	0,058
it3				
it30	0,714	0,303	0,761	0,802
it31	0,487	0,694	0,735	
it32	0,547	0,391	0,120	
it33	0,007	0,656	0,174	
it34	0,030	0,149	0,940	
it35	0,336	0,491	0,237	
it36	0,787	0,014	0,900	
it38	0,589	0,766		
it4				
it41				
it43			0,579	
it45	0,306	0,552	0,789	
it46	0,915	0,031	0,047	
it47	0,928	0,102	0,916	

it49	0,002	0,544	0,828	0,967
it5				
it51	0,055	0,999	0,817	0,848
it52	0,802	0,022	0,111	0,085
it53	0,850	0,045	0,560	0,285
it54	0,109	0,516	0,001	0,952
it55	0,501	0,557	0,735	0,708
it57	0,284	0,957	0,206	
it58	0,964	0,375	0,370	
it59	0,250	0,876	0,006	
it6				
it60	0,955	0,602	0,030	
it61	0,024	0,057	0,090	0,543
it62	0,081	0,299	0,218	0,945
it63	0,612	0,427	0,027	0,706
it64				
it64z	0,029	0,802	0,553	0,786
it66	0,280	0,587	0,143	0,754
it66z	0,011	0,469	0,067	0,004
it67		0,703	0,307	
it67z	0,955	0,146	0,087	0,042
it68	0,050	0,672	0,264	0,898
it69	0,881	0,502	0,014	0,418
it69z	0,003	0,712	0,870	
it7				
it70z	0,169	0,900	0,247	
it9	0,712	0,091	0,702	

Tabelle 3: p-Werte des UT „Funktionen Abstrahieren“ für alle Teilungskriterien, erster Berechnungsdurchgang

ItemNr	p.Median	p.Geschlecht	p.Sprache	p.Alter
it1				
it10	0,026	0,803	0,534	
it11	0,878	0,698	0,306	0,294
it12	0,165	0,488	0,410	0,967
it13	0,956	0,490	0,113	0,997

it14	0,001	0,234	0,002	0,459
it15	0,598	0,756	0,144	0,124
it16	0,711	0,156	0,493	0,443
it19	0,574	0,032	0,273	0,381
it2	0,056	0,331	0,041	
it20	0,014	0,012	0,805	0,913
it21	0,573	0,961	0,740	
it23	0,003	0,242	0,000	0,834
it24	0,056	0,181	0,094	0,580
it25	0,310	0,924	0,378	0,158
it26	0,000	0,185	0,304	
it27	0,035	0,275	0,694	
it28	0,378	0,393	0,667	0,894
it3	0,014	0,423	0,233	
it30	0,016	0,344	0,010	0,382
it32	0,162	0,405	0,750	
it34	0,004	0,617	0,220	
it36	0,851	0,139	0,002	
it37	0,196	0,967		
it38	0,544	0,904	0,576	
it39	0,707	0,886		
it4				
it40	0,122	0,744	0,449	
it41	0,938	0,131	0,046	0,428
it42	0,853	0,261	0,071	0,083
it43	0,431	0,775	0,428	
it44	0,527	0,143	0,305	
it46	0,387	0,021	0,342	
it47	0,893	0,260	0,831	
it48	0,909	0,270	0,131	0,502
it49	0,775	0,713	0,034	0,134
it5	0,120	0,260	0,760	
it51	0,837	0,968	0,235	0,239
it52	0,414	0,915	0,713	0,588
it53	0,512	0,659	0,830	0,117
it54	0,142	0,288	0,000	0,383
it55	0,182	0,490	0,814	0,919

it56	0,481	0,426	0,013	
it57	0,407	0,303	0,311	
it58	0,732	0,023	0,363	
it59	0,203	0,874	0,885	
it6	0,405	0,143	0,017	
it60	0,035	0,478	0,551	
it61	0,040	0,475	0,949	0,295
it61z	0,015	0,134	0,033	
it62	0,169	0,014	0,302	0,312
it63	0,642	0,631		0,726
it63z	0,070	0,010	0,850	
it64	0,045	0,008	0,013	0,010
it65	0,246	0,679	0,011	
it66	0,024	0,077	0,967	
it67	0,005	0,114	0,477	
it67z	0,046	0,395	0,042	0,133
it68	0,102	0,381	0,627	
it69	0,883	0,079		0,957
it69z	0,002	0,759	0,105	0,131
it7	0,146	0,204	0,122	
it70	0,182	0,943	0,256	
it70z	0,074	0,000	0,005	
it71	0,521	0,001	0,109	0,026
it72	0,025	0,973	0,177	
it72z	0,133	0,213	0,799	
it8	0,405	0,143	0,295	
it9	0,988	0,408	0,003	

Tabelle 4: p-Werte des UT „Funktionen Abstrahieren“ für alle Teilkriterien, letzter Berechnungsdurchgang

ItemNr	p.Median	p.Geschlecht	p.Sprache	p.Alter
it1				
it10	0,827	0,904	0,913	
it11	0,224	0,630	0,081	0,109
it12	0,288	0,399	0,969	0,259
it13	0,780	0,397	0,018	0,273

it15	0,867	0,916	0,334	0,302
it16	0,817	0,146	0,992	0,819
it19	0,653	0,025	0,099	0,249
it20	0,037	0,016	0,446	0,929
it21	0,145	0,851	0,885	
it24	0,605	0,288	0,116	0,599
it25	0,570	0,679	0,383	0,205
it27	0,515	0,143	0,770	
it28	0,804	0,760	0,742	0,918
it30	0,476	0,156	0,021	0,435
it32	0,043	0,216	0,845	
it34	0,097	0,979	0,350	
it37		0,933		
it38	0,417	0,673	0,493	
it39	0,512	0,992		
it4				
it40	0,044	0,928	0,646	
it41	0,529	0,137	0,085	0,172
it42	0,354	0,239	0,160	0,059
it43	0,179	0,958	0,189	
it44	0,812	0,107	0,140	
it46	0,638	0,017	0,601	
it47	0,288	0,195	0,761	
it48	0,332	0,229	0,252	0,426
it49	0,732	0,819	0,068	0,127
it5	0,152	0,229	0,550	
it51	0,288	0,770	0,215	0,019
it52	0,986	0,823	0,637	0,787
it53	0,732	0,438	0,878	0,007
it55	0,211	0,917	0,760	0,434
it56	0,141	0,205	0,015	
it57	0,180	0,169	0,317	
it58	0,818	0,008	0,333	
it59	0,679	0,534	0,878	
it60	0,098	0,805	0,621	
it61	0,269	0,425	0,895	0,310
it62	0,041	0,016	0,314	0,307

it63	0,387	0,493		0,624
it63z	0,471	0,005	0,425	
it64	0,505	0,003	0,021	0,015
it65	0,738	0,380	0,023	
it66	0,481	0,041	0,880	
it67	0,558	0,052	0,643	
it68	0,318	0,354	0,421	
it69	0,470	0,070		0,607
it69z	0,006	0,691	0,048	0,034
it7	0,367	0,265	0,041	
it70	0,385	0,810	0,129	
it8	0,780	0,279	0,153	

TABELLEN AUSGESCHLOSSENER ITEMS

Tabelle 5: Übersicht ausgeschlossener Items UT "Angewandtes Rechnen"

Berechnungsdurchgang	Ausgeschlossene Items	Auffällig in den Kriterien
0	68z	Median
1	56	Median
2	65	Sprache + Alter
3	72z	Median
4	12	Median
5	21	Alter
6	44	Median + Geschlecht
7	8	Geschlecht + Sprache
8	16	Median
9	37	Median
10	42	Median
11	kein Item mehr entfernt	-

Tabelle 6: Übersicht ausgeschlossener Items UT "Funktionen Abstrahieren"

Berechnungsdurchgang	Ausgeschlossene Items	Auffällig in den Kriterien
0	23	Median
1	14	Median
2	6	Sprache
3	70z	Geschlecht + Sprache
4	2	Median + Sprache
5	9	Sprache
6	71	Geschlecht
7	26	Median
8	67z	Median
9	72	Median
10	61z	Sprache
11	3	Median
12	54	Sprache
13	36	Sprache
14	72z	Median
15	kein Item mehr entfernt	-

Konfidenzintervall-Ellipsen erster Berechnungsdurchgang UT „Angewandtes Rechnen“

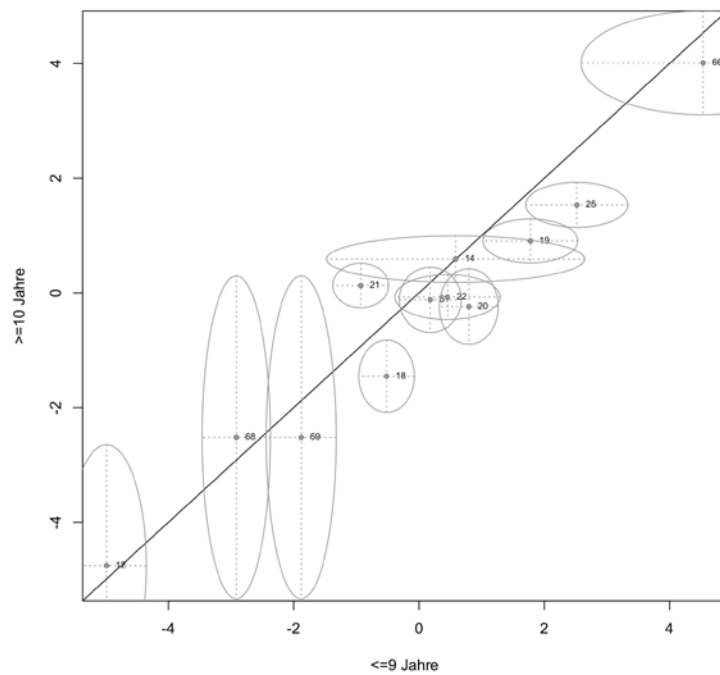


Abbildung 1: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 1

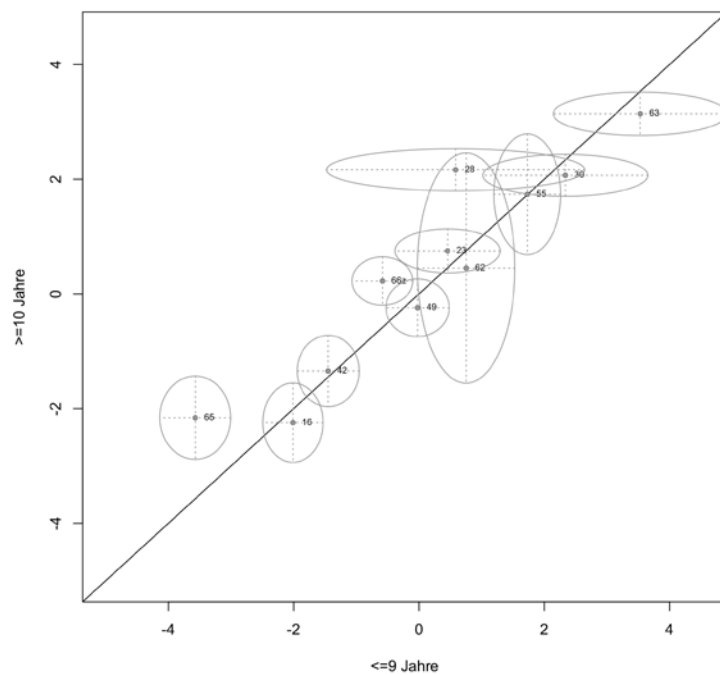


Abbildung 2: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 2

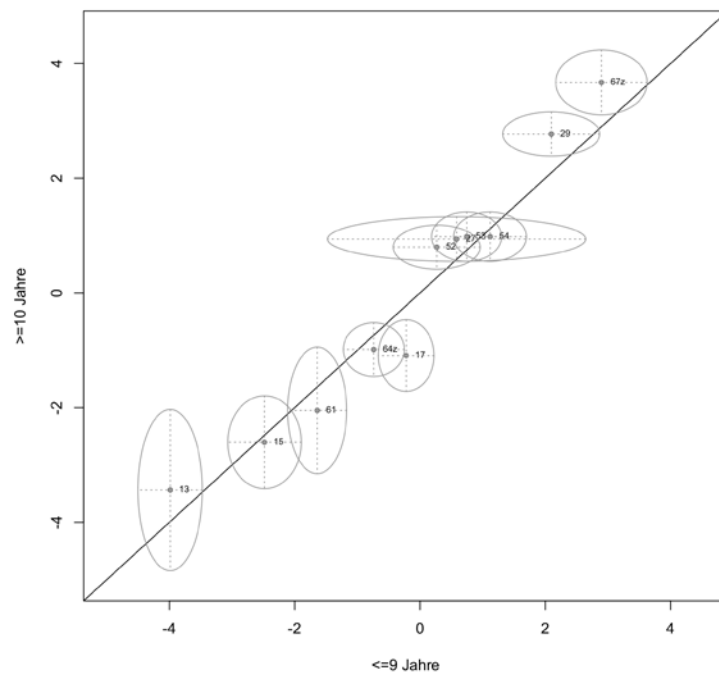


Abbildung 3: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 3

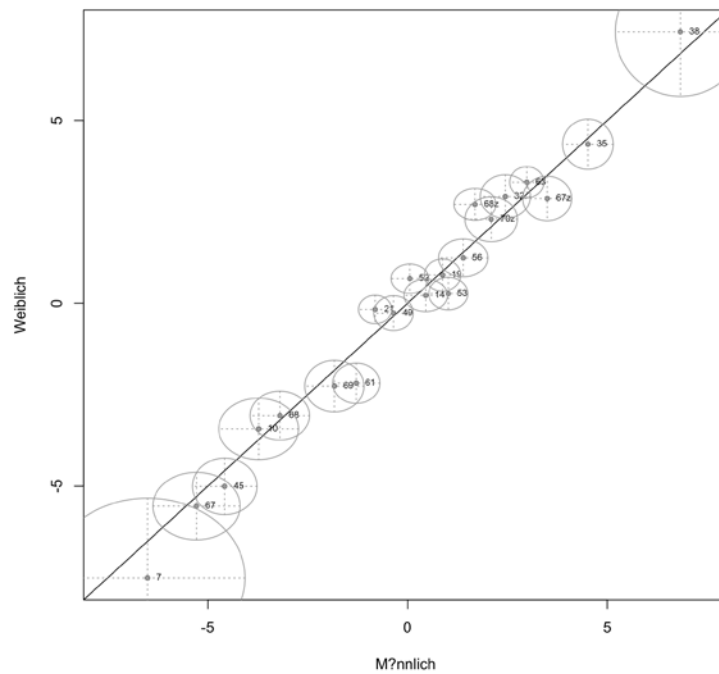


Abbildung 4: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 1

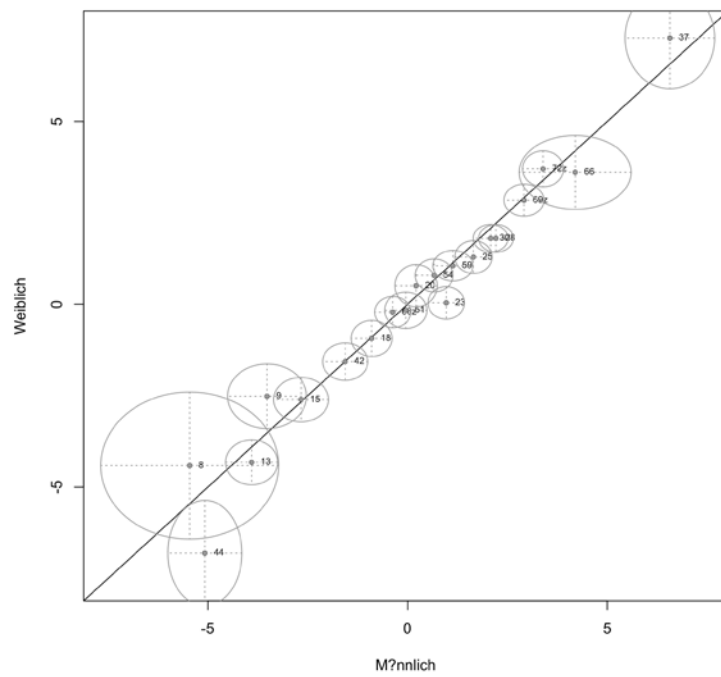


Abbildung 5: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 2

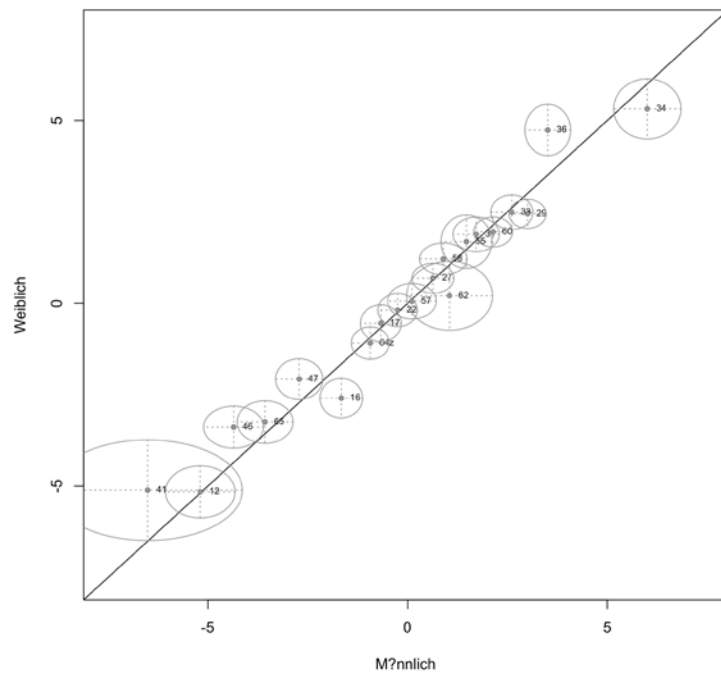


Abbildung 6: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 3

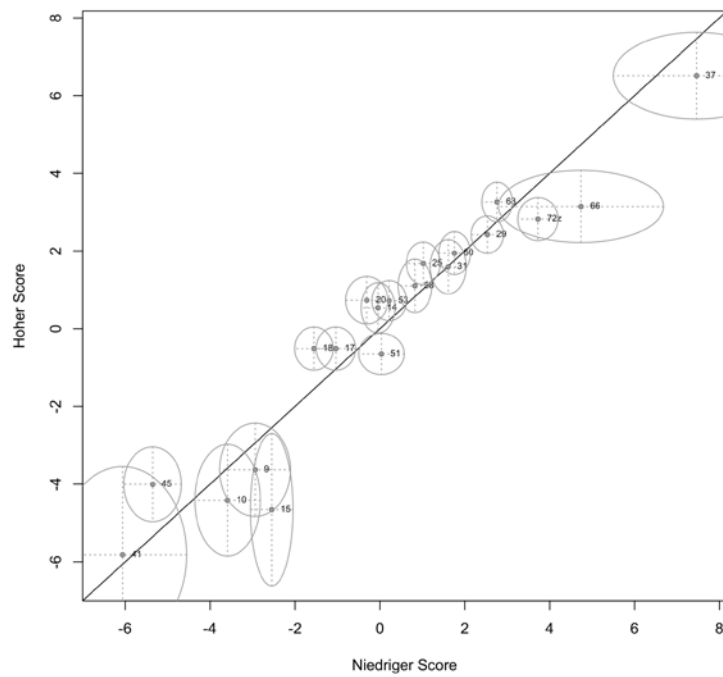


Abbildung 7: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 1

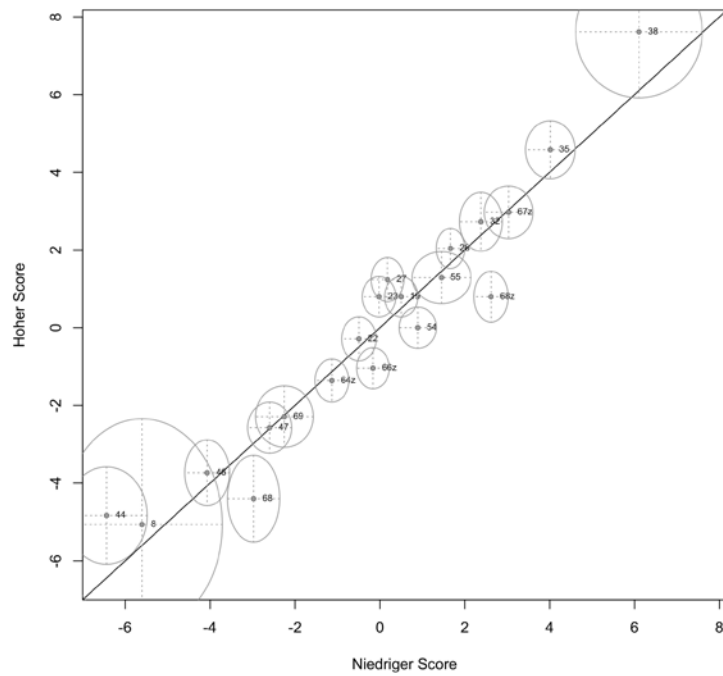


Abbildung 8: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 2

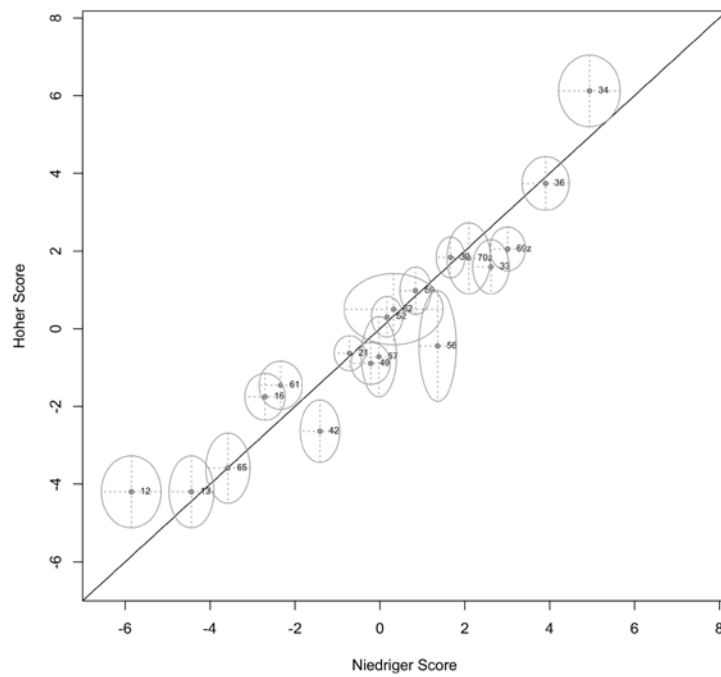


Abbildung 9: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 3

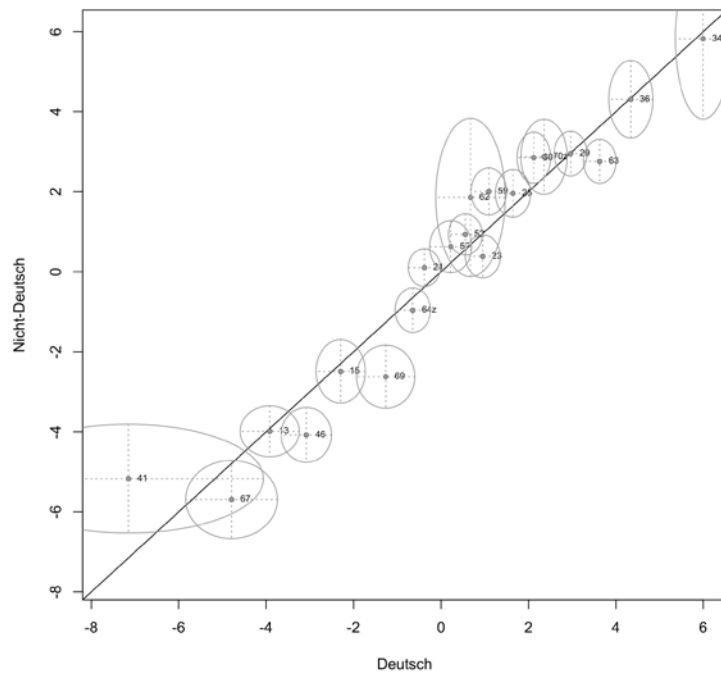


Abbildung 10: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Muttersprache mit Konfidenzintervall-Ellipsen 1

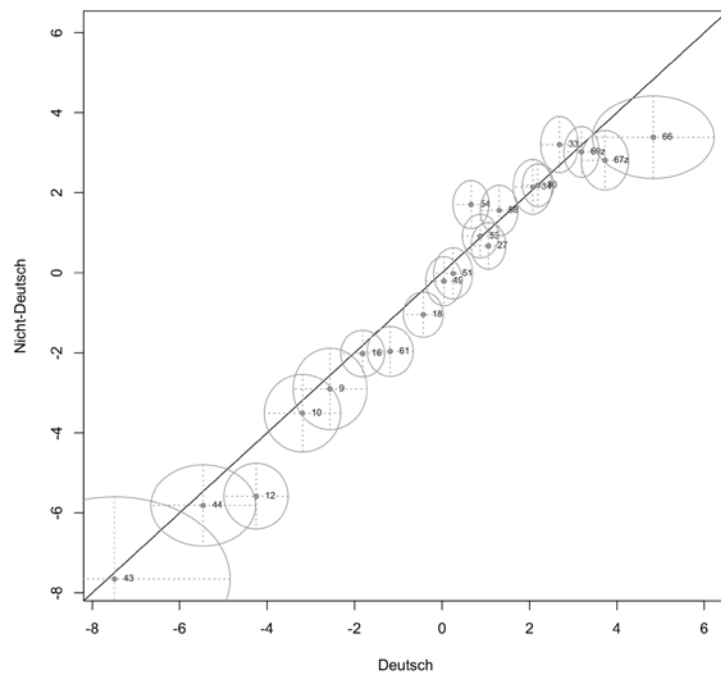


Abbildung 11: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Muttersprache mit Konfidenzintervall-Ellipsen 2

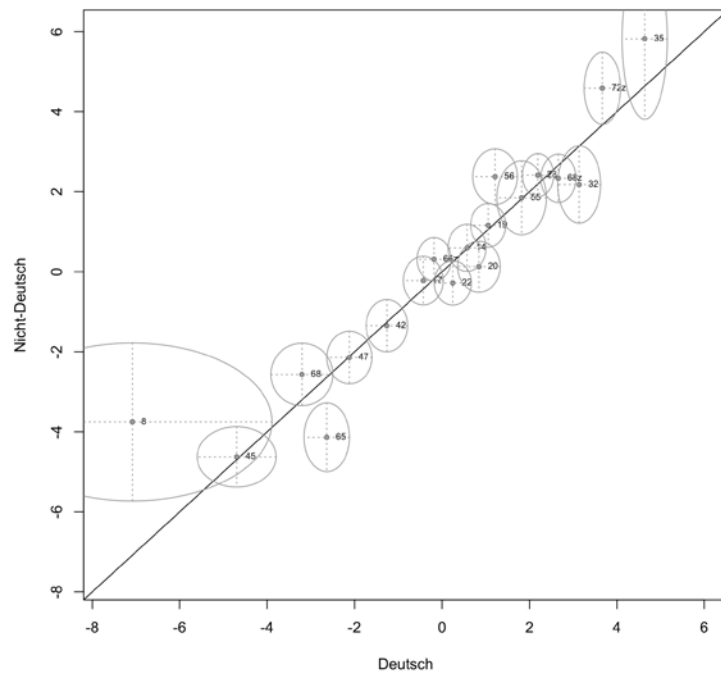


Abbildung 12: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Muttersprache mit Konfidenzintervall-Ellipsen 3

Konfidenzintervall-Ellipsen letzter Berechnungsdurchgang UT „Angewandtes Rechnen“

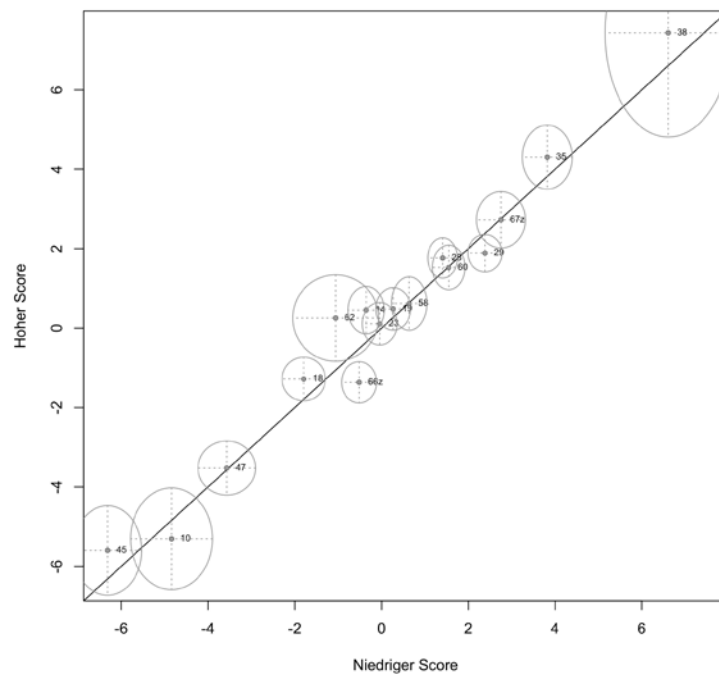


Abbildung 13: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 1 letzter Berechnungsdurchgang

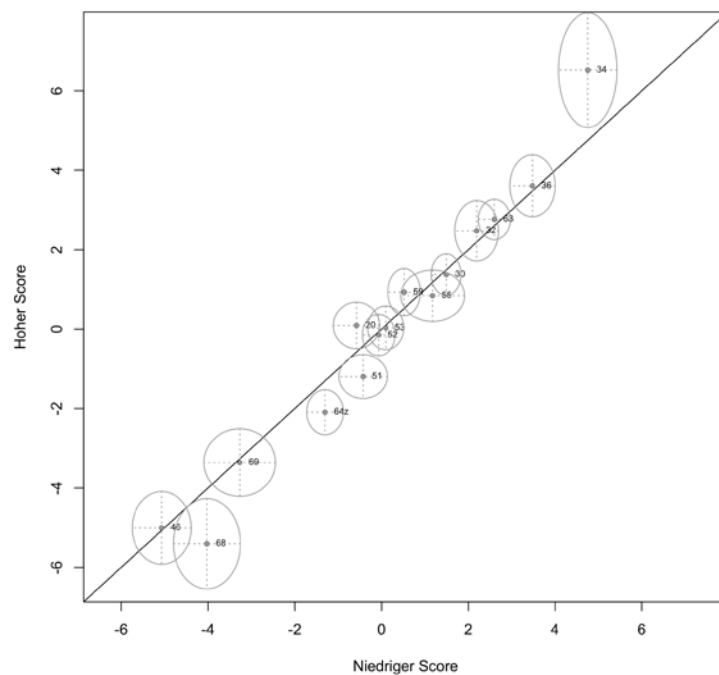


Abbildung 14: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 2 letzter Berechnungsdurchgang

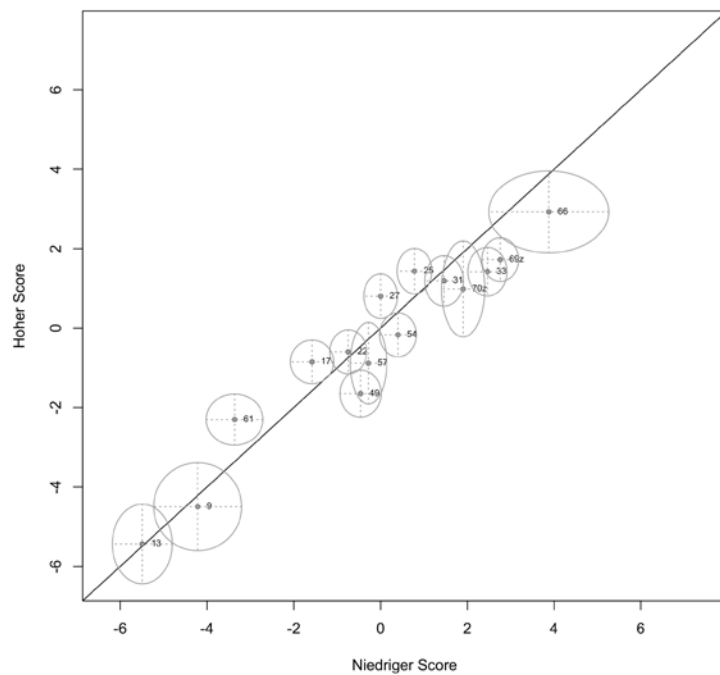


Abbildung 15: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 3 letzter Berechnungsdurchgang

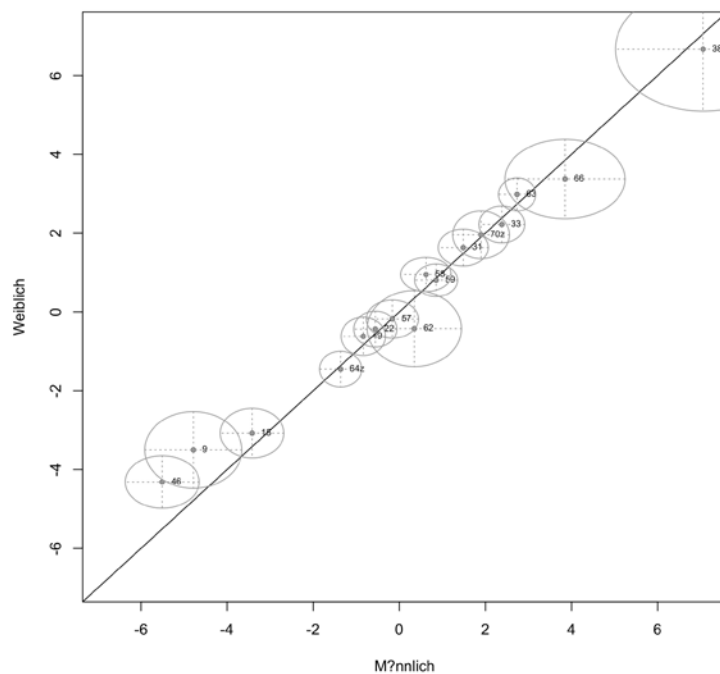


Abbildung 16: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 1 letzter Berechnungsdurchgang

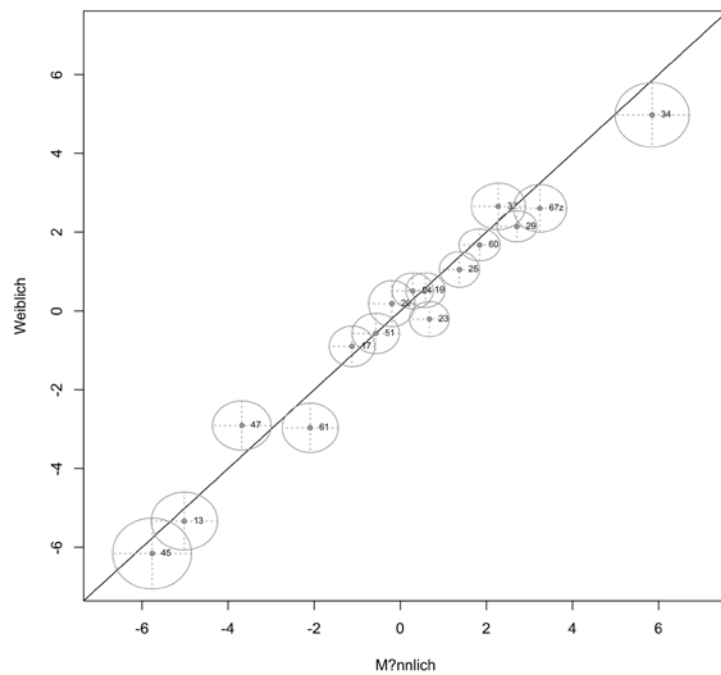


Abbildung 17: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilkriterium Geschlecht mit Konfidenzintervall-Ellipsen 2 letzter Berechnungsdurchgang

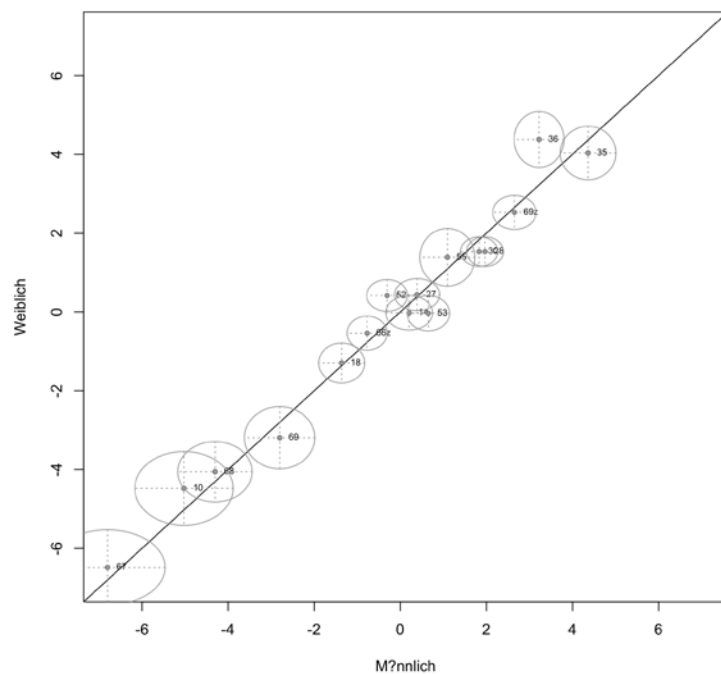


Abbildung 18: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilkriterium Geschlecht mit Konfidenzintervall-Ellipsen 3 letzter Berechnungsdurchgang

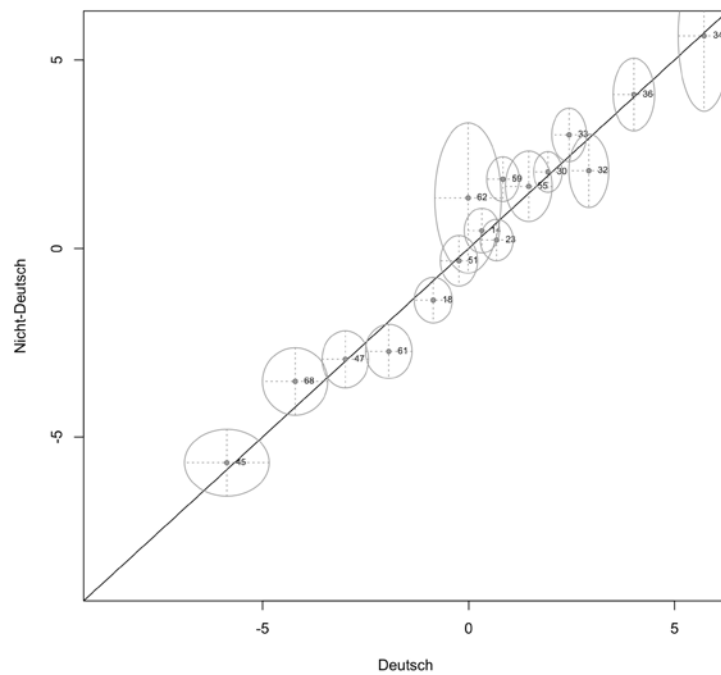


Abbildung 19: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 1 letzter Berechnungsdurchgang

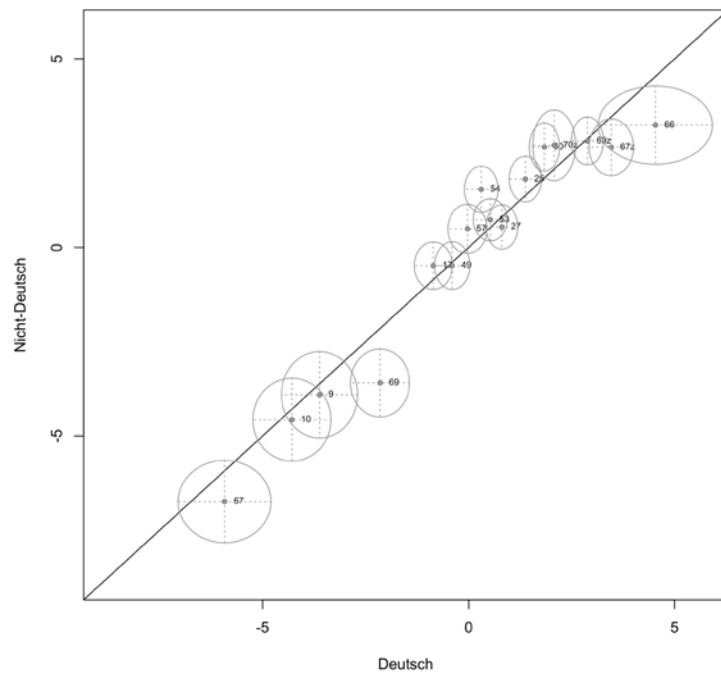


Abbildung 20: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 2 letzter Berechnungsdurchgang

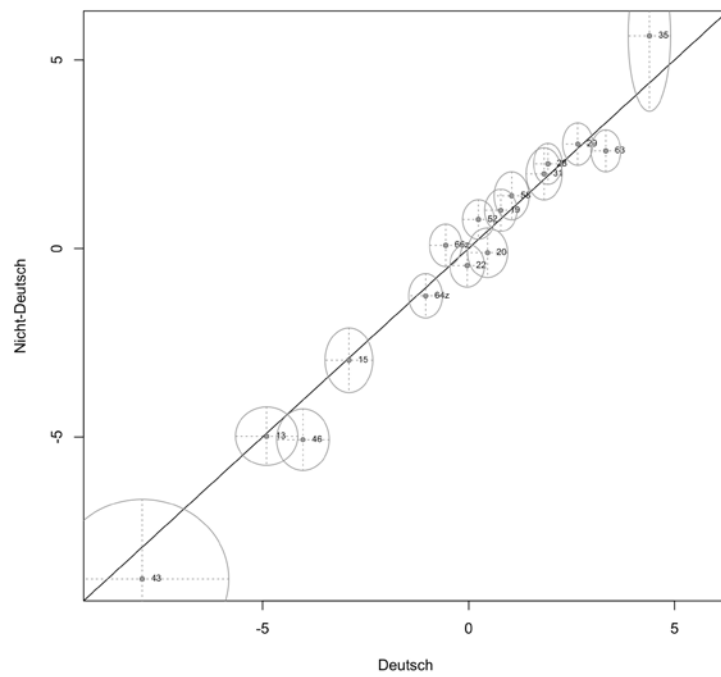


Abbildung 21: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 3 letzter Berechnungsdurchgang

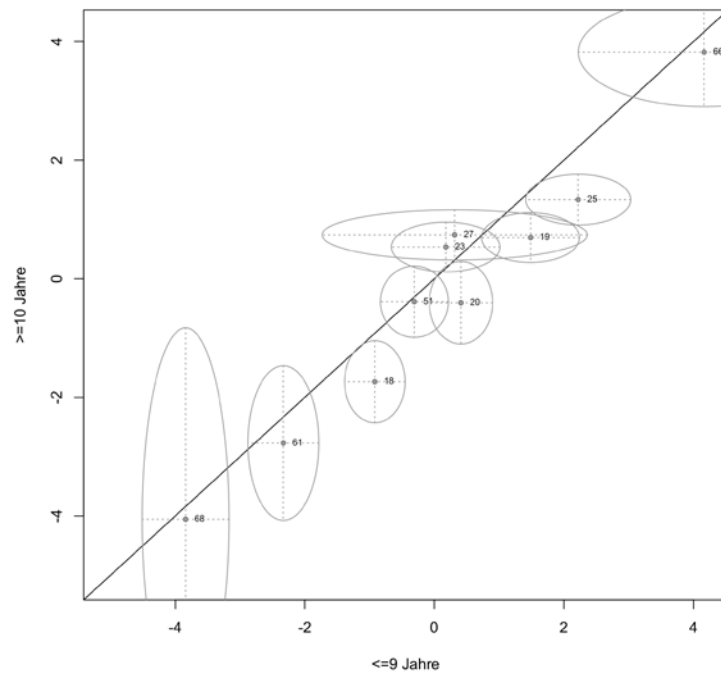


Abbildung 22: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 1 letzter Berechnungsdurchgang

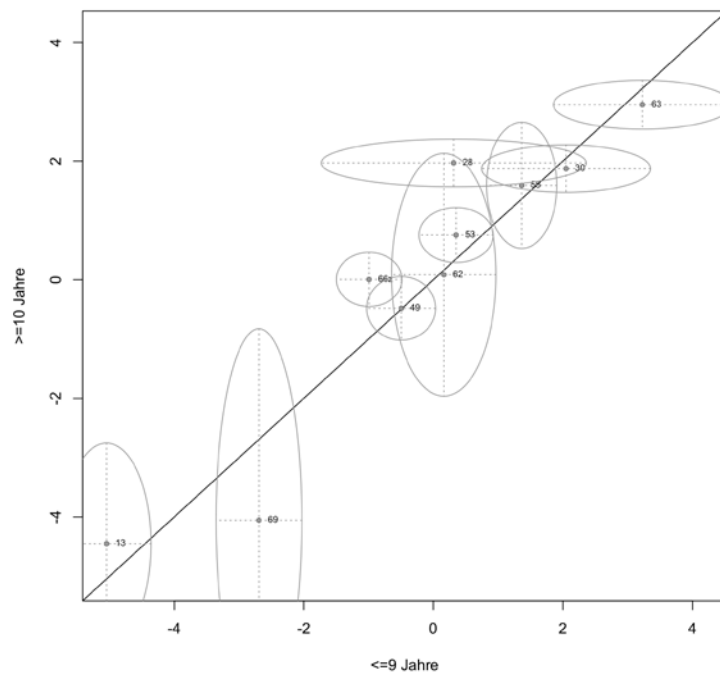


Abbildung 23: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 2 letzter Berechnungsdurchgang

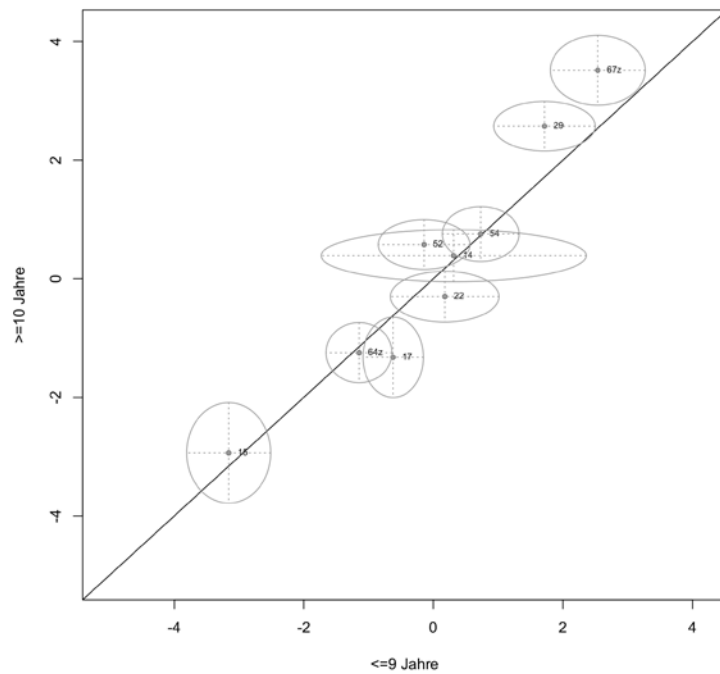


Abbildung 24: UT "Angewandtes Rechnen" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 3 letzter Berechnungsdurchgang

Konfidenzintervall-Ellipsen erster Berechnungsdurchgang UT „Funktionen Abstrahieren“

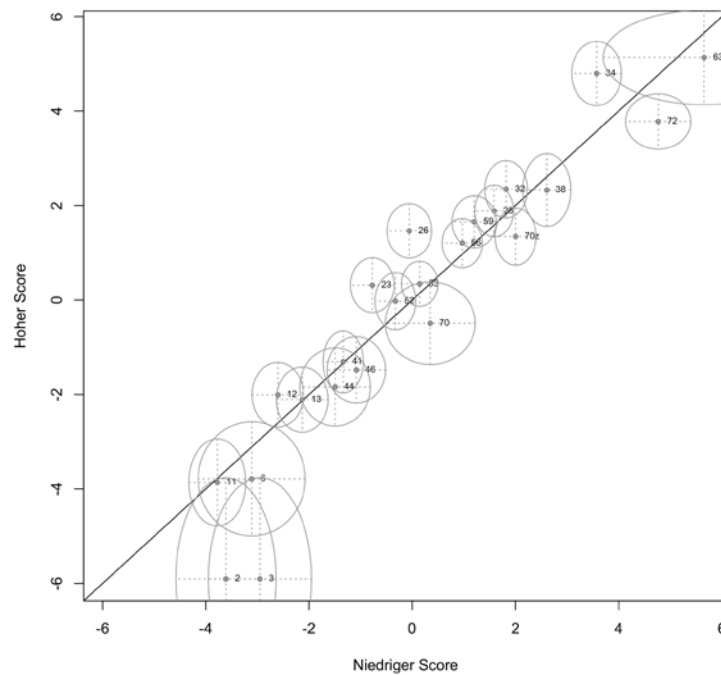


Abbildung 25: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 1 erster Berechnungsdurchgang

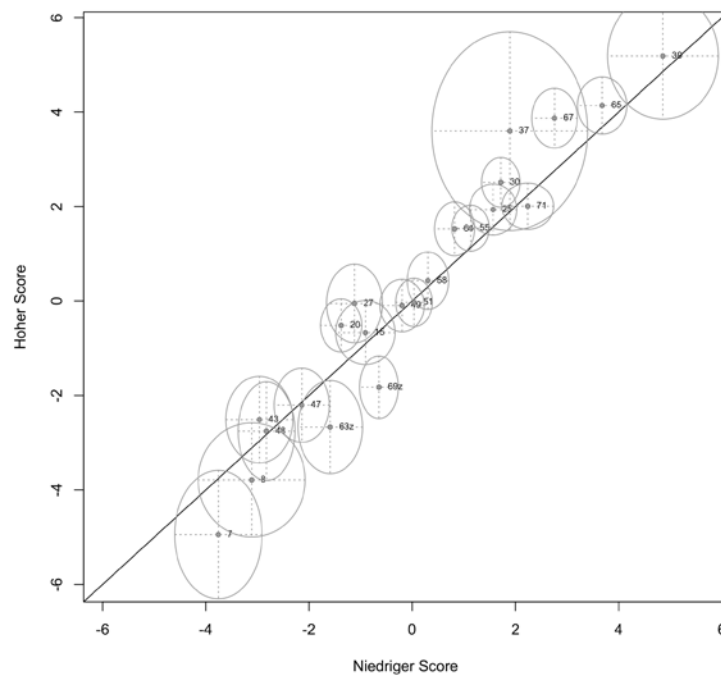


Abbildung 26: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 2 erster Berechnungsdurchgang

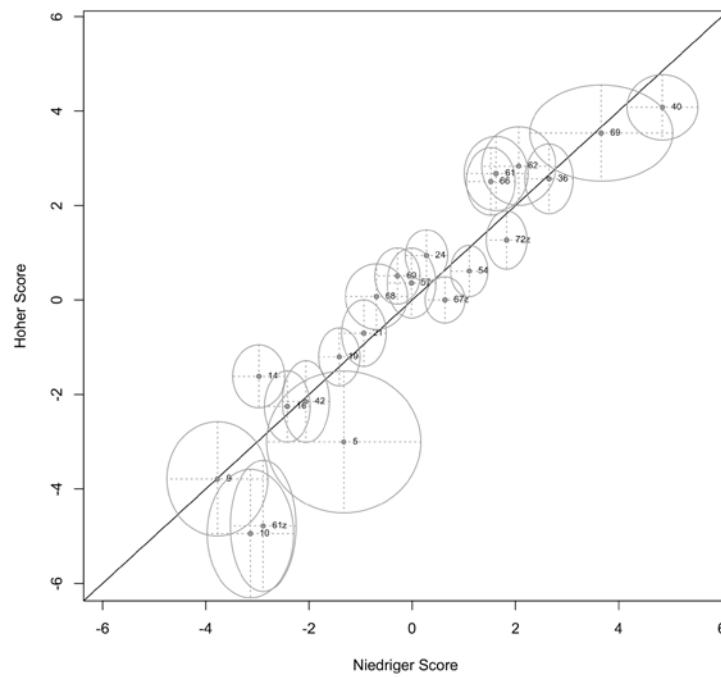


Abbildung 27: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 3 erster Berechnungsdurchgang

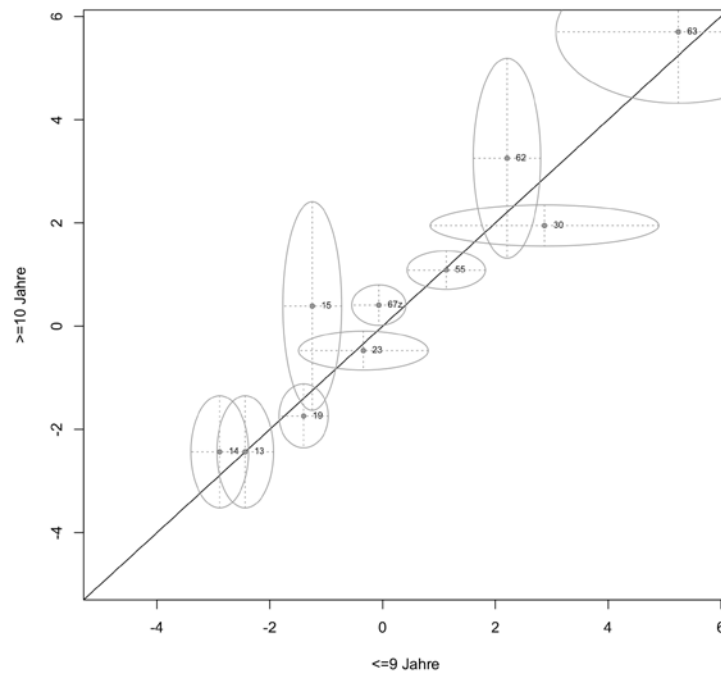


Abbildung 28: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 1 erster Berechnungsdurchgang

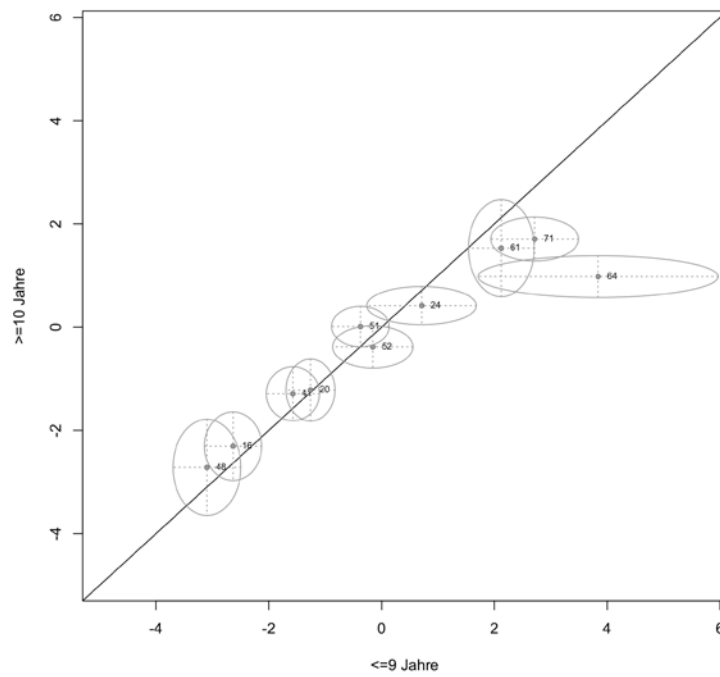


Abbildung 29: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 2 erster Berechnungsdurchgang

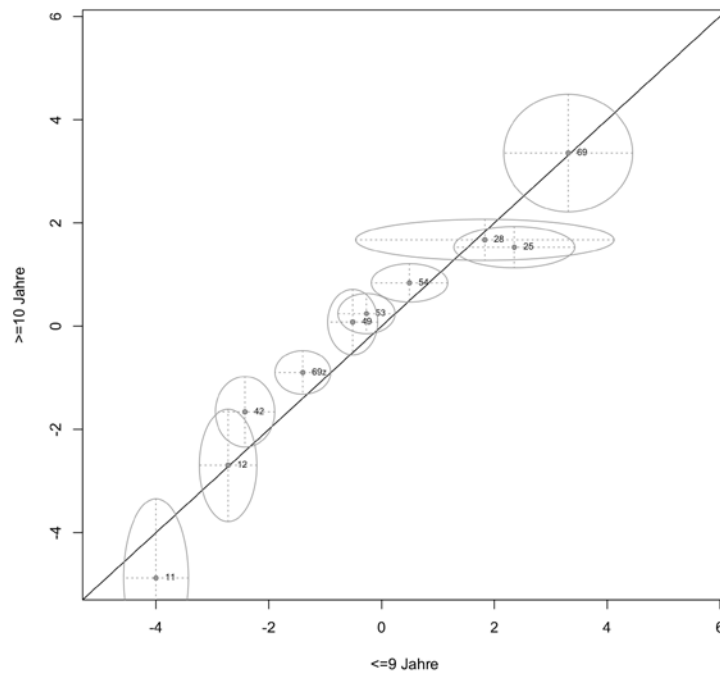


Abbildung 30: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 3 erster Berechnungsdurchgang

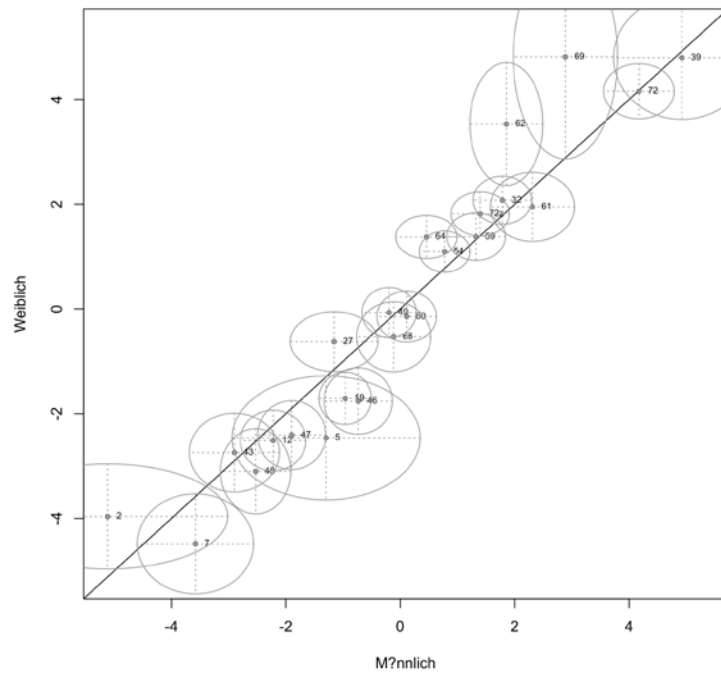


Abbildung 31: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 1 erster Berechnungsdurchgang

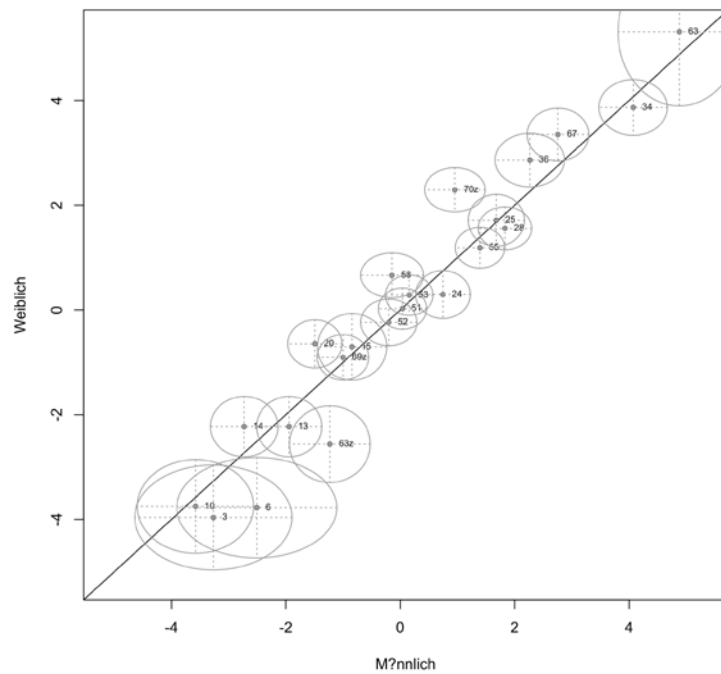


Abbildung 32: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 2 erster Berechnungsdurchgang

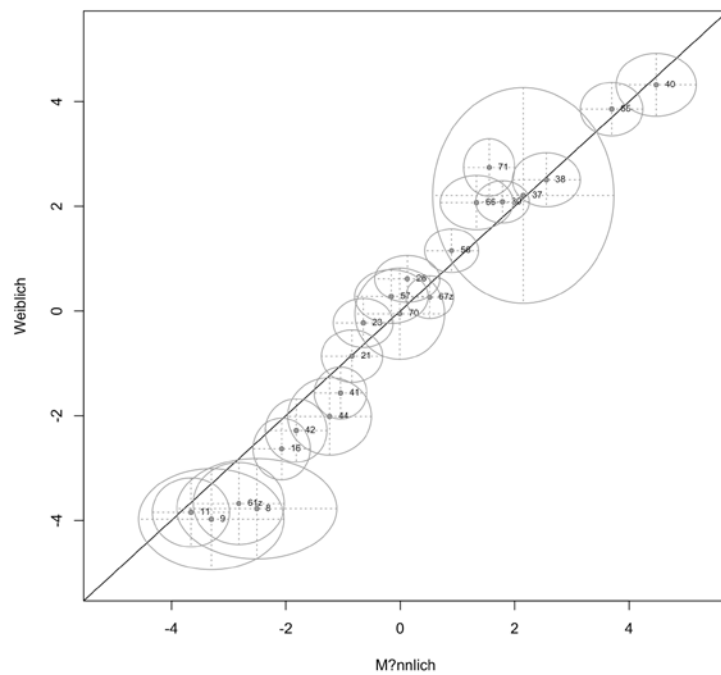


Abbildung 33: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 3. erster Berechnungsdurchgang

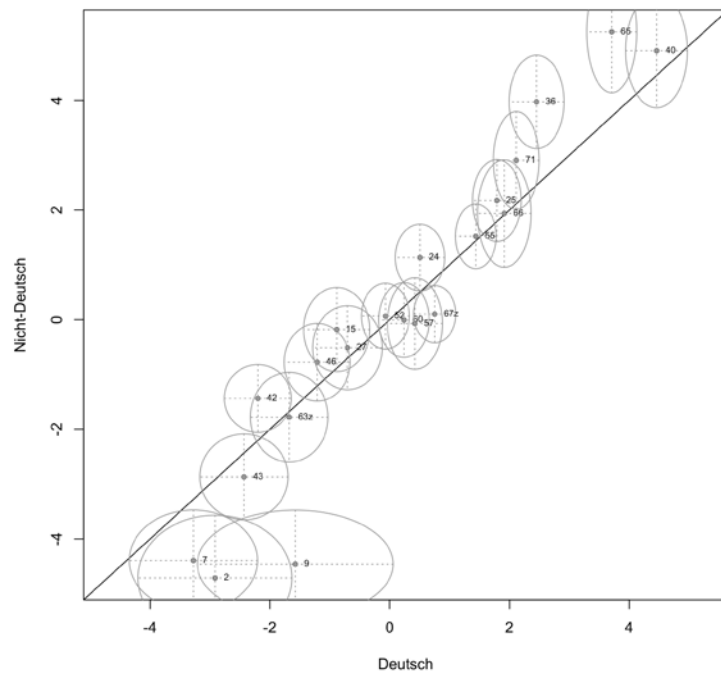


Abbildung 34: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 1. erster Berechnungsdurchgang

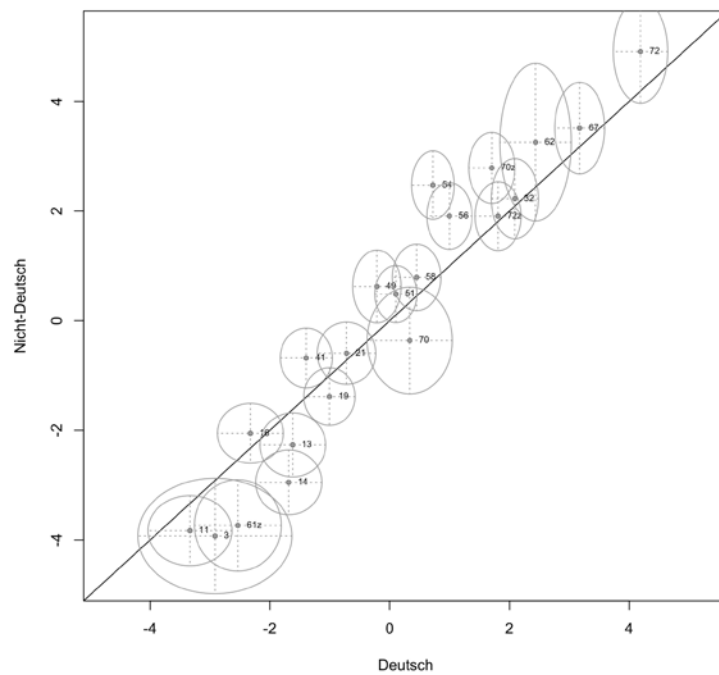


Abbildung 35: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 2 erster Berechnungsdurchgang

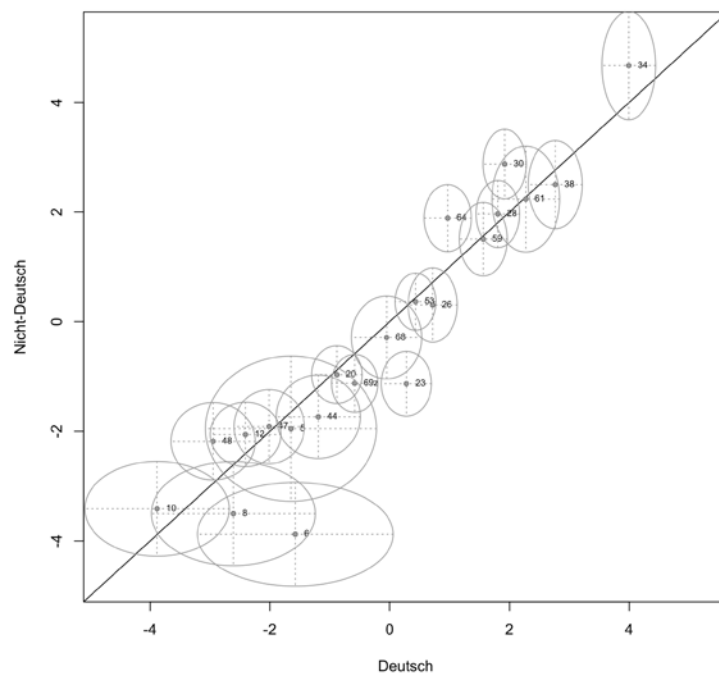


Abbildung 36: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 3 erster Berechnungsdurchgang

Konfidenzintervall-Ellipsen letzter Berechnungsdurchgang UT „Funktionen Abstrahieren“

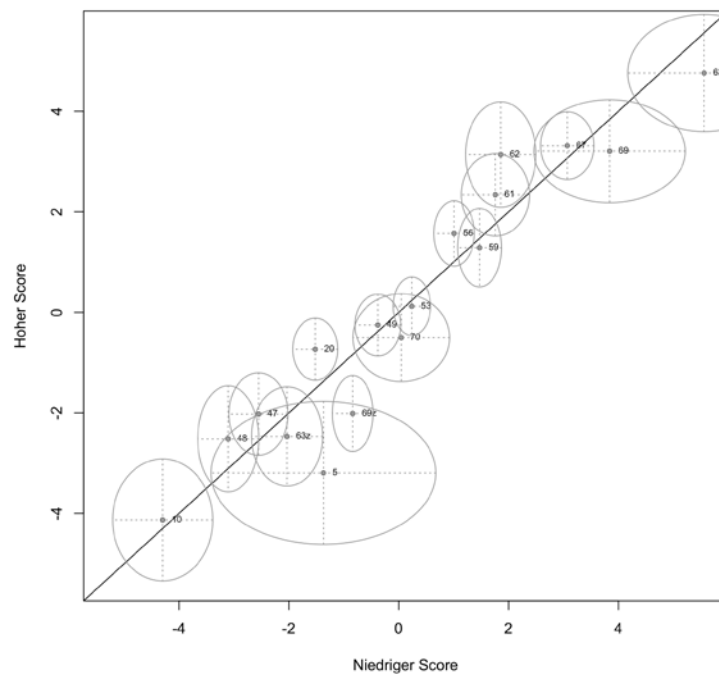


Abbildung 37: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 1 letzter Berechnungsdurchgang

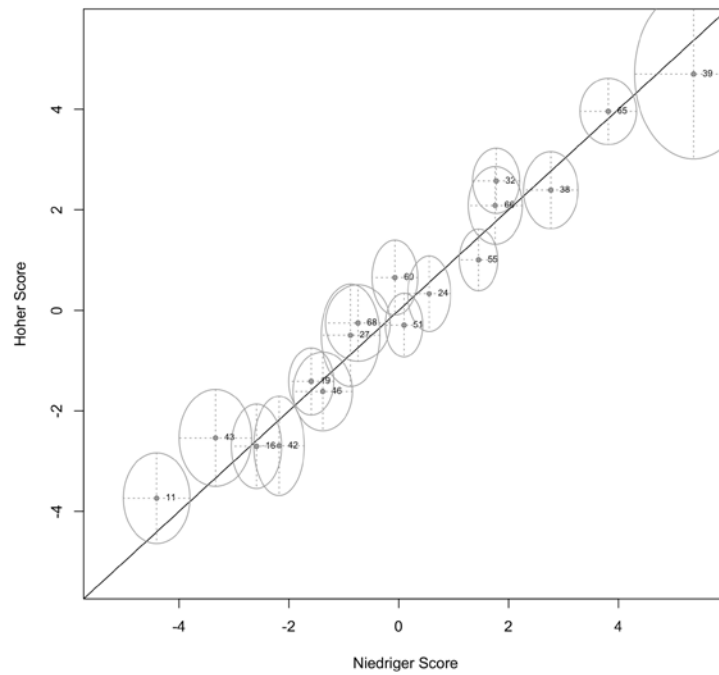


Abbildung 38: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 2 letzter Berechnungsdurchgang

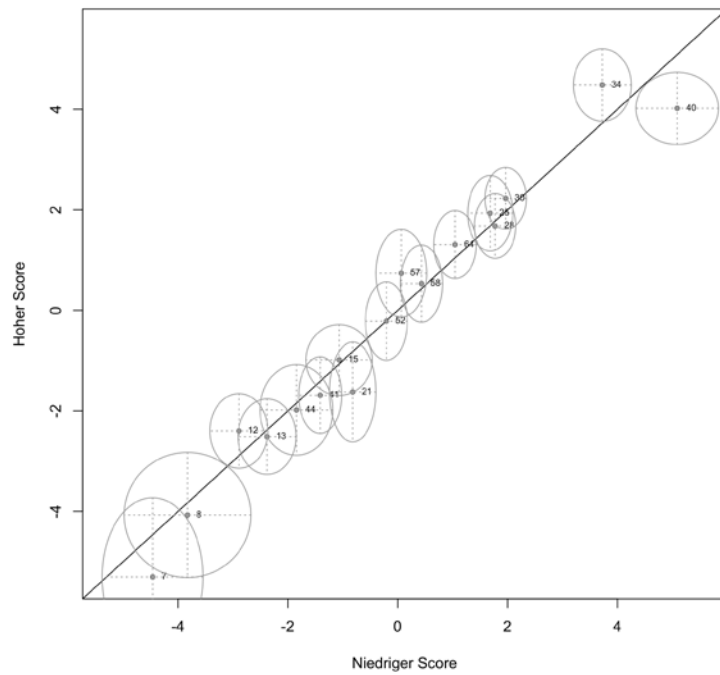


Abbildung 39: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Rohscore mit Konfidenzintervall-Ellipsen 3 letzter Berechnungsdurchgang

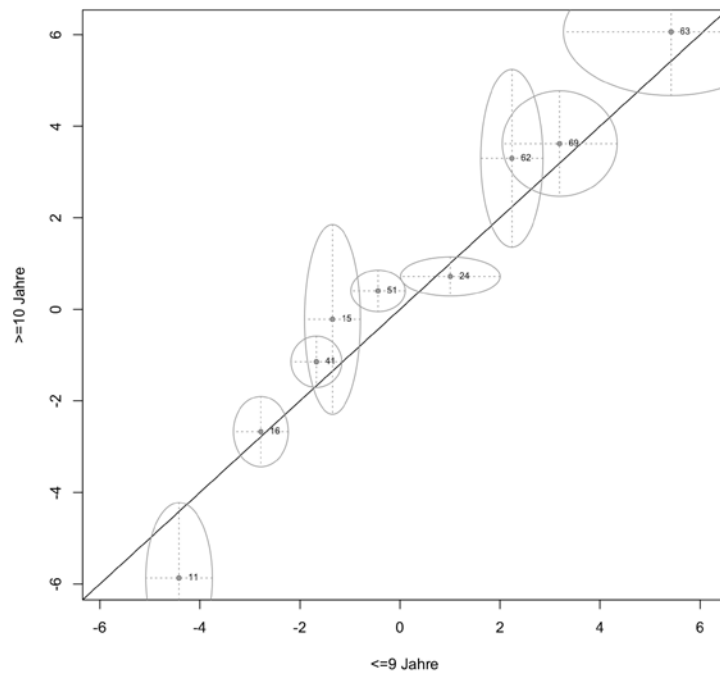


Abbildung 40: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 1 letzter Berechnungsdurchgang

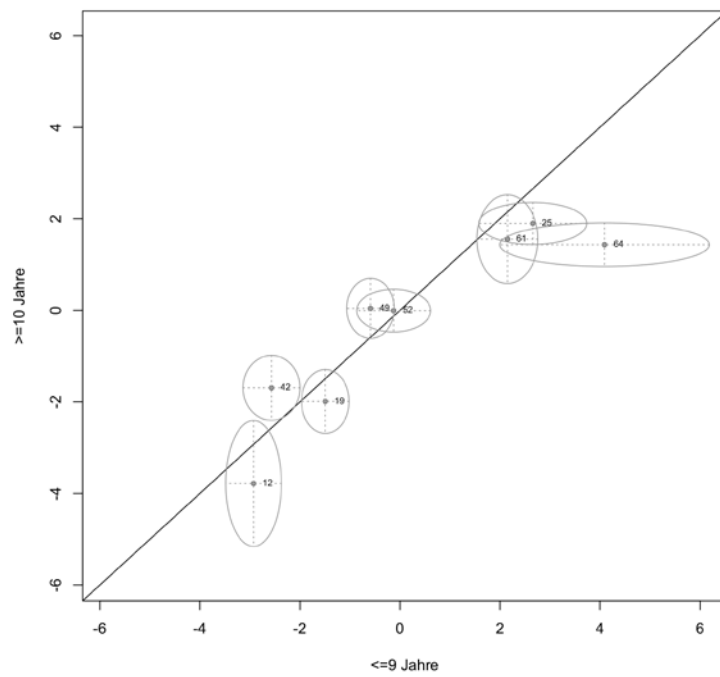


Abbildung 41: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 2 letzter Berechnungsdurchgang

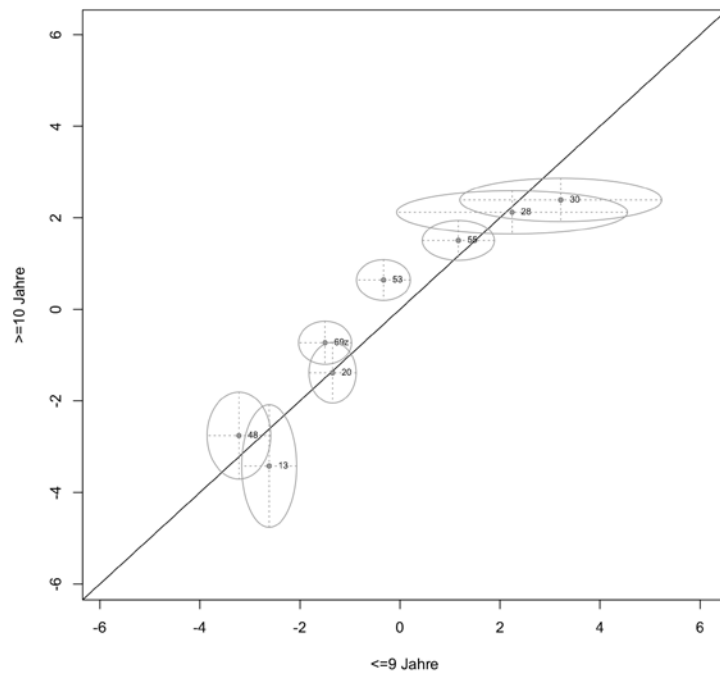


Abbildung 42: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Alter mit Konfidenzintervall-Ellipsen 3 letzter Berechnungsdurchgang

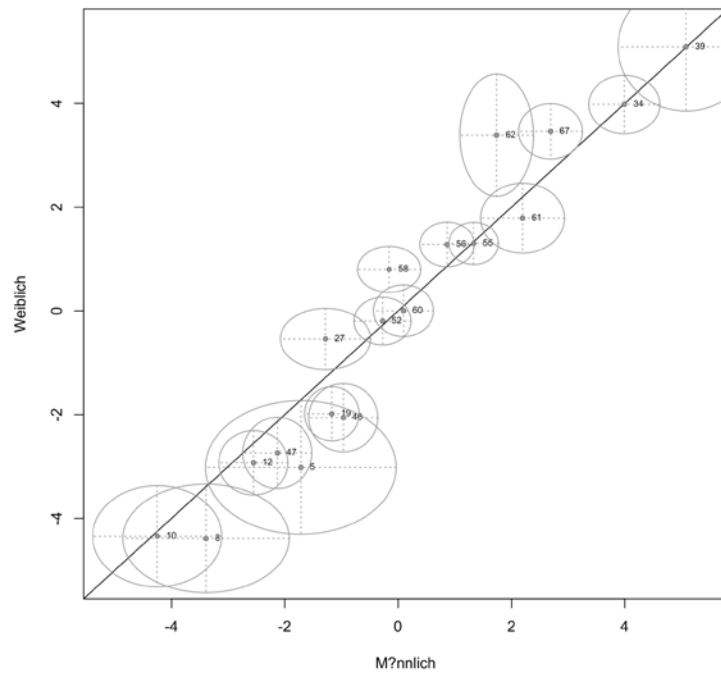


Abbildung 43: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 1 letzter Berechnungsdurchgang

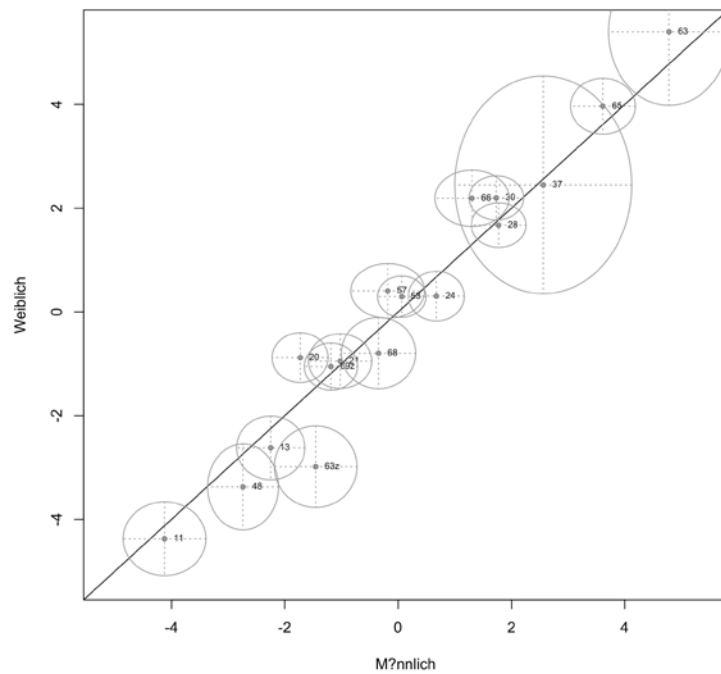


Abbildung 44: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 2 letzter Berechnungsdurchgang

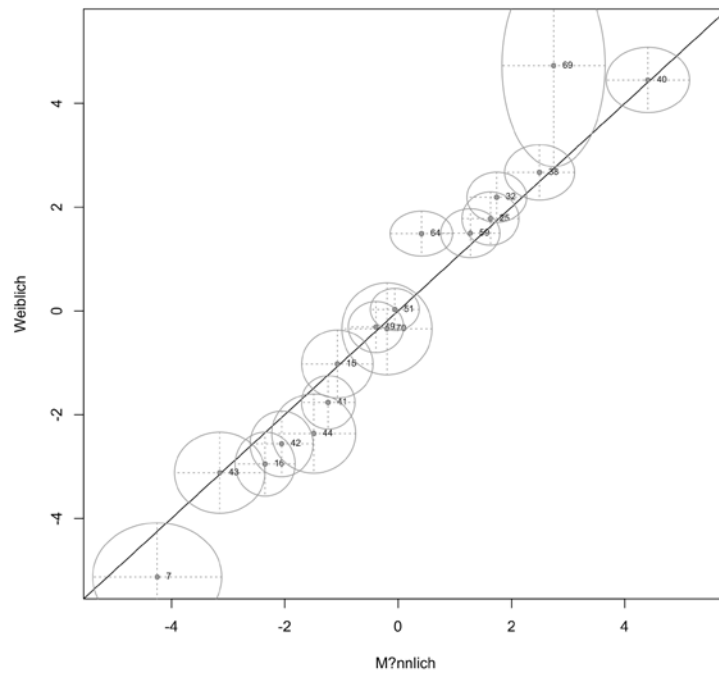


Abbildung 45: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Geschlecht mit Konfidenzintervall-Ellipsen 3 letzter Berechnungsdurchgang

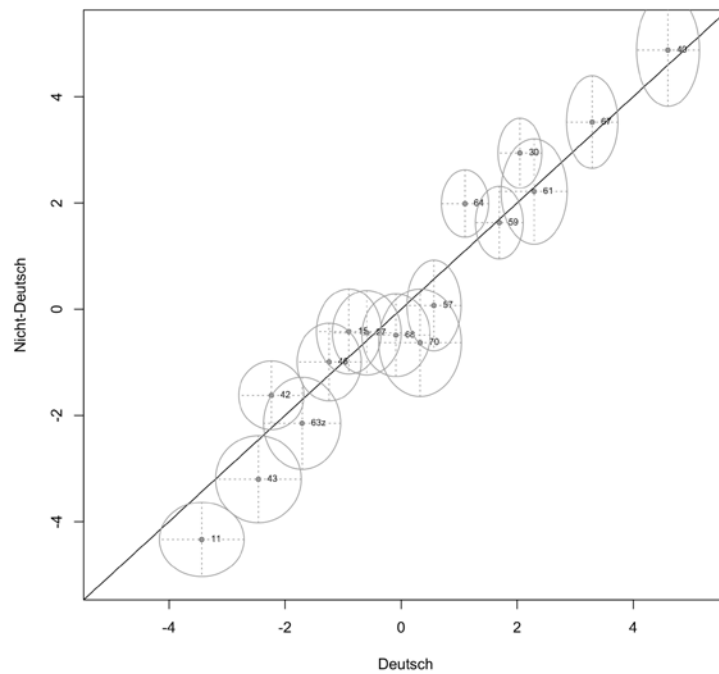


Abbildung 46: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 1 letzter Berechnungsdurchgang

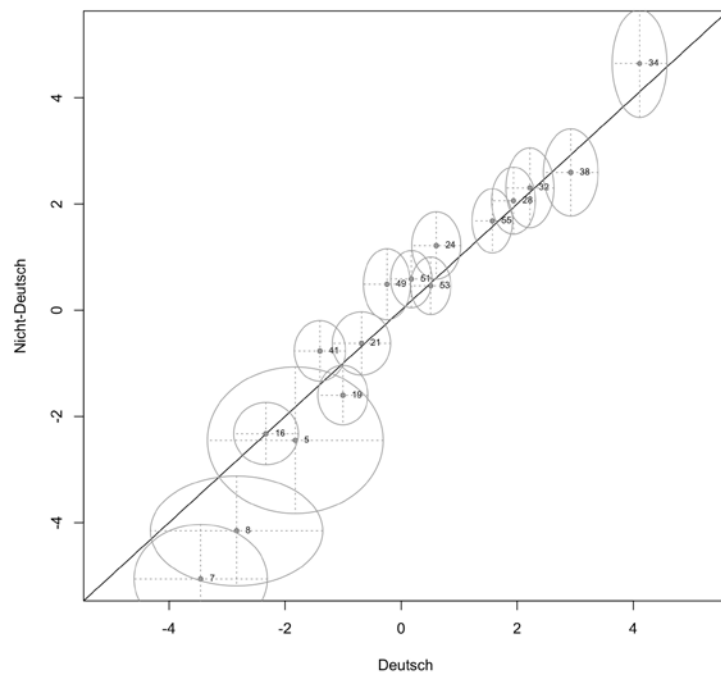


Abbildung 47: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 2 letzter Berechnungsdurchgang

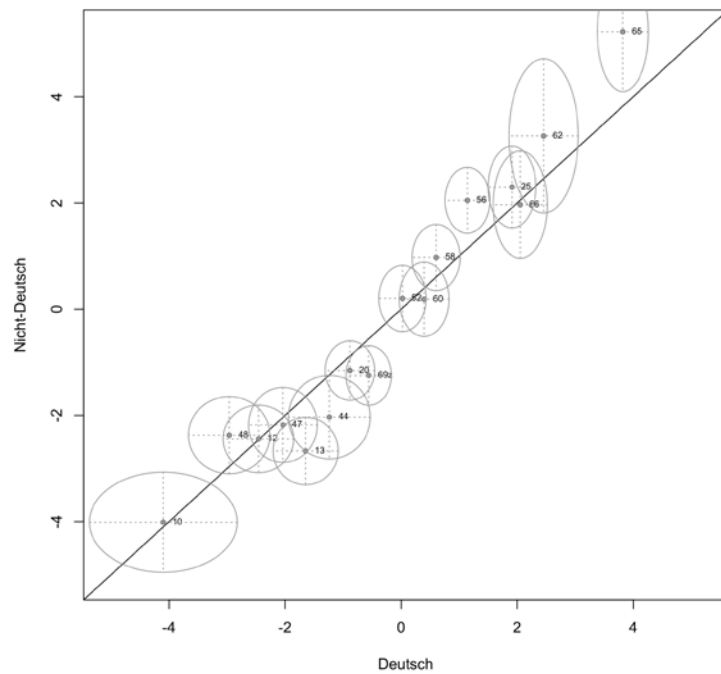


Abbildung 48: UT "Funktionen Abstrahieren" graphischer Modelltest für das Teilungskriterium Sprache mit Konfidenzintervall-Ellipsen 3 letzter Berechnungsdurchgang

Lebenslauf



Persönliche Daten:

Name: Nicole Görner
Geburtsdatum: 24.10.1974
Geburtsort: Elsterwerda
Staatsbürgerschaft: Deutsch
Wohnort: 1220 Wien, Ziegelhofstrasse 21
Telefon: 0043 664 146 98 01
E-Mail: nicole.goerner@gmx.at
Familienstand: ledig
Kinder: Lukas Schostal; geb. 23.08.2000
Lena Schostal; geb. 21.08. 2001

Ausbildungsweg:

1981-1990 Oberschule Elsterwerda
1990-1991 Leistungsklasse Elsterwerda
1991-1993 Elsterschlossgymnasium Elsterwerda
Abschluss mit Abitur
1993-1994 TU Dresden, Studium der Lebensmittelchemie
Ohne Abschluss
1994-1997 Labor: Zahntechnik GmbH Elsterwerda
Schule: Oberstufenzentrum II Cottbus
Ausbildung zur Zahntechnikerin
Abschluss mit Gesellenprüfung
2005-2013 Hauptuniversität Wien, Studium der Psychologie
6/2008 Erstes Diplomprüfungszeugnis

Beruflicher Werdegang:

3/1997-7/1997 Zahntechnik GmbH Elsterwerda
Fachbereich: Edelmetall, Keramik
8/1997-1/1998 Semper Bio Oral Axel Decker GmbH Dresden
Fachbereich: Edelmetall, Keramik
2/1998-6/1998 Bio-Oral Dental GmbH Dresden
Fachbereich: Edelmetall, Keramik
3/1999-7/2000 Prodent GmbH Wien
Fachbereich: Keramik
6/2009-6/2010 Projektassistentin am Institut für
Entwicklungspsychologie und Psychologische

Diagnostik, Arbeitsbereich Psychologische Diagnostik
im Rahmen der Diplomarbeit

9/2010-11/2010

Praktikum an der Test- und Beratungsstelle der
Universität Wien

seit 4/2013

Psychologiepraktikantin bei Frau Mag. Erika Barker-
Benfield; Praxis für Legasthenie, ADHS und
Kinderpsychologie

Karenzurlaub:

8/2000-8/2001

Geburt Lukas Schostal 23.08.2000

8/2001-2/2004

Geburt Lena Schostal 21.08.2001

Lehrgänge:

2/1996

Lehrgang für das Aufstellen von Totalprothesen
in Dorsten-Wulfen (BRD)

2000

Kurs zur Vorbereitung auf die Unternehmerprüfung
WIFI Wien

10-12/2001

Arbeits- und Sozialrecht in Theorie und Praxis
Arbeiterkammer Wien

12/2002-1/2003

EDV-Grundschulung

Internet und E-Mail

Word Einführung

Arbeiterkammer Wien

10/2003

Grundlagen erfolgreicher Rhetorik

Volkshochschule Wien-West

2/2011

Zertifizierungskurs zum ausgebildeten AID 2 Testleiter
Universität Wien

**Besondere Kenntnisse und
Fähigkeiten:**

EDV-Kenntnisse, ECDL, SPSS, R

Englisch, Russisch (Grundkenntnisse)

Führerschein B und C

Wien, 16.04.2013