



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Validierung mittels *Change*-Ansatz
am Bsp. des Untertests 1 des AID 3“

verfasst von / submitted by

Isabella Plank, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science, MSc

Wien, 2018 / Vienna, 2018

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 840

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Psychologie

Betreut von / Supervisor:

Univ.-Prof. i.R. Mag. Dr. Klaus Kubinger

Danksagung

Es gibt viele Personen, ohne die das Gelingen meiner Masterarbeit nicht möglich gewesen wäre.

Mein Dank gilt...

den Schülerinnen und Schülern, die gerne mitgemacht und mir durch ihre Wissbegier viel Freude bereitet haben,

den Eltern und Erziehungsberechtigten, die dies erst ermöglicht haben,

den Lehrerinnen und Lehrern, für ihre Bereitschaft an meiner Studie teilzunehmen,

den beiden Direktorinnen Mag. Pichler und Mag. Rieger, die mich in meinem Vorhaben vom ersten Moment an engagiert unterstützt haben,

Fr. Schindler, für ihr hilfsbereites Wesen - damals wie heute,

Univ.-Prof. i.R. Mag. Dr. Kubinger, für das Betreuen dieser Arbeit,

Thomas, für wertvolle Ideen, Rat und Feedback,

Katarina, für das Korrekturlesen der Abstracts,

dem kleinen Sam, für sein sonniges Gemüt und seine Gesellschaft beim Schreiben dieser Arbeit,

meinem Freund Mario, für seine Unterstützung und Rücksichtnahme,

und meinen Eltern Christine und Johann, die in jeglicher Hinsicht meine große Stütze sind und immer alles für mich gegeben haben.

Inhaltsverzeichnis

1	Einleitung.....	1
2	Theoretischer Hintergrund.....	3
2.1	Validität	3
2.2	Konstruktvalidität, Konstruktrepräsentation	4
2.3	Der <i>Change</i> -Ansatz	6
3	Fragestellungen und Hypothesen	9
3.1	Fragestellung	9
3.2	Hypothesen.....	9
4	Methode	10
4.1	Untersuchungsdesign	10
4.2	Stichprobenbeschreibung	11
4.3	Messinstrument	13
4.3.1	AID 3 - Adaptives Intelligenz Diagnostikum 3 (Version 3.1).....	13
4.3.2	Untertest 1 „Alltagswissen“	15
4.3.3	Der <i>Change</i> -Ansatz am Bsp. „Alltagswissen“	15
4.4	Erhebung demographischer Daten.....	19
4.5	Vorgehen	19
4.6	Informationen über die Art der Auswertung	21
4.6.1	Item-Response-Theorie	21
4.6.2	Rasch-Modell	23
4.6.3	Dreifache Varianzanalyse als Rasch-Modell-Test	25
5	Ergebnisse	27
5.1	Rasch-Modell-Analysen.....	27

5.1.1	Ergebnisse der dreifachen Varianzanalyse	27
5.1.2	Schätzung der Itemleichtigkeitsparameter	28
6	Diskussion	41
7	Verzeichnisse	46
7.1	Literaturverzeichnis	46
7.2	Abbildungsverzeichnis	49
7.3	Tabellenverzeichnis	50

Anhang

- A Abstract – deutsche Version
- B Abstract – english version
- C *Change*-Items
- D Protokollbogen
- E Elternbriefe
- F Informationsschreiben für Lehrkräfte

1 Einleitung

Das Konzept der Validität als Gütekriterium psychologisch-diagnostischer Verfahren bzw. dessen Auffassung beschäftigt Forschende seit vielen Jahrzehnten.

Im geschichtlichen Rückblick zeigen Newton und Shaw (2013), dass - trotz gegebener Standards im Denken über Validität - die Aussicht recht gering ist, Einigkeit darüber zu erlangen, wie über Validität kommuniziert wird. Genauso vielfältig sind Methoden zur Validierung.

Viele Studien zur Untersuchung der Konstruktvalidität beschäftigen sich damit, wie die Ergebnisse eines Tests mit anderen relevanten Variablen in Zusammenhang stehen. Doch auch ohne Einbezug weiterer Variablen können Aussagen zur Konstruktvalidität getroffen werden.

Entsprechende Überlegungen formulierte Embretson bereits 1983.

Sie befasste sich damit, welche kognitiven Prozesse zur Lösung einer Aufgabe erforderlich sind - und zwar in Abhängigkeit der Eigenschaften dieser Aufgabe.

In einem 2017 erschienenen Artikel von Engelhardt, Goldhammer, Naumann und Frey präsentieren die Autorin und die Autoren zwei neuartige experimentelle Ansätze zur Validierung, welche auf Embretsons Überlegungen zur Konstruktrepräsentation (1983, 1998) fußen.

Diese Ansätze sind insbesondere dann praktikabel, wenn es um Verfahren zur Erfassung komplexer Konstrukte geht – ein Umstand, der den Einsatz von Items erforderlich macht, die sich hinsichtlich ihrer konstituierenden Eigenschaften deutlich voneinander unterscheiden.

Einen ihrer beiden Zugänge bezeichnen Engelhardt et al. (2017) als *Change-Ansatz*. Dieser wird in der vorliegenden Masterarbeit dazu genutzt, eine Validierung der Testwertinterpretationen des Untertests „Alltagswissen“ der Intelligenz-Testbatterie *Adaptives Intelligenz Diagnostikum 3* (AID 3, Version 3.1; Kubinger & Holocher-Ertl, 2014) vorzunehmen.

Interessant ist der *Change-Ansatz* im Hinblick auf den AID 3 (Kubinger & Holocher-Ertl, 2014), da dieser das sehr komplexe Konstrukt Intelligenz anhand unterschiedlicher Untertests zu erfassen versucht. Dies führt zu einer Verschiedenartigkeit der innerhalb der Untertests verwendeten Items, wobei der

von Engelhardt et al. (2017) vorgestellte Ansatz theoretisch auf jedes dieser Items angewandt werden kann.

2 Theoretischer Hintergrund

2.1 Validität

Die Validität stellt das grundlegende Gütekriterium psychologisch-diagnostischer Verfahren dar. Gleichzeitig ist es am schwierigsten zu erfassen.

Kubinger (2009) folgt einem traditionellen Standpunkt, indem er angibt: „Unter *Validität* eines Tests ist zu verstehen, dass er tatsächlich jenes psychische Merkmal misst, welches er zu messen behauptet.“ (S. 55)

In jüngerer Zeit wird eine andere als diese gängige Auffassung diskutiert.

So halten es Hartig, Frey und Jude (2012) für angebracht, von der „Validität (Gültigkeit) verschiedener *möglicher Interpretationen* von Testergebnissen“ zu sprechen (S.144).

Validität ist in diesem Sinne nicht als Merkmal eines Tests per se zu verstehen, sondern bezieht sich immer nur auf jene Interpretationen und Schlussfolgerungen, die auf der Grundlage von Testergebnissen entstanden sind.

Auch Kane (2013) sieht die Notwendigkeit, eben jene Aussagen zu validieren, die basierend auf der Verwendung und Interpretation von Testwerten aufgestellt wurden.

Neben dem Grundsatz, nicht von „der Validität des Tests“ zu sprechen, adressieren Newton und Shaw (2013) einen weiteren wesentlichen Standard.

Spätestens seit 1985 wurde von Seiten der American Educational Research Association (AERA), der American Psychological Association (APA) und dem National Council on Measurement in Education (NCME) zwischen Aspekten und Arten von Validität unterschieden, wobei nur noch eine Art von Validität (Konstruktvalidität) als einheitliches Konzept postuliert wurde.

Diese Entwicklung beschreiben auch Hartig et al. (2012, S. 153) und bezeichnen Konstruktvalidität „seit Ende der 70er Jahre“ als „übergeordneten Zugang zur Bestimmung der Validität“.

Borsboom, Mellenbergh und Van Heerden (2004) plädieren ebenso dafür, nicht von verschiedenen Arten der Validität zu sprechen – ihrer Meinung nach wird

damit der Umstand adressiert, dass es verschiedene Wege gibt, Validierung zu betreiben.

Bereits 1999 waren die Bezeichnungen Inhalts-, Kriteriums- und Konstruktvalidität in den Standards der AERA, APA und NCME abgeschafft worden. Stattdessen sollte auf unterschiedliche Quellen Bezug genommen werden, aus denen Belege für Validität stammen.

2013 zeigten Newton und Shaw, dass diesem Standard in der Praxis und einschlägigen Literatur nicht Folge geleistet wird. Die Bezeichnungen für verschiedene Arten von Validität sind nicht nur sehr vielfältig, sondern allgegenwärtig, und finden sogar in Titeln von Zeitschriftenartikeln Verwendung.

Der Praxis folgend, wird im kommenden Abschnitt von Konstruktvalidität die Rede sein, wobei hiermit eher der historischen Entwicklung des Konzepts Rechnung getragen wird.

Zunächst sei noch betont, dass die Validierung eines Tests nach Hartig et al. (2012, S. 144) definiert ist als „Untersuchung der Validität von Testwertinterpretationen“.

2.2 Konstruktvalidität, Konstruktrepräsentation

Kubinger zufolge (2009, S. 57) kann einem Test Konstruktvalidität zugesprochen werden, wenn er „gewisse theoretische bzw. theoriegeleitete Vorstellungen in Bezug auf irgendein „Konstrukt““ erfüllt, wobei es sich bei Konstrukten um „nicht direkt beobachtbare „Phänomene“, wie z.B. Intelligenz“ handelt.

Hartig et al. (2012) sprechen im Zusammenhang mit psychologischen Testverfahren davon, dass die „erfassten Merkmale immer konstruierte Größen sind, die im Rahmen eines diagnostischen Anwendungskontexts und/oder einer psychologischen Theorie definiert werden.“ (S. 145)

Im Jahr 1955 unternahmen Cronbach und Meehl einen ersten einflussreichen Versuch, das zu diesem Zeitpunkt neue Konzept der Konstruktvalidität in einem

wissenschaftlichen Artikel zu charakterisieren. Verstanden wurde die Konstruktvalidität damals als Ergänzung zu anderen bis dato etablierten Untertypen der Validität.

Den Fokus legten die Autoren auf die Frage, welche Konstrukte Varianz in der Testleistung bedingen.

Hartig et al. (2012) fassen zusammen, dass Cronbachs und Meehls Ansatz der Konstruktvalidierung (1955) ein Ideal darstellt, dem in der Praxis nur selten entsprochen wird. So geht es den beiden Autoren um die schrittweise Überprüfung eines nomologischen Netzes, welches sich aus Bereichen der Theorie und Beobachtung zusammensetzt.

Aus der Theorie abgeleitete Hypothesen stellen dabei Annahmen über die Zusammenhänge von Konstrukten dar.

1983 legte Embretson dar, psychologisches Theoretisieren habe ein neues Ziel – es gehe nicht mehr nur darum, Zusammenhänge zu erforschen, sondern darum, Leistung durch zugrunde liegende Prozesse zu erklären.

Diesem veränderten Denken wollte sie Rechnung tragen und postulierte zwei separate Aspekte von Konstruktvalidität:

Konstruktrepräsentation und Nomothetische Spanne.

Eine Schwachstelle in Cronbachs und Meehls Ansatz sah Embretson (1983) darin, dass Studien, die die Beziehung von Testwerten mit anderen Variablen untersuchen, erst durchgeführt werden können, nachdem der Test bereits entwickelt worden ist.

Ihrer Auffassung nach bedarf es Methoden, Validierung in die Phase der Testentwicklung zu integrieren.

Nomothetische Spanne, welche sich mit dem Netzwerk von Beziehungen (Korrelationen) eines Testscores zu anderen Maßen beschäftigt, sah sie als zweitrangig an, und erweiterte ihre Sichtweise von Validierung durch das Konzept der Konstruktrepräsentation.

Darunter versteht sie die Identifikation theoretischer Mechanismen, welche einer Testleistung zugrunde liegen. Dieser Ansatz beschäftigt sich demnach mit Aufgabenzerlegung. Zu den Mechanismen, die verwendet werden, um Items zu lösen, zählen Prozesse, Strategien und Wissen (Embretson, 1998).

Auch heute erfolgt die Untersuchung der Konstruktvalidität häufig durch die Überprüfung theoriebasiert aufgestellter Hypothesen - und zwar in Bezug darauf, wie die Ergebnisse eines Tests mit anderen Variablen in Zusammenhang stehen.

Im Unterschied dazu führen Hartig et al. (2012) als Validierungsmethode eine Untersuchung auf Itemebene an, für die es nicht notwendig ist, zusätzliche Variablen mit einzubeziehen.

Einer dieser Ansätze, der auf den Überlegungen von Embretson (1983, 1998) basiert, ist für die vorliegende Arbeit Vorbild und wird im folgenden Abschnitt näher beschrieben.

2.3 Der *Change*-Ansatz

2015 stellten Engelhardt, Goldhammer, Naumann, Frey, Hartig, Horz, Kuchta und Wenzel ihre Studie *Konstruktvalidierung durch die Manipulation der Aufgabenschwierigkeit in simulationsbasierten ICT-Aufgaben* vor.

Eine Weiterentwicklung der Analysen präsentierten Engelhardt et al. 2017.

In diesem Artikel bezeichnen die Autorin und die Autoren eine der zwei vorgestellten Methoden zur Validierung heterogener Item-Sets als *Change*-Ansatz.

Der Ansatz geht auf Überlegungen Embretsons zur Konstruktrepräsentation (1983) zurück und sieht vor, Varianten von Items zu kreieren, indem systematisch Aufgabeneigenschaften bereits existierender Items zu Validierungszwecken manipuliert werden (Engelhardt et al., 2017, S. 683).

Als neuartig an ihrer Methode betonen die Autorin und die Autoren den Vorteil, dass er auf heterogene Item-Sets angewandt werden kann. Heterogen bedeutet in diesem Zusammenhang, dass sich die Items hinsichtlich ihrer konstituierenden Eigenschaften so voneinander unterscheiden, dass es schwer fällt, festmachen zu können, welches Item-Merkmal Unterschiede in der Itemschwierigkeit bedingen könnte.

Die Basis stellt der Ansatz der Konstruktrepräsentation (Embretson, 1983) dar, in welchem Itemeigenschaften mit Itemschwierigkeit in Zusammenhang stehen.

Hierbei wird überlegt, welche kognitiven Prozesse oder Wissensinhalte zur Lösung einer Aufgabe notwendig sind, und welche Eigenschaften einer Aufgabe diese Prozesse hervorrufen.

Der Name *Change* bezieht sich darauf, dass für ein bereits existentes Item durch Veränderung eine neue Variante (in weiterer Folge als *Change*-Item bezeichnet) erstellt wird, wobei die als schwierigkeitsbestimmend vermutete Aufgabeneigenschaft geändert wird. Schwierigkeitsbestimmend insofern, als angenommen wird, dass sie jenes Konstrukt hervorruft, welches für die Unterschiede im Testwert verantwortlich ist.

Je nach Art - also ob das Vorhandensein des konstruktrepräsentativen Merkmals in geringerer oder höherer Ausprägung erforderlich ist als beim Original - sollte die Beantwortung des *Change*-Items den Testpersonen leichter oder schwerer fallen. Dies ist dann der Fall, wenn die konstruktrepräsentativen Merkmale tatsächlich einen Einfluss auf die Lösungswahrscheinlichkeit des Items haben.

Für die Funktion der *Change*-Items resultiert wie folgt:

Das *Change*-Item soll, je nach Richtung der Veränderung, einfacher (*Change-e*-Item) oder schwieriger (*Change-s*-Item) sein als dessen Original.

Darüber hinaus sollen die *Change*-Items nach wie vor dasselbe Konstrukt messen wie die Originale.

Regeln für die Veränderung aufzustellen, ist schwer möglich, wenn nicht alle Items demselben Typus angehören (strukturell nicht gleich sind). Dies ist jedoch auch nicht notwendig, da alle Manipulationen individuell und voneinander unabhängig durchgeführt werden können. Auch mehr als eine *Change*-Variante pro Item ist denkbar, wobei sich sowohl Richtung als auch Grad der Manipulation unterscheiden können.

Da die *Change*-Items dasselbe Konstrukt erfassen wie die dazu gehörigen Originale, liegt ein weiterer Vorteil dieser Validierungsmethode darin, dass die *Change*-Items - falls in der Phase der Testentwicklung generiert - zum Testen verwendet werden können.

Dies ist insbesondere relevant für adaptives Testen, wo ein Item-Pool mit einer großen Schwierigkeits-Range benötigt wird (siehe Abschnitt 4.6.1).

3 Fragestellungen und Hypothesen

3.1 Fragestellung

Es werden - zu Validierungszwecken - Effekte der systematischen Manipulation von Aufgabeneigenschaften auf die Itemschwierigkeit untersucht.

Dies geschieht, indem schwierigkeitsbestimmende Elemente in den Original-Items des Untertests 1 "Alltagswissen" des AID 3 (Kubinger & Holocher-Ertl, 2014) verändert und die resultierenden Itemschwierigkeiten in den *Change*- und Original-Items verglichen werden.

Ein Beleg für eine gelungene Validierung liegt vor, wenn ein Unterschied in den Itemschwierigkeiten resultiert und die Items nach wie vor dasselbe Konstrukt messen.

3.2 Hypothesen

Hypothese 1:

Die *Change*-Items erfassen dasselbe Konstrukt wie die Original-Items.

Umgesetzt wird dies mittels Prüfung der Geltung des Rasch-Modells für die verwendeten Items, sodass:

H1₀(AxC): Es gibt keine Wechselwirkung zwischen den Faktoren A (z.B.: Geschlecht) und C (Items) in Bezug auf die Testwerte im „Alltagswissen“.

Hypothese 2:

Die Veränderung der Aufgabeneigenschaften hat eine Auswirkung auf die Itemschwierigkeit in beabsichtigter Richtung, sodass:

H2₁: *Change*-e-Items sind einfacher als Original-Items.

Change-s-Items sind schwieriger als Original-Items.

4 Methode

4.1 Untersuchungsdesign

Sämtliche Untertests der Intelligenztestbatterie AID 3 (Kubinger & Holocher-Ertl, 2014) – so auch der Untertest 1 „Alltagswissen“ – verfügen über eine Bandbreite an altersangepassten Items für Kinder und Jugendliche im Alter von 6;0 bis 15;11 Jahren.

Um die Anzahl an Items, die dem *Change*-Ansatz unterzogen werden sollten, von 60 (für die gesamte Alters-Range zur Verfügung stehenden) auf eine geringere Anzahl zu reduzieren, wurde die Untersuchung auf die Testung einer ausgewählten Altersgruppe begrenzt: die Altersgruppe der 10- und 11-Jährigen. Dementsprechend wurde geplant, SchülerInnen zu befragen, die zum Zeitpunkt der Erhebung die fünfte oder sechste Schulstufe besuchen.

Aufgrund der Beschaffenheit der Untersuchung wurden den Kindern einerseits Original-Items aus dem Untertest 1 „Alltagswissen“ vorgegeben; andererseits wurden ihnen Fragen gestellt, die durch gezielte Veränderung der Original-Items von der Verfasserin (durch Beteiligung und Feedback einer Arbeitsgruppe gestützt) selbst entwickelt wurden (*Change*-Items).

Da es bei einigen Items der Fall ist, dass die Originale und deren *Change*-Varianten sowie die dazu gehörigen Lösungen, einander (sehr) ähneln (oder ident sind) – dies geschah, um das damit erfasste Konstrukt beizubehalten – wurde festgelegt, einer Testperson nie alle drei Versionen ein- und desselben Items vorzugeben.

Allerdings sollte jede Testperson zwei Versionen ein- und desselben Items erhalten.

In der Folge wurde eine Aufteilung aller Items auf drei Testhefte vorgenommen - siehe Tabelle 4.

Detaillierte Angaben zu Auswahl, Modifikation, Anzahl und Reihung der verwendeten Items finden sich unter 4.3 Messinstrument.

Darüber hinaus stellt sich im Zuge der Untersuchungsplanung stets die Frage nach der erforderlichen Stichprobengröße.

Kubinger, Rasch und Yanagida (2009) legen dar, dass genaue Parameterschätzungen, sofern es das Rasch-Modell betrifft, auf einen extrem großen Stichprobenumfang angewiesen sind, wodurch es üblich ist, oft weit mehr als 200 Personen zu testen.

Auf der anderen Seite besteht die Gefahr, dass dies zu einem signifikanten Ergebnis des Modelltests führen kann – trotz praktisch vernachlässigbarer Modellabweichung.

In der vorliegenden Masterarbeit wird ein Ansatz der dreifachen Varianzanalyse für die Testung der Daten auf Rasch-Modell-Geltung (Kubinger et al., 2009, 2011b) angewandt, für den die Anzahl an Testpersonen nicht größer als zwei Mal 150 sein sollte.

Für die Planung resultierte, mehr als eine Schule zur Testung heranzuziehen, zumal nur die 5. und 6. Schulstufe von Interesse waren.

Insofern wurde mit der ersten Schule vereinbart, an insgesamt 175 Kinder (7 Klassen) herantreten zu dürfen, mit der zweiten Schule an etwa 60 Kinder (4 Klassen) – in Summe waren dies 235 angefragte Testpersonen, wobei selbstverständlich mit Abschlügen betreffend Teilnahmequote zu rechnen war.

4.2 Stichprobenbeschreibung

Insgesamt nahmen 127 SchülerInnen aus zwei Wiener Schulen unterschiedlicher Gemeindebezirke (1030 Wien, 1230 Wien) freiwillig an der Erhebung teil:

37 (29,1%) Kinder aus einer Wiener Mittelschule, 90 (70,9 %) Kinder aus einem Gymnasium.

Zum Zeitpunkt der Erhebung besuchten 54 (42,5%) der teilnehmenden Kinder die fünfte Schulstufe, 73 (57,5%) die sechste Schulstufe.

Die Stichprobe setzt sich hinsichtlich des Geschlechts aus 69 (54,3%) Schülern und 58 (45,7%) Schülerinnen zusammen.

Das Alter der teilnehmenden Kinder erstreckte sich von 10 bis 14 Jahren ($M=10,98$, $SD=0,831$). Die genaue Altersverteilung ist Tabelle 1 zu entnehmen.

99 (78%) Kinder gaben Deutsch als ihre Muttersprache an, 28 (22%) Kinder hatten eine andere Muttersprache (siehe Tabelle 2).

Tabelle 1. *Altersverteilung der Gesamtstichprobe in Jahren*

Alter	Absolute Häufigkeit	Anzahl in Prozent
10	37	29,1
11	63	49,6
12	21	16,5
13	5	3,9
14	1	0,8
Gesamt	127	100

Tabelle 2. *Verteilung der Muttersprache innerhalb der Gesamtstichprobe*

Muttersprache	Absolute Häufigkeit	Anzahl in Prozent
deutsch	99	78,0
serbisch	9	7,1
türkisch	5	3,9
afrikanisch	3	2,4
arabisch	2	1,6
chinesisch	2	1,6
thailändisch	1	0,8
andere slawisch	3	2,4
andere indogermanisch	3	2,4
Gesamt	127	100

Anmerkung.

andere slawisch = kroatisch, bulgarisch, polnisch

andere indogermanisch = albanisch, rumänisch, persisch

Die Verteilung der Testhefte erfolgte gleichmäßig über die gesamte Stichprobe hinweg und kann Tabelle 3 entnommen werden.

Tabelle 3. *Verteilung der Testhefte innerhalb der Gesamtstichprobe*

Testheft	Absolute Häufigkeit	Anzahl in Prozent
1	43	33,9
2	42	33,1
3	42	33,1
Gesamt	127	100

Die Erhebung erfolgte ohne Zwischenfälle oder Komplikationen.

Krankheits-, also abwesenheitsbedingt, konnten drei der angemeldeten Kinder nicht daran teilnehmen.

4.3 Messinstrument

4.3.1 AID 3 - Adaptives Intelligenz Diagnostikum 3 (Version 3.1)

Bei dem Adaptiven Intelligenz Diagnostikum 3 (Version 3.1) (Kubinger & Holocher-Ertl), kurz AID 3, handelt es sich um die 2014 erschienene 3. Auflage einer Intelligenz-Testbatterie für Kinder und Jugendliche im Alter zwischen 6;0 und 15;11 Jahren.

Ihr liegt der Anspruch zugrunde, „ziemlich viele (komplexe und basale) Fähigkeiten, die für „intelligentes“ Verhalten verantwortlich scheinen, zu erfassen“ (Kubinger & Holocher-Ertl, 2014, S. 7).

Diese Vielfältigkeit erlaubt eine Profilinterpretation in Bezug auf etliche Dimensionen von „Intelligenz“, wobei „Intelligenz“ als „Gesamtheit aller kognitiven Voraussetzungen, die notwendig sind, um Wissen zu erwerben und Handlungskompetenzen zu entwickeln“ (S. 26) definiert ist.

Eingesetzt wird der AID 3 hauptsächlich in der Schul- und Laufbahnberatung, sowie bei konkreten Fragestellungen der Schulpsychologie, im Rahmen derer er auf eine förderungsorientierte Diagnostik abzielt.

Nicht nur das Erkennen relativer Stärken eines Kindes wird ermöglicht, sondern auch das Identifizieren individueller Schwächen und somit jener Bereiche, die gezielter Förderung bedürfen.

Der AID 3 umfasst 12 Untertests und 5 Zusatztests, welchen operationale Definitionen zugrunde liegen - mittels dieser wurde begrifflich festgelegt, welche Fähigkeiten mit dem jeweiligen Untertest erfasst werden sollen.

Eine wesentliche Besonderheit des AID 3 liegt in der, innerhalb der Item-Response-Theorie verankerten, adaptiven Testvorgabe (siehe dazu Abschnitt 4.6.1).

Umgesetzt wurde diese in Form des branched testings, welches altersspezifische Startgruppen vorsieht. Der Testperson werden - zunächst alters- und in weiterer Folge antwort-, also leistungsabhängig - nur jene Aufgabengruppen vorgegeben, die ihrem Leistungsniveau entsprechen.

Wie dem Testmanual (Kubinger & Holocher-Ertl, 2014) zu entnehmen ist, wurde der AID 3 hinsichtlich seiner inhaltlichen Geltung mittels Experten-Ratings validiert.

Auch seine Konstruktvalidität in Bezug auf ein hierarchisches Modell zu Specific Learning Disorders wurde gezeigt.

Für die Vorgänger-Version, den AID, besteht diskriminante Konstruktvalidität (d.h. kein oder niedriger Zusammenhang) in Bezug auf etliche Leistungstests sowie Persönlichkeitsfragebögen.

In der vorliegenden Arbeit wird ein neuerer, experimenteller Ansatz der Validierung, der *Change*-Ansatz, auf den Untertest 1 „Alltagswissen“ angewandt, weshalb in weiterer Folge auch nur dieser Untertest näher beschrieben wird.

4.3.2 Untertest 1 „Alltagswissen“

Der Untertest 1 „Alltagswissen“ des AID 3 gehört zu jenen Untertests, bei dem die Anforderung verbal-akustisch eingebettet ist.

Die Testperson bekommt mündlich Fragen gestellt, die sie (ebenfalls mündlich) beantworten soll.

Das Antwortformat ist frei (es stehen keine Antwortmöglichkeiten zur Verfügung).

Die Testleiterin/der Testleiter vermerkt auf einem Protokollbogen, ob das Item gelöst wurde (1) oder nicht (0).

Die Bereiche, aus denen die Fragen stammen, umfassen die Themen Familie, Schule und Umwelt, Physik, Biologie, Wirtschaft und Politik, Geschichte und Geografie, Sport und Kunst.

Gemäß operationaler Definition soll mit diesem Untertest die Fähigkeit der Testperson überprüft werden, „sich Sachkenntnisse über Inhalte anzueignen, die in der heutigen Gesellschaft alltäglich sind“ (Kubinger & Holocher-Ertl, 2014, S. 8).

Erworbenes Wissen wird im AID als „Beleg für das Funktionieren wesentlicher Kognitionen“ aufgefasst, und stellt somit auch die Basis für „künftige kognitionsbedingte Leistungen“ (S. 8) dar.

Jene Aufgaben, die für diesen Untertest generiert worden sind, sollen somit in der Lage sein, „die Fähigkeit zu prüfen, erfahrungs- bzw. kenntnisbedingt über bestimmte Sachinhalte Bescheid zu wissen oder darüber reflektieren zu können“ (S.8), wobei in weiterer Folge ausgeführt wird, dass derart definiertes Verhalten Interesse in der Person voraussetzt.

4.3.3 Der *Change*-Ansatz am Bsp. „Alltagswissen“

4.3.3.1 Auswahl der Items

Die drei Aufgabengruppen vertikal zur (altersspezifischen) Startgruppe 10-11 wurden als Fixstarter gewählt: Blöcke 5, 12 und 17 (siehe Abbildung 1).

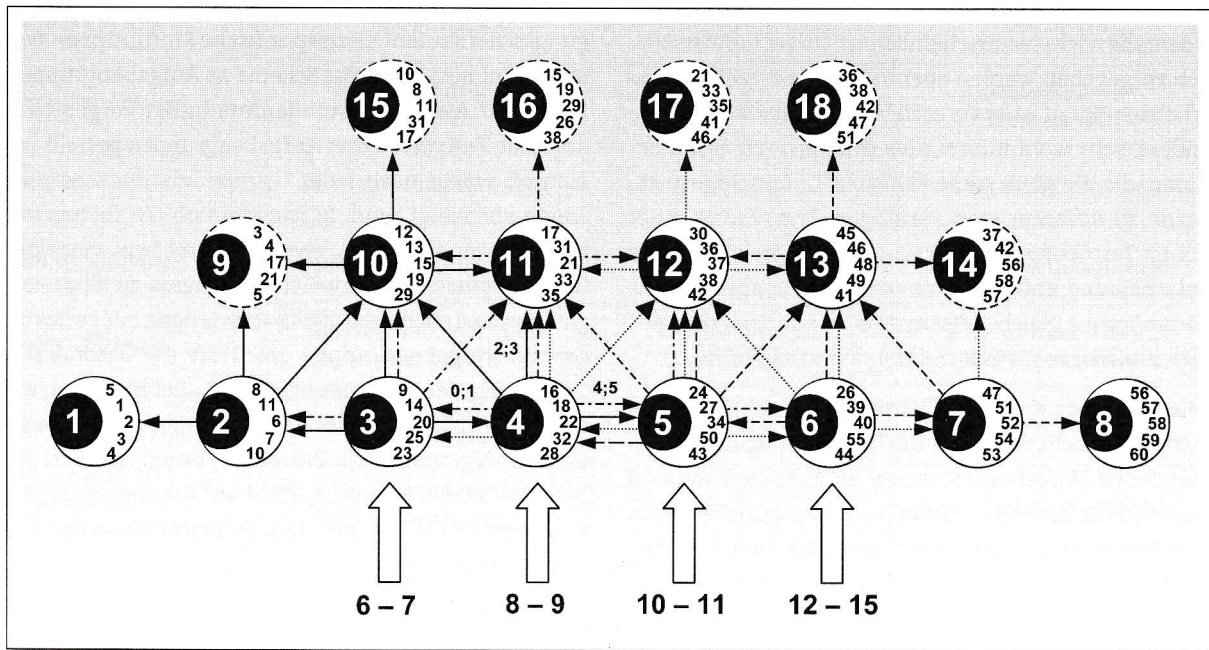


Abbildung 1. Das Prinzip der Testvorgabe nach dem branched-testing für mehrere Untertests des **AID 3**. (Aus Kubinger & Holocher-Ertl, 2014, S. 40)

Da jede dieser Gruppen 5 Items umfasst, ergab dies in Summe 15 Items. Diese Items sowie alle Items der umliegenden Blöcke 4, 6, 11, 13, 16 und 18 (wobei es hier einige doppelt vorkommende gibt) wurden zunächst gemäß Testmanual nach Rangplätzen gereiht, sodass eine Liste von Items mit steigender Itemschwierigkeit die Folge war. Zusätzlich zu den 15 Kern-Items wurden 9 weitere Items umliegender Rangplätze für die geplante Modifikation ausgewählt. Diese verbliebenen 24 Items wurden, nach steigender Itemschwierigkeit, von 1 bis 24 nummeriert.

4.3.3.2 Change der Items

Der *Change*-Ansatz sieht eine Veränderung der Items zu Validierungszwecken vor, indem eine bestimmte Strategie verfolgt wird (Engelhardt et al., 2017). Jene Elemente im Item, welche konstruktrepräsentativ sind und somit jene kognitiven Prozesse hervorrufen, die zur Lösung der Aufgabe entscheidend sind, werden gezielt modifiziert.

Die dadurch neu entstandenen Item-Varianten (*Change*-Items) sollen einfacher oder schwieriger sein als ihr entsprechendes Original, jedoch nach wie vor dieselbe Fähigkeit erfassen.

Bei der Veränderung der 24 Original-Items wurde keiner Regel nachgegangen, sondern jedes Item individuell bearbeitet.

Grundsätzlich sind beim *Change*-Ansatz sowohl die Richtung, der Grad als auch die Anzahl der Modifikationen pro Item beliebig wählbar bzw. durch die Beschaffenheit des Originals gelenkt. Ist ein Item beispielsweise bereits sehr einfach, so wird es günstig sein, dieses Item schwieriger zu machen.

Für die vorliegende Validierungsstudie wurde aus zeitökonomischen Gründen (der Vorgabe) geplant, für jedes Original jeweils eine einfachere und eine schwierigere *Change*-Version zu entwickeln.

Dieses Vorhaben wurde für 22 der 24 Items realisiert.

Die Umsetzung erfolgte - durch Beteiligung und Feedback einer Arbeitsgruppe gestützt - durch die Autorin selbst.

Für Item 8 gibt es lediglich eine einfachere, keine schwierigere, Version; für Item 4 gibt es lediglich eine schwierigere, keine einfachere, Version.

Somit existieren für jedes Item (mit Ausnahme von Item 4 und 8) drei Varianten: Original (o), *Change*-einfacher (e) und *Change*-schwieriger (s).

Dies ergibt in Summe 70 Items.

Folgendes, in der Untersuchung verwendetes, Beispiel demonstriert die Umsetzung des *Change*-Ansatzes:

Original (Kubinger & Holocher-Ertl, 2014, S. 104): Nenne die Währungseinheit (das Geld) der USA?

Change-einfacher: In den USA kann man nicht mit Euro bezahlen. Wie heißt das Geld dort?

Change-schwieriger: Einen Euro kann man in Cent unterteilen. In welche Geldeinheiten kann ein Dollar unterteilt werden?

Sämtliche *Change*-Item-Varianten sowie Verweise zu den Original-Items des AID 3 (Kubinger & Holocher-Ertl, 2014) sind Anhang C zu entnehmen.

4.3.3.3 Testhefte

Wie in Abschnitt 4.1. Untersuchungsdesign beschrieben, erfolgte eine Aufteilung der insgesamt 70 Items auf drei Testhefte.

Dies hatte zur Folge, dass die Testhefte 1 und 2 insgesamt 46 Items enthalten, und Testheft 3 insgesamt 44 Items umfasst; siehe Tabelle 4.

Tabelle 4. *Verteilung der Items auf die drei Testhefte (TH)*

		Anzahl Items				gesamt	
Item-Nr.	1-12	13-24	1-12	13-24	1-12	13-24	
Variante	einfacher	einfacher	Original	Original	schwieriger	schwieriger	
TH 1	11		11	12		12	46
TH 2		12	11	12	11		46
TH 3	10	12			10	12	44

Bezüglich der Vorgabe-Reihenfolge der Items innerhalb eines Testhefts gilt, dass für jedes Itempaar, d.h. zwei Versionen desselben Items, individuell festgelegt wurde, welche Version an erster, und welche an zweiter Stelle im Testheft steht. Waren beispielsweise mögliche Lösungen für ein (einfacheres) Item in der Formulierung der schwierigeren Variante (schwieriger, da korrekte Antwortmöglichkeiten somit bereits vorweg genommen wurden) enthalten, so wurde die schwierigere Variante an zweiter Stelle dargeboten, um dem Merken von Lösungen vorzubeugen.

Im Gegensatz dazu wurde in einigen Fällen die schwierigere Frage zuerst gestellt, um zu verhindern, dass das Kind die Lösung (der einfacheren Frage) kennt, sich diese merkt, und dann bei der nachfolgenden, schwierigeren, Version reproduziert.

4.4 Erhebung demographischer Daten

An demographischen Variablen wurde für jede Testperson deren Alter, Geschlecht und Muttersprache erhoben.

Auf dem Protokollbogen (siehe Anhang D) wurde überdies vermerkt, welche Schulstufe und Schule das Kind besucht, und welches Testheft verwendet wurde.

4.5 Vorgehen

Nach Kontaktaufnahme mit dem Schulleiter eines Gymnasiums in Wien wurde für Anfang Juli 2017 ein persönlicher Gesprächstermin vereinbart (*Anmerkung:* Mit September 2017 erfolgte ein Wechsel der Schulleitung). Das formale Ansuchen, die Erhebung an dieser Schule durchführen zu dürfen, wurde im Schulgemeinschaftsausschuss am 10. Oktober 2017 bewilligt.

Auch an der zweiten teilnehmenden Schule, einer Wiener Mittelschule, wurde in einem telefonisch vereinbarten persönlichen Termin am 18. September 2017 um Bewilligung der wissenschaftlichen Erhebung angesucht. Diese wurde in der Sitzung des Schulforums am 21. September 2017 erteilt.

In einem nächsten Schritt wurden an beiden Schulen Informationsschreiben für die Lehrkräfte (siehe Anhang F) sowie Elternbriefe (siehe Anhang E) an die SchülerInnen der fünften und sechsten Schulstufen ausgegeben. Neben allen, die Erhebung betreffenden, relevanten Informationen enthielt der Elternbrief einen Abschnitt, um rückzumelden, ob die/der Erziehungsberechtigte der Teilnahme des Kindes an der Untersuchung zustimmt. Nur jene Kinder, deren Eltern ihr Einverständnis erklärt und unterschrieben hatten, nahmen an der Erhebung teil.

Auf dem Elternbrief des Gymnasiums gab es zusätzlich - auf Wunsch des zum

Zeitpunkt des Erstkontakts eingesetzten Direktors - die Möglichkeit, eine kurze Rückmeldung anzufordern.

Nach erfolgter Elternbrief-Laufzeit wurden die unterschriebenen Einverständniserklärungen von Lehrkräften eingesammelt und am Tag der Testung an die Testleiterin übergeben.

Die Erhebung selbst fand in der Wiener Mittelschule von 6.-9. November 2017 und im Gymnasium von 20.-24. November 2017 im Zeitfenster der 1.-6. Schulstunde statt.

Sie erfolgte für jedes Kind einzeln in einem ruhigen, von der Schule zur Verfügung gestellten, separaten Raum und wurde von der Verfasserin der Arbeit selbst geleitet.

Zu Beginn wurde dem Kind der Zweck der Testung geschildert sowie Geschlecht, Alter, Muttersprache und Schulstufe des Kindes erhoben. Zudem wurde auf die Freiwilligkeit der Teilnahme und Anonymität der Daten hingewiesen.

Jedes Kind wurde dahingehend instruiert, dass einfachere und schwierigere Fragen folgen. Die anfängliche Überlegung, darauf hinzuweisen, dass ähnliche Fragen vorkommen, wurde verworfen, da die Instruktionsphase sonst zu lange gedauert hätte. Stattdessen sollte ein Kind beim Ansprechen einer Ähnlichkeit (etwa: „Diese Frage hast du mir schon gestellt!“) für seine Aufmerksamkeit gelobt und darauf hingewiesen werden, dass manche Fragen ähnlich lauten, aber immer ein bisschen anders sind.

Im Verlauf der Testung selbst wurden jedem Kind mündlich 44 bzw. 46 Fragen gestellt. Die Testdauer schwankte zwischen 10-15 Minuten; Schulpausen wurden eingehalten.

Nach Beendigung wurde jedem Kind für seine Teilnahme gedankt, und es durfte sich eine kleine Süßigkeit aussuchen.

4.6 Informationen über die Art der Auswertung

4.6.1 Item-Response-Theorie

Die Item-Response-Theorie (IRT; item response theory) stellt eine Teildisziplin innerhalb der psychologischen Testtheorie dar.

Sie widmet sich der Frage, wie eine Reaktion auf ein Item (Aufgabe, Frage, Feststellung) eines psychologischen Tests oder Fragebogens zustande kommt, und bietet Modelle, um dieses Zustandekommen zu beschreiben (Kubinger, 2009; Kubinger, Rasch & Yanagida, 2011a).

Innerhalb der Item-Response-Theorie wird zwischen manifesten und latenten Variablen unterschieden.

Bei ersteren handelt es sich um das beobachtbare Antwortverhalten (beispielsweise: die Reaktion auf ein Item im AID 3), wohingegen es sich bei einer latenten Variable um ein Konstrukt (beispielsweise: Intelligenz) handelt, welches nicht direkt beobachtbar ist, aber das manifeste Verhalten bestimmt.

Der Zusammenhang zwischen diesen Variablen wird in der IRT beschrieben, wobei dem Testergebnis, als Indikator, mit bestimmter Wahrscheinlichkeit latente Merkmalsausprägungen zugeordnet werden. Das Antwortverhalten, bzw. die Wahrscheinlichkeit, ein Item zu lösen, ist demnach abhängig von der Ausprägung des latenten Merkmals in der Person.

Diese Ausprägung wird in IRT-Modellen als Personenparameter ξ bezeichnet.

Der Parameter, welcher das Item charakterisiert, ist der Itemschwierigkeitsparameter σ .

Beide stellen einen numerischen Kennwert dar.

Durch die Verwendung von Modellen, die innerhalb der Item-Response-Theorie verankert sind, ergibt sich eine Vielzahl an Vorzügen.

Sind Daten gewonnen, so lassen sich diese mittels Modelltests auf deren Geltung betreffend der im Rahmen der IRT-Modelle postulierten Annahmen überprüfen.

Dies betrifft unter anderem das testtheoretische Gütekriterium der Skalierung – Kubinger (2009, S. 82) versteht darunter, dass „die laut Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen adäquat abbilden“ –und die Dimensionalität der Skala (im Sinne von Itemhomogenität). Da die Zusammenfassung von Reaktionen auf verschiedene Items zu einem einzigen Testwert nur dann sinnvoll ist, wenn das Antwortverhalten auf ein einziges latentes Merkmal zurückzuführen ist, ist die Prüfung dieser Voraussetzung bei einer derartigen Praxis unumgänglich. Besteht hier Modellgeltung - messen also alle Items dasselbe Konstrukt - können empirisch geprüfte Angaben zur Konstruktvalidität gemacht werden.

Ist ein Test innerhalb der IRT verankert, ergibt sich als weiterer Vorteil die Möglichkeit des adaptiven Testens (Kubinger, 2009).

Im Gegensatz zur üblichen Testvorgabe, bei der alle Items nacheinander bearbeitet werden, werden beim adaptiven Testen nur jene Aufgaben vorgegeben, die dem Fähigkeitsniveau der Person entsprechen, und somit maximal informativ sind.

Das erste Item ist dabei ein mittelschweres, wobei mit zunehmender Testlänge eine immer genauere Schätzung der Fähigkeit der betreffenden Person vorgenommen werden kann. Um passende Items parat zu haben, bedarf es eines großen Pools davon.

Erfolgt die Itemvorgabe nicht in Bezug auf jedes einzelne Item maßgeschneidert (tailored testing, wie zuvor geschildert), sondern im Hinblick auf Itemgruppen, spricht man von branched testing. Diese Methode lässt sich in Paper-Pencil-Verfahren, wie dem AID 3 (Kubinger & Holocher-Ertl, 2014), umsetzen.

Innerhalb der IRT existieren verschiedene Testmodelle – sie unterscheiden sich unter anderem dahingehend, wie viele bzw. welche Parameter als auf das Antwortverhalten Einfluss nehmend inkludiert werden.

Auch Itemformate finden Berücksichtigung - so gibt es für Tests, deren Itemformat dichotom ist, dichotome Modelle.

4.6.2 Rasch-Modell

Das (dichotome logistische Test-) Modell von Rasch, kurz Rasch-Modell, gilt als Grundmodell der IRT und wurde von dem dänischen Statistiker Georg Rasch (1901-1980) entwickelt (Kubinger, 2018).

Kubinger et al. (2011a) folgend, beschreibt es die Wahrscheinlichkeit, dass eine Person ein Item eines Leistungstests löst, wobei nur zwei Ausprägungen für die Variable des manifesten, beobachtbaren Merkmals in Frage kommen: -/0 für „nicht gelöst“, +/1 für „gelöst“.

Welche dieser Ausprägungen zum Tragen kommt, ist abhängig von zwei latenten Merkmalen: der Fähigkeit der Person v , im Ausmaß ξ_v , und der Schwierigkeit des Items i , im Ausmaß σ_i .

Diese beiden Parameter - Personenparameter und Itemparameter - sind zunächst unbekannt, können jedoch auf Basis des Modells mittels Maximum-Likelihood-Verfahren computergestützt geschätzt werden.

Kubinger (2018) formuliert die (Rasch-)Modellannahmen folgendermaßen:

- (1) Das Ausmaß an interessierender Fähigkeit je Person ist durch einen einzigen Parameter zu charakterisieren; die Fähigkeit stellt also ein eindimensionales Merkmal dar.
- (2) Der Grad der Schwierigkeit je Aufgabe ist durch einen einzigen Parameter zu charakterisieren; die Schwierigkeit stellt also ein eindimensionales Merkmal dar.
- (3) Beide Parameter messen auf derselben Skala.
- (4) Die Leistungen (gelöst oder eben nicht gelöst) sind für jede Person über alle Aufgaben hinweg «lokal stochastisch unabhängig»; d.h., ob eine best. Person eine best. Aufgabe löst oder nicht löst, hängt – abgesehen vom Zufall – nur von ihrer Fähigkeit und der Schwierigkeit der Aufgabe ab, nicht aber davon, welche anderen Aufgaben sie bereits gelöst hat oder noch lösen wird.

Die Gleichung, welche die Wahrscheinlichkeit modelliert, dass Person v mit der Fähigkeit ξ_v ein Item i mit der Schwierigkeit σ_i löst, lautet:

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

Die Wahrscheinlichkeit, ein Item nicht zu lösen, resultiert als Gegenwahrscheinlichkeit.

Darüber hinaus wird aus der Formel ersichtlich, dass beide Parameter - konkret die Differenz $\xi_v - \sigma_i$ - für die Lösung des Items entscheidend sind.

Entspricht das Fähigkeitsausmaß einer Person der Anforderung des Items, so resultiert die Lösungswahrscheinlichkeit $P(x_{vi} = 1) = 0,50$ (Kubinger, 2018).

Für die vorliegende Arbeit ist die Überprüfung der gewonnenen Daten auf Rasch-Modell-Geltung insbesondere deshalb relevant, „weil dann einerseits eindimensional, wichtiger aber, mittels der „Anzahl gelöster Aufgaben“ gemessen bzw. verrechnet werden kann.“ (Kubinger et al., 2011a, S. 557).

Eindimensionalität bedeutet, dass der Test nur eine einzige latente Eigenschaft erfasst.

Die Überprüfung der Rasch-Modell-Geltung für einen entwickelten Test bzw. dessen Items wird mittels Modelltests umgesetzt.

Ermöglicht wird dies durch jene Eigenheit des Rasch-Modells, welche als Spezifische Objektivität bzw. Stichprobenunabhängigkeit bezeichnet wird.

Kubinger (2009, S. 89) spezifiziert dies folgendermaßen:

Der Unterschied in den Fähigkeiten ξ_v und ξ_w zwischen je zwei Personen v und w kann unabhängig davon bestimmt werden, welche Aufgaben des Tests dafür herangezogen werden; bzw. umgekehrt und wichtiger, der Vergleich je zweier Aufgaben i und j bezüglich σ_i und σ_j ist unabhängig davon möglich, welche Stichprobe dafür verwendet wird.

Hält spezifische Objektivität der Vergleiche nicht, gilt auch das Rasch-Modell nicht.

Unter den zur Verfügung stehenden Tests zur Überprüfung der Modellgeltung stellt Andersens (bedingter) Likelihood-Ratio-Test (1973) den etabliertesten dar. Kubinger (2018) erläutert, dass der hierfür erforderliche Stichprobenumfang extrem groß ist und daher trotz vernachlässigbarer Modellabweichung häufig zu einem signifikanten Modelltest führt.

Im Folgenden findet daher jene Alternative zu Andersens Likelihood-Ratio-Test Anwendung, die im kommenden Abschnitt näher beschrieben wird.

4.6.3 Dreifache Varianzanalyse als Rasch-Modell-Test

Kubinger et al. (2009, 2011b) fanden (als Nebenprodukt der Entwicklung eines Ansatzes zur Planung der Stichprobengröße beim Kalibrieren eines Leistungstests nach dem Rasch-Modell) eine alternative Methode, um Items einer Prüfung auf Rasch-Modell-Geltung zu unterziehen.

Dabei machten sie sich die Eigenheit der Spezifischen Objektivität des Rasch-Modells zunutze.

Umgelegt auf ein Varianzanalyse-Design bedeutet diese, „dass es keine Wechselwirkung zwischen bestimmten Personengruppen, z.B. in Bezug auf das Geschlecht, und bestimmten Aufgaben gibt.“ (Kubinger et al., 2011a, S. 331)

Auf dieser Basis kann eine dreifache Varianzanalyse mit gemischter Klassifikation als Rasch-Modell-Test in folgender Weise herangezogen werden:

Als externes Teilungskriterium für die Daten dient das Geschlecht der Testpersonen – dies ist Faktor A, mit den beiden Faktorstufen männlich und weiblich. Der Faktor ist fest und teilt die Daten in zwei Gruppen auf.

Die Testpersonen selbst gelten als Faktorstufen des zufälligen Faktors **B**¹, und sind in A eingenistet ($A \succ B$), da jede Testperson nur einer der Faktorstufen von A (entweder männlich oder weiblich) zugeordnet werden kann.

Den dritten Faktor stellen die Items dar – dieser Faktor C ist fest, da genau jene Items interessieren, die entwickelt wurden und nun überprüft werden sollen.

¹ fett gedruckt, da es sich um einen zufälligen Faktor handelt

(A $\times$ B) ist kreuzklassifiziert mit C.

Das Resultat ist eine dreifache Varianzanalyse der Variante (A $\times$ B) $\times$ C mit gemischter Klassifikation.

Obwohl es sich bei dem manifesten Merkmal Testleistung um eine Variable mit dichotomer Ausprägung handelt, was zur Verletzung der Normalverteilungsannahme führt, konnte von Kubinger et. al (2009) gezeigt werden, dass die dreifache Varianzanalyse hier angewendet werden kann.

Der Ansatz sieht vor, zu überprüfen, ob eine Wechselwirkung zwischen den Faktoren A und C, also zwischen Personengruppen und Aufgaben, vorliegt. Ist dies nicht der Fall, hält Spezifische Objektivität, und somit auch das Rasch-Modell.

Die wesentliche Voraussetzung, um den Ansatz nutzen zu können, ist jene, dass sich die Hauptwirkung A zuvor als nicht signifikant herausgestellt haben muss.

Kubinger et al. (2011a) erinnern daran, „dass nur ein einziger Modelltest durchgeführt wird, also nur in Bezug auf ein einziges Teilungskriterium, weil auch die Untersuchungsplanung bloß auf ein untersuchungsbezogenes Risiko 1. Art (von z.B. $\alpha = 0,05$) vorgenommen wurde.“ (S.562)

5 Ergebnisse

5.1 Rasch-Modell-Analysen

Zunächst werden die Ergebnisse der Prüfung auf Geltung des Rasch-Modells mittels dreifacher Varianzanalyse nach Kubinger et al. (2009, 2011b) für alle verwendeten Items (Originale sowie *Change*-Varianten) dargestellt.

Die Analyse erfolgte mittels IBM SPSS Statistics (24.0).

Im Anschluss werden die Itemleichtigkeitsparameter aller Items angeführt - geschätzt in der freien Statistik-Software-Umgebung R mittels Programmpaket eRm (extended Rasch modeling, 0.15-7) von Mair, Hatzinger und Maier (2016).

5.1.1 Ergebnisse der dreifachen Varianzanalyse

Der Rasch-Modelltest wurde in Bezug auf das Teilungskriterium Geschlecht durchgeführt, das Risiko 1. Art auf $\alpha = 0,05$ festgesetzt.

Das Ergebnis für die Hauptwirkung des Faktors Geschlecht lautet:

$F(1, 125) = 0,32, p = 0,57$.

Da es nicht signifikant ist, darf die Wechselwirkung in weiterer Folge geprüft werden.

Für die Interaktion der Faktoren Geschlecht x Items ergibt sich:

$F(69, 5462) = 1,19, p = 0,13$.

Auch hier liegt ein nicht signifikantes Ergebnis vor.

Die erwünschte Nullhypothese $H_{10}(A \times C)$ kann angenommen werden.

5.1.2 Schätzung der Itemleichtigkeitsparameter

Da im vorliegenden Kontext eine falsche Antwort mit 0 für *nicht gelöst* und eine richtige mit 1 für *gelöst* verrechnet wird, sind die Itemleichtigkeitsparameter wie folgt zu interpretieren:

Ein hoher Itemleichtigkeitsparameter signalisiert eine hohe Lösungshäufigkeit und somit eine niedrige Itemschwierigkeit.

Umgekehrt steht ein niedriger Itemleichtigkeitsparameter für eine geringe Lösungshäufigkeit und hohe Itemschwierigkeit.

In den Tabellen 5 bis 28 werden die jeweiligen Itemleichtigkeitsparameter der drei Versionen eines Items (*Change*-einfacher = e, Original = o, *Change*-schwieriger = s) angeführt, sowie der dazu gehörige Standardschätzfehler und das 95%-Konfidenzintervall. Dies ist der Bereich, in dem der Parameter, unter Berücksichtigung von Zufallsfehlern, mit einer Wahrscheinlichkeit von 95% liegt.

Tabelle 5. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 1 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
1e	2,127	0,305	[1,529; 2,725]
1o	1,877	0,294	[1,300; 2,454]
1s	-0,779	0,267	[-1,303; -0,255]

Wie in Tabelle 5 ersichtlich, resultiert für Item 1 eine Reihung der Itemschwierigkeiten gemäß der erwünschten Alternativhypothese H2₁:

Item 1e ist einfacher als das Original, welches wiederum einfacher ist als 1s.

Bei Betrachtung der Konfidenzintervalle wird deutlich, dass die Intervalle von Item 1e und 1o überlappen, die Itemschwierigkeiten sich also nicht signifikant voneinander unterscheiden.

Item 1s hingegen weist einen signifikanten Unterschied betreffend Schwierigkeit zu 1o auf.

Tabelle 6. Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 2 (95% KI)

Item	Schätzwert	Std.Schätzfehler	95% KI
2e	2,127 ²	0,305	[1,529; 2,725]
2o	2,822	0,368	[2,101; 3,543]
2s	-1,691	0,317	[-2,312; -1,069]

Item 2e erweist sich, entgegen der Intention beim Generieren des Items, als schwieriger als sein entsprechendes Original, wobei die Konfidenzintervalle überlappen.

Die Itemschwierigkeit von 2s unterscheidet sich signifikant von jener des Items 2o, und zwar in die erwünschte Richtung – es ist signifikant schwieriger als das Original.

Tabelle 7. Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 3 (95% KI)

Item	Schätzwert	Std.Schätzfehler	95% KI
3e	4,177	0,598	[3,005; 5,348]
3o	-1,185	0,283	[-1,739; -0,630]
3s	-2,017	0,344	[-2,691; -1,343]

Tabelle 7 lässt sich entnehmen, dass Item 3e signifikant einfacher ist als 3o.

² Werte, die entgegengesetzt der geplanten Richtung ausgefallen sind, sind grau hervorgehoben.

Die schwierigere Variante zeigt ebenso die erwünschte Tendenz, schwieriger zu sein als das Original, jedoch kann aufgrund überlappender Konfidenzintervalle nicht von einem signifikanten Unterschied gesprochen werden.

Tabelle 8. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 4 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
4o	-0,041	0,354	[-0,734; 0,652]
4s	-0,550	0,356	[-1,264; 0,165]

Item 4s ist in Einklang mit der Alternativhypothese H_{21} schwieriger als sein Gegenstück, der Unterschied ist jedoch unter Einbezug der Konfidenzintervalle nicht signifikant.

Tabelle 9. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 5 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
5e	2,773	0,360	[2,067; 3,479]
5o	2,247	0,318	[1,724; 2,770]
5s	0,551	0,250	[0,062; 1,041]

Wie in Tabelle 9 ersichtlich, resultiert für Item 5 eine Reihung der Itemschwierigkeiten gemäß der erwünschten Richtung:

Item 5e ist einfacher als das Original, welches wiederum einfacher ist als 5s.

Bei Betrachtung der Konfidenzintervalle wird deutlich, dass die Intervalle von Item 5e und 5o überlappen, die Itemschwierigkeiten sich also nicht signifikant voneinander unterscheiden.

Item 5s hingegen weist einen signifikanten Unterschied betreffend Schwierigkeit zu 5o auf.

Tabelle 10. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 6 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
6e	2,320	0,319	[1,694; 2,945]
6o	2,055	0,305	[1,458; 2,653]
6s	-0,310	0,254	[-0,808; 0,187]

Item 6e erweist sich als einfacher als sein entsprechendes Original, wobei die Konfidenzintervalle überlappen.

Die Itemschwierigkeit von 6s unterscheidet sich signifikant von jener des Items 6o, und zwar in die erwünschte Richtung – es ist signifikant schwieriger als das Original.

Tabelle 11. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 7 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
7e	1,409	0,268	[0,884; 1,935]
7o	1,119	0,263	[0,604; 1,635]
7s	-1,408	0,298	[-1,992; -0,823]

Tabelle 11 lässt sich entnehmen, dass für Item 7 eine Reihung der Itemschwierigkeiten gemäß der erwünschten Alternativhypothese H2₁ resultiert: Item 7e ist einfacher als das Original, welches wiederum einfacher ist als 7s.

Bei Betrachtung der Konfidenzintervalle wird deutlich, dass die Intervalle von Item 7e und 7o überlappen, die Itemschwierigkeiten sich also nicht signifikant voneinander unterscheiden.

Item 7s hingegen weist einen signifikanten Unterschied betreffend Schwierigkeit zu 7o auf.

Tabelle 12. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 8 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
8e	1,705	0,397	[0,926; 2,484]
8o	1,865	0,408	[1,065; 2,665]

Item 8e erweist sich, entgegen der Intention beim Generieren des Items, als schwieriger als sein entsprechendes Original, wobei die Konfidenzintervalle überlappen.

Tabelle 13. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 9 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
9e	0,814	0,252	[0,319; 1,308]
9o	0,852	0,257	[0,349; 1,355]
9s	0,246	0,248	[-0,241; 0,733]

Tabelle 13 kann entnommen werden, dass auch bei diesem Item eine Richtungsumkehr statt fand – Item 9e ist schwieriger als das Original, wenngleich nicht signifikant schwieriger. Für 9s ist die Tendenz wie erwünscht - es ist schwieriger als das Original, doch auch hier überlappen die Konfidenzintervalle.

Tabelle 14. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 10 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
10e	-0,107	0,251	[-0,600; 0,385]
10o	0,160	0,250	[-0,330; 0,650]
10s	-1,408	0,298	[-1,992; -0,823]

Bei Item 10 ergibt sich für das (intendiert) einfachere Item eine höhere Itemschwierigkeit als für das Original - bei überlappenden Konfidenzintervallen. Item 10s ist, wie erwünscht, schwieriger als 10o. Da die Konfidenzintervalle nicht überlappen, kann auf einen signifikanten Unterschied der Itemschwierigkeiten geschlossen werden.

Tabelle 15. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 11 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
11e	0,865	0,254	[0,366; 1,363]
11o	1,877	0,294	[1,300; 2,454]
11s	0,864	0,255	[0,365; 1,363]

Aufgrund einander lediglich berührender Konfidenzintervalle zeigt sich Item 11e als signifikant schwieriger als Item 11o, was der Alternativhypothese H2₁ widerspricht. Es ist signifikant schwieriger als das Original. Item 11s ist ebenso signifikant schwieriger als 11o, wobei es in etwa dieselbe Schwierigkeit aufweist wie das *Change-e*-Item.

Tabelle 16. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 12 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
12e	1,866	0,289	[1,300; 2,433]
12o	0,596	0,252	[0,101; 1,091]
12s	-3,732	0,606	[-4,920; -2,544]

Wie in Tabelle 16 ersichtlich, sind alle Modifikationen von Item 12 gelungen. Die Reihung der Itemschwierigkeiten ist wie erwünscht, bei nicht überlappenden Konfidenzintervallen. Item 12e ist demnach signifikant einfacher als das Original, welches signifikant einfacher ist als 12s.

Tabelle 17. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 13 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
13e	-0,440	0,257	[-0,943; 0,064]
13o	-0,672	0,262	[-1,186; -0,158]
13s	-1,491	0,305	[-2,089; -0,893]

Die Itemschwierigkeiten der Versionen von Item 13 weisen durchwegs überlappende Konfidenzintervalle auf, tendieren jedoch insofern in die erwünschte Richtung, als 13e tendenziell einfacher ist als 13o und 13s tendenziell schwieriger als 13o.

Tabelle 18. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 14 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
14e	2,358	0,330	[1,711; 3,005]
14o	-1,265	0,287	[-1,828; -0,703]
14s	-2,137	0,357	[-2,837; -1,437]

Tabelle 18 lässt sich entnehmen, dass Item 14e signifikant einfacher ist als 14o. Die schwierigere Variante zeigt ebenso die erwünschte Tendenz, schwieriger zu sein als das Original, jedoch kann aufgrund überlappender Konfidenzintervalle nicht von einem signifikanten Unterschied gesprochen werden.

Tabelle 19. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 15 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
15e	-1,592	0,310	[-2,200; -0,984]
15o	-2,049	0,343	[-2,722; -1,376]
15s	-2,931	0,456	[-3,824; -2,038]

Die Itemschwierigkeiten der Versionen von Item 15 weisen durchwegs überlappende Konfidenzintervalle auf, tendieren jedoch insofern in die erwünschte Richtung, als 15e einfacher ist als 15o und 15s schwieriger ist als 15o.

Tabelle 20. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 16 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
16e	0,613	0,250	[0,122; 1,104]
16o	-1,185	0,283	[-1,739; -0,630]
16s	-1,898	0,336	[-2,556; -1,240]

Tabelle 20 lässt sich entnehmen, dass Item 16e signifikant einfacher ist als 16o. Die schwierigere Variante zeigt ebenso die erwünschte Tendenz, schwieriger zu sein als das Original, jedoch kann aufgrund überlappender Konfidenzintervalle nicht von einem signifikanten Unterschied gesprochen werden.

Tabelle 21. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 17 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
17e	-0,247	0,253	[-0,743; 0,249]
17o	0,059	0,248	[-0,427; 0,546]
17s	-2,553	0,406	[-3,348; -1,757]

Bei Item 17 ergibt sich für das (intendiert) einfachere Item eine höhere Itemschwierigkeit als für das Original - bei überlappenden Konfidenzintervallen. Item 17s ist, wie erwünscht, signifikant schwieriger als 17o.

Tabelle 22. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 18 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
18e	-0,440	0,257	[-0,943; 0,064]
18o	-0,342	0,254	[-0,840; 0,157]
18s	-0,913	0,274	[-1,450; -0,377]

Wie in Tabelle 22 ersichtlich, haben alle Versionen von Item 18 überlappende Konfidenzintervalle bezüglich ihrer Schwierigkeit. Entgegen der eigentlichen Intention zeigt Item 18e die Tendenz, schwieriger zu sein als das Original. Für 18s ist die Tendenz wie erwünscht.

Tabelle 23. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 19 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
19e	1,794	0,290	[1,225; 2,363]
19o	-2,296	0,369	[-3,020; -1,572]
19s	-2,413	0,386	[-3,170; -1,655]

Tabelle 23 lässt sich entnehmen, dass Item 19e signifikant einfacher ist als 19o. Die s-Variante zeigt ebenso die erwünschte Tendenz, schwieriger zu sein als das Original, jedoch kann aufgrund überlappender Konfidenzintervalle nicht von einem signifikanten Unterschied gesprochen werden.

Tabelle 24. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 20 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
20e	-0,851	0,270	[-1,380; -0,321]
20o	-0,537	0,259	[-1,044; -0,030]
20s ³	-	-	-

Item 20e erweist sich, entgegen der Intention beim Generieren des Items, als schwieriger als sein entsprechendes Original, wobei die Konfidenzintervalle überlappen.

Tabelle 25. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 21 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
21e	3,355	0,448	[2,477; 4,233]
21o	-0,955	0,272	[-1,489; -0,421]
21s	-1,788	0,327	[-2,429; -1,148]

Wie in Tabelle 25 ersichtlich, ist Item 21e signifikant einfacher als 21o.

Die schwierigere Variante zeigt ebenso die erwünschte Tendenz, schwieriger zu sein als das Original, wobei die Konfidenzintervalle überlappen, sodass nicht von einem signifikanten Unterschied gesprochen werden kann.

³ Da keine der Testpersonen das Item löste, konnte für Item 20s kein Itemleichtigkeitsparameter geschätzt werden.

Tabelle 26. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 22 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
22e	1,966	0,301	[1,377; 2,556]
22o	-0,406	0,256	[-0,907; 0,095]
22s	-0,560	0,261	[-1,071; -0,048]

Item 22e ist, entsprechend der gewünschten Alternativhypothese H_{21} , signifikant einfacher als das Original. Item 22o tendiert dazu, einfacher zu sein als 22s, wobei der Unterschied nicht signifikant ist.

Tabelle 27. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 23 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
23e	-0,616	0,291	[-1,185; -0,046]
23o	-0,672	0,262	[-1,186; -0,158]
23s	-1,898	0,336	[-2,556; -1,240]

Tabelle 27 kann entnommen werden, dass Item 23e und 23o annähernd gleiche Itemschwierigkeiten aufweisen. Item 23s hingegen ist bezüglich seiner Itemschwierigkeit signifikant vom Original verschieden, und zwar in die intendierte Richtung.

Tabelle 28. *Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 24 (95% KI)*

Item	Schätzwert	Std.Schätzfehler	95% KI
24e	-0,060	0,250	[-0,550; 0,431]
24o	-1,030	0,275	[-1,570; -0,490]
24s ⁴	-	-	-

Item 24e ist in Einklang mit der Alternativhypothese H_{21} einfacher als sein Original, der Unterschied ist jedoch unter Einbezug der Konfidenzintervalle nicht signifikant.

⁴ Da keine der Testpersonen das Item löste, konnte für Item 24s kein Itemleichtigkeitsparameter geschätzt werden.

6 Diskussion

Im Hinblick auf das angestrebte Untersuchungsziel, eine Validierung des Untertests „Alltagswissen“ des AID 3 (Kubinger & Holocher-Ertl, 2014) mittels *Change*-Ansatz vorzunehmen, wurden die Effekte der systematischen Manipulation von Aufgabeneigenschaften auf die Itemschwierigkeit untersucht. Wesentlich hierbei ist, dass nur dann von einer gelungenen Validierung gesprochen werden kann, wenn ein Unterschied in den Itemschwierigkeiten in intendierter Richtung resultiert und die Items nach wie vor dasselbe Konstrukt messen.

Um letzteres zu überprüfen, wurden die Items einer dreifachen Varianzanalyse als Rasch-Modelltest unterzogen, die ergab, dass keine Wechselwirkung zwischen Geschlecht und Items in Bezug auf die Testwerte im „Alltagswissen“ vorliegt.

Umgelegt auf das Rasch-Modell bedeutet dies, dass es für die gegebenen Daten gilt.

Die Ergebnisse sprechen dafür, dass die *Change*-Items dasselbe Konstrukt erfassen wie die Originale.

Die Untersuchung der Auswirkung auf die Itemschwierigkeiten durch den *Change*-Ansatz wurde vorgenommen, indem in einem ersten Schritt die Itemleichtigkeitsparameter aller verwendeten Items geschätzt wurden.

In weiterer Folge wurden für die drei Versionen eines Items die Itemschwierigkeiten der *Change*-Items mit jenen der Originale verglichen.

Es zeigte sich, dass die *Change*-schwieriger-Transformationen erfolgreicher waren als die *Change*-einfacher-Transformationen; erfolgreich insofern, als es bei den *Change*-schwieriger-Items kein einziges Item gab, das sich nicht hypothesenkonform (H_{21} : *Change*-s-Items sind schwieriger als Original-Items) verhielt.

Alle *Change*-schwieriger-Items (23 Items) zeigten die Tendenz, schwieriger zu sein, als die jeweiligen Originale.

Davon waren zehn Items⁵ hinsichtlich ihrer Itemschwierigkeit signifikant verschieden vom Original. Zwei Items (20s, 24s) wurden von keiner Testperson gelöst.

Elf Items⁶ tendierten in die richtige Richtung, jedoch kann aufgrund überlappender Konfidenzintervalle nicht von einem signifikanten Unterschied in der Itemschwierigkeit gesprochen werden.

Bei den *Change*-einfacher-Transformationen (23 Items) zeigte sich ein anderes Bild.

Hier musste für neun Items die Alternativhypothese (H2₁: *Change*-e-Items sind einfacher als Original-Items) verworfen werden.

Sieben⁷ dieser Items waren tendenziell schwieriger als das Original, eines (11e) wies hinsichtlich seiner Itemschwierigkeit sogar einen signifikanten Unterschied zu seinem Original auf und ein Item (23e) besaß annähernd dieselbe Schwierigkeit wie das dazu gehörige o-Item.

Insgesamt sieben Items⁸ zeigten hinsichtlich ihrer Tendenz die richtige Richtung – sie waren einfacher als die Originale, wobei der Unterschied nicht signifikant ausfiel.

Dies war nur bei sieben weiteren Items⁹ der Fall: Diese Items erwiesen sich hinsichtlich ihrer Schwierigkeit als signifikant verschieden vom Original, und zwar in die erwünschte Richtung.

Auffallend ist, dass jene *Change*-Varianten, die sich, hypothesenkonform und signifikant, von ihrem Original unterscheiden, unter den e-Items größtenteils bei höheren Item-Nummern⁹ zu finden sind (also bei jenen Items, wo das Original bereits eine höhere Schwierigkeit aufwies), wohingegen es bei den gelungenen *Change*-schwieriger-Versionen⁵ eher umgekehrt der Fall ist.

Zurückzuführen ist dies darauf, dass der *Change*-Ansatz prinzipiell freie Richtungsgestaltung zulässt, es aber am günstigsten ist, die Schwierigkeit des Originals einzukalkulieren: Es ist weitaus sinnvoller und viel versprechender im

⁵ Items: 1s, 2s, 5s, 6s, 7s, 10s, 11s, 12s, 17s, 23s

⁶ Items: 3s, 4s, 9s, 13s, 14s, 15s, 16s, 18s, 19s, 21s, 22s

⁷ Items: 2e, 8e, 9e, 10e, 17e, 18e, 20e

⁸ Items: 1e, 5e, 6e, 7e, 13e, 15e, 24e

⁹ Items: 3e, 12e, 14e, 16e, 19e, 21e, 22e

Sinne der Zielerreichung, Items, die bereits sehr einfach sind, mit dem *Change*-Ansatz schwieriger zu machen, und umgekehrt.

Diesen Aspekt betonen auch Engelhardt et al. in ihrer Vorstellung des *Change*-Ansatzes (2017).

Dennoch erlaubt diese Methode der Validierung ein flexibles Vorgehen, welches auf jeden Item-Typ umgelegt werden kann.

Im Folgenden werden mögliche Erklärungen für nicht hypothesenkonforme Ergebnisse gegeben, sodass – bei Anwendung des *Change*-Ansatzes auf ähnliches Item-Material – bestimmte Fehler zukünftig vermieden werden können.

Zunächst sei darauf hingewiesen, dass aufgrund des Designs der vorliegenden Studie erforderlich war, jeder Testperson zwei Varianten eines Items vorzugeben. Wenngleich die Kinder - bei Nachfrage - darauf hingewiesen wurden, dass jedes Item unterschiedlich sei, machte sich bei einigen Kindern eine gewisse Irritation bemerkbar, dass etwas Ähnliches schon einmal gefragt worden war.

Dies war insbesondere der Fall bei Item 2 und Item 9 (alle Items: siehe Anhang C).

Die Vorgangsweise könnte dazu geführt haben, dass das Kind seine Antwort änderte, also etwas Anderes sagte als bei der zuerst geführten Version – möglicherweise, weil es der Meinung war, die erste Antwort sei falsch gewesen, weshalb die Frage noch einmal gestellt werde.

Die Entscheidung für ein derartiges Design fiel überlegt, mit der Überzeugung, dass die Testpersonen es verkraften, wenn ihnen Varianten doppelt dargeboten werden. Im Nachhinein ist jedoch, insbesondere beim Vorliegen ähnlicher Varianten, zu einem anderen Design zu raten.

Für Item 10 und 18 liegt die Vermutung nahe, dass die Idee, die Items durch eine Zusatzinformation einfacher zu machen, zu einem gegenteiligen Effekt geführt hat, da deren Länge dadurch erhöht worden war. Die Kapazität des Arbeitsgedächtnisses könnte in höherem Ausmaß gefordert worden sein, wodurch das Item nicht einfacher, sondern schwieriger wurde.

Für die fünf verbliebenen Items mit Richtungsumkehr (die Items sollten einfacher werden als die Originale, erwiesen sich aber als schwieriger/gleich schwierig) können keine pauschalen Fehler betreffend Konstruktion benannt werden.

Rückblickend liegt jedenfalls nahe, dass sie tatsächlich schwieriger waren als die entsprechenden Originale. Die Gründe dafür werden nachfolgend einzeln besprochen.

Bei Item 8 kommt im Ausgangs-Statement des Originals das Wort „Affe“ vor. Ausgehend davon soll die Testperson spezifischeres Wissen dazu abrufen.

In der *Change*-einfacher-Variante hingegen taucht der Begriff „Affe“ gar nicht auf – die Testperson muss selbst darauf kommen.

Für die *Change*-schwieriger Variante war bei Item 11 als Lösung „1 Jahr“ gesucht. Die Frage nach einem Zeitraum, der noch dazu „1“ beträgt, ist rückblickend leichter als die einfachere Variante, bei der keine Zahl gesucht war, sondern die Testperson sich ausgewählter artikulieren musste.

Auch bei Item 17 wurde das Original häufiger gelöst als jene Version, die einfacher sein sollte. Da die e-Variante zuerst gestellt wurde, zeigte sich, dass etliche Kinder die e-Frage „Was ist ein Triathlon?“ nicht beantworten konnten.

Wenn man ihnen danach die o-Variante, in welcher der Begriff vorkam, vorlas, hatten sie sogar spezifischeres Wissen darüber. Das lässt den Schluss zu, dass der Begriff „Triathlon“ bei etlichen Testpersonen nur im passiven Wortschatz vorhanden war, was die e-Variante tatsächlich schwieriger machte.

Bei Item 20 stellte die in der *Change*-einfacher-Variante (siehe Anhang C) gegebene Zusatzinformation für die Kinder keine Hilfe dar, sondern war - im Gegenteil - sogar irreführend, da viele Kinder „Metzger“ antworteten.

Darüber hinaus könnte das e-Item auch zu lang gewesen sein.

Auf die Frage nach der Religion des Papstes in Item 23e antworteten auffällig viele Kinder „Judentum“. Schließlich verbalisierte ein Kind die von der Verfasserin

vermutete Ursache dafür, indem es sagte: „Der muss Jude sein, weil er so ein Kapperl¹⁰ auf hat.“ Auch für diese Frage zeigte sich demnach, dass sie tatsächlich schwieriger war als das Original.

Konnte für gewisse Items kein hypothesenkonformes Ergebnis erzielt werden, so ist dies, wie eben dargelegt, auf erklärbare Konstruktionsfehler zurückzuführen.

Als Hinweis für zukünftige AnwenderInnen kann demnach gelten, die Konstruktion des *Change*-Items immer auf die Beschaffenheit des Originals abzustimmen. An der Erstellung einer bestimmten (beispielsweise einfacheren) Variante festzuhalten, kann auch zu unerwünschten, gegenteiligen Ergebnissen führen, wenn sich das Original nicht gut dafür eignet. Auch zu lange Items sollten vermieden werden.

Bezüglich der gewonnenen Itemparameterschätzungen ist festzuhalten, dass die Stichprobe mit 127 Testpersonen zu klein sein könnte, um auf ihrer Basis genaue Schätzwerte zu erlangen. Dies wurde unter Punkt 4.1 angesprochen.

Ein größerer Stichprobenumfang war nicht möglich, da nur zwei Schulen ihre Bereitschaft zur Teilnahme erklärten.

Dennoch konnte als Validierungsschritt gezeigt werden, dass die als schwierigkeitsbestimmend erachteten Inhalte und Anforderungen für die Lösung der Items bedeutsam waren. Darüber hinaus wurde für die gewonnenen Daten Rasch-Modell-Geltung nachgewiesen.

Die erstmalige Anwendung des *Change*-Ansatzes als Validierungsmethode für Items des AID 3 (Kubinger & Holocher-Ertl, 2014) wird demnach prinzipiell als gelungen angesehen.

¹⁰ im Fachjargon: Pileolus

7 Verzeichnisse

7.1 Literaturverzeichnis

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Andersen, E. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, S. 123-140. doi: 10.1007/BF02291180

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111 (4), S. 1061-1071. doi: 10.1037/0033-295X.111.4.1061

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52 (4), S. 281-302. doi: 10.1037/h0040957

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological bulletin*, 93 (1), S. 179-197. doi: 10.1037/0033-2909.93.1.179

Embretson, S. E. (1998). A cognitive design system approach for generating valid tests: Application to abstract reasoning. *Psychological Methods*, 3 (3), S. 380-396. doi: 10.1037/1082-989X.3.3.380

- Engelhardt, L., Goldhammer, F., Naumann, J., & Frey, A. (2017). Experimental validation strategies for heterogenous computer-based assessment items. *Computers in Human Behavior*, 76, S. 683-692. doi: 10.1016/j.chb.2017.02.020
- Engelhardt, L., Goldhammer, F., Naumann, J., Frey, A., Hartig, K., Horz, H., Kuchta, K., Wenzel, S.F.C. (2015, März). *Konstruktvalidierung durch die Manipulation der Aufgabenschwierigkeit in simulationsbasierten ICT-Aufgaben*. Vortrag auf der 3. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Bochum, Deutschland.
- Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger, & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage). Heidelberg: Springer.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, S. 1-73. doi: 10.1111/jedm.12000
- Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2., überarbeitete und erweiterte Auflage). Göttingen: Hogrefe.
- Kubinger, K. D. (2018). *Rasch-Modell*. In M. A. Wirtz (Hrsg.), *Dorsch – Lexikon der Psychologie*. Abgerufen am 22. 02. 2018. URL: <http://portal.hogrefe.com/dorsch/de/startseite/stichwort-detailseite/desktop/1/keyword/rasch-modell/>.
- Kubinger, K. D., & Holocher-Ertl, S. (2014). *Adaptives Intelligenz Diagnostikum - Version 3.1 (AID 3)*. Göttingen: Hogrefe.
- Kubinger, K. D., Rasch, D., & Yanagida, T. (2009). On designing data-sampling for Rasch model calibrating an achievement test. *Psychology Science Quarterly*, 51, S. 370-384. doi: 10.1.1.469.926

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

Kubinger, K. D., Rasch, D., & Yanagida, T. (2011b). A new approach for testing the Rasch model. *Educational Research and Evaluation*, 17 (5), S. 321-333. doi: 10.1080/13803611.2011.630529

Mair, P., Hatzinger, R., & Maier, M. J. (2016). eRm: Extended Rasch Modeling. 0.15-7. <http://erm.r-forge.r-project.org/>

Newton, P. E., & Shaw, S. D. (2013). Standards for talking and thinking about validity. *Psychological Methods*, 18 (3), S. 301-319. doi: 10.1037/a0032969

7.2 Abbildungsverzeichnis

Abbildung 1: Das Prinzip der Testvorgabe nach dem branched-testing für mehrere Untertests des AID 3.....	16
--	----

7.3 Tabellenverzeichnis

Tabelle 1: Altersverteilung der Gesamtstichprobe in Jahren.....	12
Tabelle 2: Verteilung der Muttersprache innerhalb der Gesamtstichprobe.....	12
Tabelle 3: Verteilung der Testhefte innerhalb der Gesamtstichprobe.....	13
Tabelle 4: Verteilung der Items auf die drei Testhefte (TH).....	18
Tabelle 5: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 1 (95% KI).....	28
Tabelle 6: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 2 (95% KI).....	29
Tabelle 7: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 3 (95% KI).....	29
Tabelle 8: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 4 (95% KI).....	30
Tabelle 9: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 5 (95% KI).....	30
Tabelle 10: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 6 (95% KI).....	31
Tabelle 11: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 7 (95% KI).....	31
Tabelle 12: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 8 (95% KI).....	32

Tabelle 13: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 9 (95% KI).....	32
Tabelle 14: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 10 (95% KI).....	33
Tabelle 15: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 11 (95% KI).....	33
Tabelle 16: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 12 (95% KI).....	34
Tabelle 17: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 13 (95% KI).....	34
Tabelle 18: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 14 (95% KI).....	35
Tabelle 19: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 15 (95% KI).....	35
Tabelle 20: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 16 (95% KI).....	36
Tabelle 21: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 17 (95% KI).....	36
Tabelle 22: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 18 (95% KI).....	37
Tabelle 23: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 19 (95% KI).....	37

Tabelle 24: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 20 (95% KI).....	38
--	----

Tabelle 25: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 21 (95% KI).....	38
--	----

Tabelle 26: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 22 (95% KI).....	39
--	----

Tabelle 27: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 23 (95% KI).....	39
--	----

Tabelle 28: Schätzungen der Itemleichtigkeitsparameter aller Varianten von Item 24 (95% KI).....	40
--	----

Anhang

A Abstract – deutsche Version

In der vorliegenden Masterarbeit wird die Strategie des *Change*-Ansatzes genutzt, um eine Validierung der Testwertinterpretationen des Untertests „Alltagswissen“ der Intelligenz-Testbatterie *Adaptives Intelligenz Diagnostikum 3* (AID 3, Version 3.1; Kubinger & Holocher-Ertl, 2014) vorzunehmen.

Der *Change*-Ansatz – nach einer Studie von Engelhardt, Goldhammer, Naumann und Frey (2017) – stellt eine experimentelle Validierungs-Methode dar, die sich besonders für jene Verfahren eignet, die komplexe Fähigkeiten anhand heterogener Item-Sets erfassen.

Er sieht vor, die Effekte der systematischen Manipulation von Aufgabeneigenschaften – mittels Kreation neuer Varianter existenter Items – auf die Itemschwierigkeit zu untersuchen.

Für die vorliegende Studie wurden 24 Original-Items des Untertests „Alltagswissen“ ausgewählt und von jedem Original (mit Ausnahme zweier Aufgaben) eine einfachere und eine schwierigere Version erstellt, sodass in Summe 70 Items resultierten.

Diese Items wurden in Form von Testheften 127 Schülerinnen und Schülern der fünften und sechsten Schulstufe (zwischen 10 und 14 Jahren) zur Bearbeitung vorgegeben.

Im Ergebnis zeigt sich, dass durch die *Change*-Items nach wie vor dasselbe Konstrukt erfasst wird wie durch die Originale, was die Basis für die angestrebte Validierung darstellt.

Darüber hinaus hatte die Veränderung der Aufgabeneigenschaften überwiegend den gewünschten Effekt betreffend Itemschwierigkeit: *Change*-einfacher-Items tendierten mehrheitlich dazu, einfacher zu sein als die entsprechenden Originale. Die erwünschte Tendenz, schwieriger zu werden, zeigte sich auch für alle *Change*-schwieriger-Modifikationen.

Für neun Items mit gegensätzlicher Entwicklung (anders als intendiert) werden mögliche Erklärungen gegeben.

Insgesamt wird der Versuch, den *Change*-Ansatz erstmalig als Validierungsmethode für AID 3- Items einzusetzen, als gelungen angesehen.

B Abstract – english version

The present master's thesis aims to validate the interpretations based on the test score of the subtest „Everyday Knowledge“ of the intelligence test battery *Adaptives Intelligenz Diagnostikum 3* (AID 3, Version 3.1; Kubinger & Holocher-Ertl, 2014) via the *change*-approach.

Proposed in 2017 by Engelhardt, Goldhammer, Naumann and Frey, the *change*-approach represents an experimental validation method, which is particularly useful for procedures that measure complex skills assessed by heterogeneous item sets.

The idea behind the *change*-approach is to investigate the effects of systematic manipulation of task characteristics on item difficulty by creating new variants of existing items.

In the course of the present study, 24 original „Everyday Knowledge“ items were selected to undergo the *change* process. An easier and a harder version were created for each of these items (except two), resulting in a total of 70 items.

After being assigned to test booklets, these items were administered to 127 students in fifth and sixth grade, aged 10 to 14 years.

The results show that original and *change* items assess the same trait, which is fundamental to validation.

Furthermore, for most of the items, changing task characteristics resulted in the desired effect regarding item difficulty: The majority of *change*-easier-items tended to be easier, and all of the *change*-harder-items tended to be harder than the originals. Explanations are offered for nine items with reversed (different than intended) results.

Overall, the attempt of applying the *change-approach* as a validation tool for AID 3-items for the first time is regarded as successful.

C *Change-Items*

Anhang C beinhaltet alle in der Untersuchung verwendeten Items.

Die Verweise zu den Original-Items des AID 3 beziehen sich auf die Testanweisung (Kubinger & Holocher-Ertl, 2014).

Item 1

Original: Aufgabengruppe 11, Item 3

Change-einfacher: In den USA kann man nicht mit Euro bezahlen. Wie heißt das Geld dort? (*Dollar*)

Change-schwieriger: Einen Euro kann man in Cent unterteilen. In welche Geldeinheiten kann ein Dollar unterteilt werden? (*Dime; Cent; Mill*)

Item 2

Original: Aufgabengruppe 5, Item 1

Change-einfacher: Vor ca. 70 Jahren war der letzte Weltkrieg; wie viele Weltkriege hat es gegeben? (*zwei*)

Change-schwieriger: Im 1. und 2. Weltkrieg kämpften viele Länder. Nenne ein europäisches Land, das in beiden Kriegen offiziell neutral blieb. (*Dänemark, Liechtenstein, Spanien, Schweden, Schweiz*)

Item 3

Original: Aufgabengruppe 6, Item 1

Change-einfacher: Der amerikanische Kontinent besteht aus Nordamerika und ...? (*Südamerika*)

Change-schwieriger: Brasilien erstreckt sich bis in den Süden von Südamerika. Nenne einen weiteren südamerikanischen Staat, der im Süden liegt. (*Chile; Argentinien; Uruguay*)

Item 4

Original: Aufgabengruppe 5, Item 2

Change-schwieriger: Sonne, Wind und Wasser sind nur einige von vielen möglichen Energiequellen. Nenne eine weitere. (*Erdgas (Propan, Methan); Kohle; Torf; Holz; Wellen- und Gezeitenkraft; Meeresströmung; Osmosekraft; Erdwärme (Geothermie); Biomasse; Müll; Kern-/Atomkraft; Deponie-/Klär-/Biogas*)

Item 5

Original: Aufgabengruppe 4, Item 5

Change-einfacher: Wie heißt der Kontinent, auf dem du und ich leben? (*Europa*)

Change-schwieriger: Welcher Kontinent liegt östlich von Europa? (*Asien*)

Item 6

Original: Aufgabengruppe 16, Item 3

Change-einfacher: Nenne ein Nachbarland von Österreich. (*Deutschland; Schweiz; Tschechien; Slowakei; Slowenien; Italien; Liechtenstein; Ungarn*)

Change-schwieriger: Nenne einen Staat, der im Osten an Österreich grenzt. (*Slowakei; Ungarn*)

Item 7

Original: Aufgabengruppe 12, Item 1

Change-einfacher: Was kann man tun, um nicht schwanger zu werden? (*enthaltssam sein; verhüten*)

Change-schwieriger: Welches Verhütungsmittel ist der einzige Schutz vor einer Ansteckung mit HIV? (*Kondom*)

Item 8

Original: Aufgabengruppe 9, Item 5

Change-einfacher: Welche Tiere sind die engsten Verwandten des Menschen?
(*Affen; Menschenaffen; Menschenartige; Hominidae; Hominiden*)

Item 9

Original: Aufgabengruppe 4, Item 4

Change-einfacher: Wie viel Kilogramm hat ungefähr ein Liter Wasser? (*1kg*)

Change-schwieriger: 2 Kilogramm Wasser sind ungefähr wie viel Liter? (*2l*)

Item 10

Original: Aufgabengruppe 11, Item 4

Change-einfacher: Das Saxophon ist für einen Musikstil typisch, der anfänglich vor allem von Afroamerikanern gespielt wurde. Welcher Musikstil ist das? (*Jazz*)

Change-schwieriger: Welcher Musikstil entstand in den USA im frühen 20. Jahrhundert und ist u.a. für seine Improvisationen bekannt? (*Jazz*)

Item 11

Original: Aufgabengruppe 5, Item 3

Change-einfacher: Die Erde dreht sich nicht nur um die Sonne, sondern sie dreht sich auch ... ? (*um sich selbst; um die eigene Achse*)

Change-schwieriger: Wie lange benötigt die Erde, um die Sonne ein Mal zu umkreisen? (*1 Jahr*)

Item 12

Original: Aufgabengruppe 11, Item 5

Change-einfacher: Die Stadt Tokyo liegt nicht in China, sondern ...? (*Japan*)

Change-schwieriger: In welchem Land liegt die Stadt Osaka? (*Japan*)

Item 13

Original: Aufgabengruppe 12, Item 2

Change-einfacher: Welcher Handwerker repariert einen Wasserrohrbruch, genauso wie eine verstopfte Toilette? (*Installateur; Klempner*)

Change-schwieriger: Welcher Handwerker verlegt auf einer Baustelle die Wasserleitungen? (*Installateur; Klempner*)

Item 14

Original: Aufgabengruppe 12, Item 3

Change-einfacher: Was sind die Olympischen Spiele? (*sportliche Wettkämpfe*)

Change-schwieriger: Was gab den Olympischen Spielen ihren Namen? (*der Berg Olymp*)

Item 15

Original: Aufgabengruppe 12, Item 4

Change-einfacher: Die höchste menschliche Stimmlage wird als Sopran bezeichnet. Nenne eine andere Stimmlage. (*Mezzosopran; Alt; Tenor; Bariton; Bass*)

Change-schwieriger: Die höchste menschliche Stimmlage ist der Sopran. Wie heißt die entsprechende Stimmlage bei Männern? (*Tenor*)

Item 16

Original: Aufgabengruppe 6, Item 2

Change-einfacher: Die Sonne ist für Ebbe und Flut mit verantwortlich. Welcher Himmelskörper noch? (*Mond*)

Change-schwieriger: Welche (physikalische) Kraft spielt eine Rolle bei der Entstehung von Ebbe und Flut? (*Anziehungskräfte; Gravitationskraft*)

Item 17

Original: Aufgabengruppe 6, Item 3

Change-einfacher: Was ist ein Triathlon? (*Sport; Sportart; sportlicher Bewerb / Wettbewerb*)

Change-schwieriger: Wie viele Kilometer muss man beim klassischen Triathlon (Langdistanz) ungefähr laufen? (*42 km; ungefähr 40 km*)

Item 18

Original: Aufgabengruppe 13, Item 5

Change-einfacher: In den meisten südamerikanischen Staaten ist Spanisch die Landessprache. In Brasilien jedoch nicht - welche Sprache ist in Brasilien die Landessprache? (*portugiesisch*)

Change-schwieriger: In welchem südamerikanischen Staat wird portugiesisch als Landessprache gesprochen? (*Brasilien*)

Item 19

Original: Aufgabengruppe 17, Item 4

Change-einfacher: Wozu werden Diamanten verwendet? (*Schmuck; Industriediamant; zum Bohren; Schneiden; Glasschneiden; Plattenspielnadel; Gravieren*)

Change-schwieriger: Woraus besteht ein (natürlicher) Diamant? (*Kohlenstoff*)

Item 20

Original: Aufgabengruppe 12, Item 5

Change-einfacher: Vincent van Gogh ist u.a. dafür bekannt, dass er sich ein Ohr abgeschnitten haben soll. Welchen Beruf hatte er? (*Maler; Künstler (des Expressionismus); Zeichner; Lithograph*)

Change-schwieriger: Vincent van Gogh war ein berühmter Maler. Welchem Kunststil werden seine Bilder größtenteils zugeordnet? ((*Anfänge des Expressionismus; Post-Impressionismus*))

Item 21

Original: Aufgabengruppe 5, Item 5

Change-einfacher: Was fließt in Arterien und Venen durch unseren Körper? (*Blut*)

Change-schwieriger: Mithilfe des Blutes wird der Sauerstoff im menschlichen Körper verteilt. Welche Aufgabe besitzt der Blutkreislauf noch? (*Nährstoffverteilung; Kohlendioxydableitung; Schlackenstoffableitung*)

Item 22

Original: Aufgabengruppe 6, Item 5

Change-einfacher: Nenne mir ein Streichinstrument. (*Geige; Kontrabass; Cello; Bratsche*)

Change-schwieriger: Nenne mir einen Unterschied zwischen Bratsche und Cello. (*Größe, Gewicht, Länge der Saiten, Trägheit der Saiten, Bogen, Haltung des Spielers*)

Item 23

Original: Aufgabengruppe 13, Item 1

Change-einfacher: Welche Religion hat der Papst? (*römisch-katholisch*)

Change-schwieriger: Welcher Konfession gehören Gläubige an, deren Kirche aus der Reformationsbewegung um Martin Luther entstand? (*evangelisch; protestantisch*)

Item 24

Original: Aufgabengruppe 5, Item 4

Change-einfacher: Wofür steht die Abkürzung EU? (*Europäische Union*)

Change-schwieriger: Wo ist der Sitz des Europäischen Parlaments? (*Straßburg*)

D Protokollbogen

PROTOKOLLBOGEN

Datum: _____

Testheft:	☐ 1	☐ 2	☐ 3

Rückmeldung:	☐ ja ... Code:	☐ nein	

Geburtsdatum:	Alter (Jahre, Monate):		

☐ männlich	☐ weiblich		

Muttersprache:	☐ deutsch		
	☐ andere:		

Schule:	☐ Wiener Mittelschule		
	☐ Gymnasium		

Klasse/Schulstufe:	_____		

Aufgaben

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	

E Elternbriefe

ELTERNBRIEF

Oktober 2017

Geschätzte Erziehungsberechtigte, werte Eltern!

Als **Altkalksburgerin** wende ich mich mit der Bitte um **Unterstützung** bei der Datenerhebung im Zuge meiner **Masterarbeit** an der Fakultät für Psychologie der **Universität Wien** an Sie.

Inhalt meiner Studie ist die Evaluierung der **Intelligenz-Testbatterie AID 3** (Erfassung kognitiver Fähigkeiten).

Ziel ist es, einen Beitrag zur Überprüfung der testtheoretischen Güte dieses psychologisch-diagnostischen Verfahrens, welches vor allem in der Schulleistungsdiagnostik eingesetzt wird, zu leisten.

Zu diesem Zweck bin ich **auf Ihre Unterstützung und die Mithilfe Ihres Kindes angewiesen** - nur durch Ihren Beitrag ist das Gelingen meiner Studie möglich!

Die Testung selbst wird von mir geleitet und findet für jedes Kind einzeln während der Unterrichtszeit in den Räumlichkeiten der Schule statt. Sie nimmt etwa 15 Minuten in Anspruch.

Dabei werden Fragen gestellt, die einerseits aus einem der zwölf Untertests des AID 3 stammen, und andererseits von mir selbst daraus entwickelt wurden.

Da das Ziel der vorliegenden Studie in der Evaluierung des Untertests liegt, sind Rückschlüsse auf die individuelle Leistung Ihres Kindes **nicht** von Interesse und auch nicht Gegenstand der Auswertung.

Sollten Sie dennoch an einer kurzen schriftlichen Ergebnismeldung interessiert sein, kann diese gesondert für Sie erstellt werden, nachdem sämtliche Daten erhoben worden sind. In diesem Fall wird Ihrem Kind zwecks Zuordnung ein Code zugewiesen.

Die Ergebnisse sowie sämtliche während der Untersuchung gewonnene Daten werden **streng vertraulich** behandelt und **anonymisiert**.

Die **Auswertung der Daten** erfolgt **ausschließlich für wissenschaftliche Zwecke**.

Die Ergebnisse der Testungen werden **NICHT** an die Schule weitergegeben.

Unter isabella_plank@hotmail.com stehe ich gerne für Ihre **Rückfragen** zur Verfügung.

Ich bitte Sie - mit Ihrer **Unterschrift** auf der folgenden Seite - der Teilnahme Ihres Kindes an der geplanten Studie zuzustimmen.

Mit freundlichen Grüßen,

Isabella Plank, BSc

Ich erkläre mich damit einverstanden, dass mein/e Sohn/Tochter

NAME: KLASSE:

an der Testung im Rahmen der Masterarbeit von Isabella Plank, BSc teilnimmt.

Ich möchte eine kurze schriftliche Ergebnisrückmeldung erhalten.

- ☐ ja
- ☐ nein

.....
Datum

.....
Unterschrift eines Erziehungsberechtigten

Sehr geehrte Erziehungsberechtigte!

Mein Name ist Isabella Plank und ich studiere Psychologie an der Universität Wien. Derzeit schreibe ich meine **Abschlussarbeit**. In dieser überprüfe ich die Gültigkeit eines seit vielen Jahren verwendeten Tests, der das Alltagswissen von Kindern erfasst.

Die **Befragung** wird von mir selbst geleitet und findet für jedes Kind einzeln während der Unterrichtszeit in der Schule statt. Sie dauert ungefähr 15 Minuten. Die Fragen beziehen sich auf Dinge, die Ihrem Kind aus dem Alltag bekannt sind.

Die Ergebnisse und alle Daten, die während der Untersuchung gewonnen werden, werden **streng vertraulich** behandelt und **anonymisiert**. Die **Auswertung der Daten** erfolgt **ausschließlich für wissenschaftliche Zwecke**.

Die Ergebnisse der Befragung werden **NICHT** an die Schule weitergegeben.

Unter isabella_plank@hotmail.com stehe ich gerne für Ihre **Rückfragen** zur Verfügung.

Ich bitte Sie – durch Ihre **Unterschrift** auf der folgenden Seite – der Teilnahme Ihres Kindes an der geplanten Befragung zuzustimmen.

Vielen Dank, mit freundlichen Grüßen,

Isabella Plank

Isabella Plank, BSc

Ich erkläre mich damit einverstanden, dass mein/e Sohn/Tochter

NAME: KLASSE:

an der Befragung teilnimmt.

.....
Datum

.....
Unterschrift eines Erziehungsberechtigten

F Informationsschreiben für Lehrkräfte

INFORMATIONSSCHREIBEN LEHRKRÄFTE

Oktober 2017

Geschätzte Lehrerinnen und Lehrer!

Mit der Bitte um **Unterstützung** bei der Datenerhebung im Zuge meiner **Masterarbeit** an der Fakultät für Psychologie der Universität Wien wende ich mich an Sie und Ihre SchülerInnen.

Gegenstand meiner Arbeit ist die **Intelligenz-Testbatterie *Adaptives Intelligenz Diagnostikum 3. „Generation“*** (AID 3, Version 3.1; Kubinger und Holocher-Ertl, 2014), welche der Erfassung kognitiver Fähigkeiten bei Kindern und Jugendlichen im Alter von 6-15 Jahren dient.

Im deutschsprachigen Raum findet der AID 3 häufig Anwendung bei schulpsychologischen Fragestellungen sowie in der Laufbahn- und Bildungsberatung.

Um sicherzugehen, dass psychologisch-diagnostische Verfahren wie der AID 3 Qualitätsstandards genügen und nützliche Instrumente für die Schulpsychologie darstellen, ist deren fortlaufende Überprüfung und Evaluierung erforderlich.

Im Zuge meiner Masterarbeit wird mittels eines experimentellen Ansatzes untersucht, ob einer der zwölf Untertests des AID 3 inhaltliche Gültigkeit besitzt. Zu diesem Zweck sollen Daten von **SchülerInnen der 5. und 6. Schulstufe** erhoben werden.

Selbstverständlich werden die Eltern vorab per **Elternbrief** über meine Studie informiert und darum gebeten, durch ihre Unterschrift ihr Einverständnis zur Teilnahme ihres Kindes zu geben.

Ich bitte Sie darum, diese Briefe an die Eltern Ihrer SchülerInnen weiterzuleiten und die unterschriebenen **Einverständniserklärungen** am Tag der Testung an mich zu retournieren.

Die **Testung** findet für jedes Kind einzeln während der Unterrichtszeit statt und wird von mir geleitet. Das Kind wird aus der Klasse abgeholt und nach erfolgter Testung in einem separaten Raum wieder dorthin zurück gebracht. Die Anwesenheit einer Lehrkraft während der Testung ist nicht erforderlich.

Jedem Kind werden (mündlich) etwa 35 altersangepasste Wissensfragen gestellt, die es (mündlich) zu beantworten hat. Pro Kind beträgt die Testdauer etwa 15 Minuten, wobei dies je nach Fähigkeiten individuell schwanken kann. Pausen sind jederzeit möglich.

Die **Auswertung** der Daten erfolgt ausschließlich für **wissenschaftliche Zwecke**, die Testergebnisse werden im Anschluss an die Testung **anonymisiert**.

Aufgrund der Zielsetzung der Studie sind Rückschlüsse auf individuelle Personen und ihre Leistungen nicht von Interesse und auch nicht Gegenstand der Auswertung. Ich bitte um Ihr Verständnis, dass die Testergebnisse **nicht** an die Schule (Lehrkräfte, Direktion) weitergegeben werden.

Bei Fragen stehe ich gerne unter isabella_plank@hotmail.com zur Verfügung.

Vielen Dank für Ihre Unterstützung, mit freundlichen Grüßen,

Isabella Plank

Isabella Plank, BSc