



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Dimensionalitätsvergleich der Reasoning-Tests *Gleichungen*,
Figurale Analogien und *Figurales Zweidimensionales Reasoning*“

verfasst von / submitted by

Benedikt Otto Johann Winter, BSc

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

Wien, 2016 / Vienna, 2016

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:

Psychologie

Betreut von / Supervisor:

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

Mitbetreut von / Co-Supervisor:

Danksagung

Zunächst möchte ich meinem Betreuer, Herrn Univ.-Prof. Dr. Mag. Klaus Kubinger für sein Vertrauen betreffs der Bearbeitung des Themas danken. Ebenso für die konstruktiven Anregungen zur Verbesserung der schriftlichen Arbeit.

Besonderer Dank gebührt meiner Sub-Betreuerin, Frau Mag. Serra Ünal. Ihr möchte ich für die Koordination des Projektes danken, außerdem für die Beantwortung jeglicher Fragen und die Hilfestellung bei der Organisation der Studie.

Für die Genehmigung und Ermöglichung der Testungen gilt mein Dank Herrn MinRat Mag. Dr. Norbert Leitner (BM.I SIAK). Darüber hinaus der Standortleiterin (BZ Wien) Frau Oberst Manuela Türk, sowie Herrn Erwin Theuritzbacher für die Organisation und Vereinbarung der Testtermine.

Für die Hilfe bei der Durchführung der Testungen danke ich Gülten Arslan, Mag. Joseph-Sebastian Steinlechner, sowie Nicole Prochaska.

Nicole möchte ich außerdem für ihre langjährige Freundschaft, den emotionalen Beistand, den fachlichen Austausch, sowie für das Korrekturlesen der Arbeit danken.

Spezieller Dank gilt auch Mag. Larissa Bartok für die Einführung in die Programmiersprache R. Darüber hinaus danke ich ihr für die Beantwortung meiner Fragen zu ihrem Verfahren F2R, wie auch für die Erstellung einer eigenen Testversion für diese Studie.

Meiner Schwester Rebecca Winter möchte ich dafür danken, dass sie trotz der uns trennenden Distanz stets an meiner Seite war und mich während des gesamten Studiums emotional gestützt hat.

Für das Akzeptieren meiner Entscheidung, in Wien zu studieren, sowie für die bedingungslose Unterstützung auf meinem Weg möchte ich meinen Eltern Hans und Linda Winter danken. Ohne sie wäre meine Studium nicht möglich gewesen.

Abstract (Deutsch)

Überprüft wurde die Konformität des fluiden Reasoning-Tests „Figurales Zweidimensionales Reasoning“ (F2R), eines 5x5 Matrizentests, mit dem Rasch-Modell. Zwei Items wurden aufgrund des Teilungskriteriums Testform ausgeschlossen, da die Itemschwierigkeit durch Positionsveränderung der gesuchten Zelle beeinflusst zu werden scheint. Ein drittes Item wurde anhand des Scores im F2R entfernt. Dadurch konnte a posteriori Modellgültigkeit erreicht werden. Unter Berücksichtigung der konventionellen Teilungskriterien (ohne Testform), genügte der Ausschluss eines Items. Der F2R besitzt somit das Potenzial, Rasch-Modell-konform zu sein. Neben der Erhebung der fluiden Intelligenz durch den F2R wurde die kristallisierte Intelligenz über die Tests „Gleichungen“ (numerisch) und „Figurale Analogien“ (figural) erfasst. Die Korrelation zwischen Gleichungen und F2R $r_s = 0,355$ zeigte sich signifikant stärker als erwartet. Dies lässt fluide Anteile am Lösungsprozess im Test Gleichungen vermuten. Allerdings erklärt selbst der stärkste dieser Zusammenhänge nur knapp 13% der Varianz der Leistungen im F2R. So konnte gezeigt werden, dass für die drei Verfahren diskriminante Validität gegeben ist. Sie sind daher unabhängig und erfassen, wie beabsichtigt, unterschiedliche Konstrukte.

Schlüsselwörter: schlussfolgerndes Denken, fluide Intelligenz, kristallisierte Intelligenz, Reasoning, Rasch-Modell, Diskriminante Validität, *Gleichungen*, *Figurale Analogien*, *Figurales Zweidimensionales Reasoning*, *F2R*

Abstract (English)

The conformity of the fluid reasoning test „Figurales Zweidimensionales Reasoning“ (F2R), a 5x5 matrices test, with the Rasch model has been reviewed. Based on the split criteria test form two items had to be removed as changing the asked object within the matrix seems to influence item difficulty. Another item was deleted because of the criteria score F2R. Thereby a posteriori validity of the Rasch model could be achieved. Using only conventional split criteria (without test form), removal of one item was sufficient. Concluding the F2R has potential to show conformity with the Rasch model. Despite measuring fluid intelligence with the F2R, crystallized intelligence has been examined by using the tests “Gleichungen” (numeric) as well as “Figurale Analogien” (figural). The correlation between Gleichungen and F2R showed as significantly stronger than expected $r_s = .355$. This suggests fluid aspects within the solving process of Gleichungen. Albeit even the strongest of the correlations only showed able to explain 13% of variance in F2R Score. Thus discriminant validity of the three test could be shown. They appear independent and measure different constructs, just as intended.

Keywords: reasoning, fluid intelligence, crystallized intelligence, Rasch-Modell, discriminant validity, *Gleichungen*, *Figurale Analogien*, *Figurales Zweidimensionales Reasoning*, *F2R*

Inhaltsverzeichnis

1. Einleitung	13
2. Theoretischer Hintergrund	14
2.1. Intelligenztheorien.....	14
2.1.1. Theorie der fluiden und kristallisierten Intelligenz	14
2.1.2. Reasoning	16
2.2. Das dichotome logistische Modell von Rasch	17
2.3. Verfahren.....	19
2.3.1. Gleichungen (GL; Kubinger & Gamsjäger, in Vorb.).....	19
2.3.2. Figurales Zweidimensionales Reasoning (F2R; Kubinger & Bartok, in Vorb.) ..	20
2.3.3. Figurale Analogien (FA; Ünal & Kubinger, in Vorb.).....	21
2.3.4. Anstrengungsvermeidungstest (AVT; Rollett & Rollett, 2000).....	22
3. Empirischer Teil	23
3.1. Hypothesen.....	23
3.2. Methode.....	25
3.2.1. Stichproben- und Untersuchungsplanung	25
3.2.2. Durchführung der Untersuchung	26
3.2.3. Stichprobenbeschreibung	27
3.3. Ergebnisse	29
3.3.1. Rasch-Modell-Analysen F2R	29
3.3.2. Zusammenhang zwischen den Verfahren.....	43
3.3.3. Berechnung Ergebnisbasiertes Risiko 2. Art.....	45
4. Diskussion und Ausblick	46
5. Zusammenfassung	50
5.1. Literaturverzeichnis.....	51
5.2. Abbildungsverzeichnis	56
5.3. Tabellenverzeichnis.....	57
A. Ansuchen an das Bundesministerium für Inneres	58

1. Einleitung

Die Vielfältigkeit der modernen Welt erfordert mehr denn je eine rasche Auffassungsgabe, das Erkennen von Zusammenhängen und das Lösen neuartiger Problemstellungen. Dies gilt gleichermaßen für den Beruf, das Studium wie auch für den Alltag. Die Fähigkeit des logischen Denkens findet sich daher sehr häufig in den Anforderungsprofilen etlicher Berufe (vgl. AMS, 2016). Auch in den Aufnahmeverfahren für eine Vielzahl von Studiengänge werden logisch-schlussfolgernde Fähigkeiten geprüft, beispielsweise für den Studiengang Humanmedizin (MedUni Wien, 2016).

Allgemein haben sich Intelligenztests als valide Prädiktoren für den Erfolg in Bildung und Beruf erwiesen (vgl. z.B. Süß, 2003a). Dies begründet die Bedeutung sowie den Bedarf qualitativ hochwertiger Tests zur Messung der Intelligenz, bzw. der Teilkomponente logisches Denken, für die Berufs- und Bildungswahlberatung, allgemein für die Auswahl Diagnostik.

Unter der Leitung von Herrn Univ.-Prof. Dr. Mag. Klaus Kubinger wurde in den vergangenen Jahren eine Reihe solcher Tests zum logisch-schlussfolgernden Denken entwickelt (Gleichungen; Gamsjäger, 2012; Figurale Analogien; Ünal, 2012; Figurales Zweidimensionales Reasoning; Bartok, 2015). Sie stellen den Anspruch, modernen wissenschaftlichen Standards – sprich: der modernen Testtheorie – zu genügen. Um dies sicherzustellen ist folglich die Evaluation und Überprüfung der Verfahren notwendig.

Die vorliegende Arbeit beschäftigt sich mit dem Dimensionalitätsvergleich der drei neu entwickelten Verfahren. Der Tests Figurales Zweidimensionales Reasoning (Kubinger & Bartok, in Vorb.) soll auf Konformität mit dem dichotom logistischen Modell von Rasch überprüft werden. Zusätzlich werden die Zusammenhänge zwischen den Leistungen in den Verfahren betrachtet, um die diskriminante Validität zu bestimmen.

Im Theorieteil soll zunächst ein kurzer Überblick über die zu erfassenden Konstrukte und Fähigkeiten der Verfahren gegeben werden. Ebenso werden die relevanten Aspekte der Testtheorie dargelegt, anhand welcher die Tests untersucht werden sollen. Im Anschluss werden Datenerhebung, Stichprobe, die Ergebnisse der Rasch-Modell-Analysen und die Zusammenhänge der Verfahren beschrieben und beurteilt.

2. Theoretischer Hintergrund

Umfassende Darstellungen des geschichtlichen Hintergrundes der psychologischen sowie der Intelligenzdiagnostik finden sich bereits in großer Zahl in Sammelwerken, Lehrbüchern und Enzyklopädien, auf welche an dieser Stelle verwiesen werden soll. Die Geschichte der Diagnostik findet sich aufgearbeitet beispielsweise bei Vanecek (2003) oder Zimbardo und Gerrig (2008).

Die Ursprünge der ersten Intelligenztests und die gängigen Intelligenztheorien finden sich z.B. bei Süß (2003b) sowie Stemmler, Hagemann, Amelang und Bartussek (2010). Die für die Arbeit unmittelbar relevanten Intelligenztheorien werden nachfolgend dargestellt.

Für den testtheoretischen Teil der Arbeit sei für die klassische Testtheorie auf Moosbrugger und Hartig (2003), für die „moderne“, probabilistische Testtheorie auf Kubinger (1989, 2003) verwiesen. Das dichotome logistische Modell von Rasch (1960, 1980) wird, aufgrund seiner Bedeutung für die Testkonstruktion sowie Kalibrierung, später bündig beschrieben.

2.1. Intelligenztheorien

2.1.1. Theorie der *fluiden und kristallisierten Intelligenz*

Die wichtigste Intelligenztheorie für die vorliegende Arbeit stellt die der *fluiden und kristallisierten Intelligenz* von Cattell (1963) dar. Die beiden Konstrukte, die Cattell (1963) anhand von Faktorenanalysen extrahierte, bilden die Grundlage für die drei eingesetzten Tests.

Cattell beschrieb kristallisierte Intelligenz („crystallized intelligence“, g_c) als Produkt vorangegangener Lernerfahrungen in bestimmten Wissensgebieten, dadurch auch abhängig von Kultur, Umwelt und Bildung der Person. Als Beispiele führte er Leistungen in Geographie oder Geschichte an, welche durch erworbenes Wissen bedingt sind. Kristallisierte Intelligenz ist demnach hauptsächlich durch kulturabhängige Gewohnheiten wie auch kulturspezifische Bildungssysteme bestimmt (Cattell, 1963).

Erfasst werden kann sie daher am besten durch Aufgaben, welche kulturspezifische Inhalte und Konzepte aufweisen (Horn & Cattell, 1966), wie beispielsweise

Sprachverständnis. Wortschatz-, Rechen- und Allgemeinwissenstests lassen sich ebenfalls dem latenten Konstrukt g_c zuordnen (Zimbardo & Gerrig, 2008).

Fluide Intelligenz („fluid intelligence“, g_f) hingegen wird benötigt, um sich neuartigen Situationen anzupassen und abstrakte Probleme zu lösen. Aufgrund dessen helfen vorangegangene Lernerfahrungen hier nicht. Die fluide Intelligenz ist hauptsächlich durch biologische Faktoren, wie Vererbung, oder Verletzungen des zentralen Nervensystems, beeinflusst (Horn & Cattell, 1966). Am stärksten vertreten ist g_f in Aufgaben, welche Induktion, semantische Beziehungen oder figurale Schlussfolgerungen benötigen.

Allgemein sind Tests mit für alle Personen gleichermaßen neuartigem Material am besten geeignet, um g_f zu erfassen. (Horn & Cattell, 1966). Bereits Cattell (1963) brachte g_f daher in Verbindung mit *culture fair* Tests, zu denen man auch Matrizentests zählt (Süß, 2003a).

Durch die Unterschiedlichkeit der beiden Faktoren zeigt sich auch ein jeweils anderer Entwicklungsverlauf. Fluide Intelligenz erreicht seinen Höhepunkt mit ca. 14 – 15 Jahren und fällt ab etwa 22 wieder ab (Cattell, 1963). Kristallisierte Intelligenz hingegen steigt in Abhängigkeit von der Kultur und der dort üblichen Lern- bzw. Ausbildungszeit bis zum 18. Lebensjahr oder sogar darüber hinaus an. Zusätzlich beginnt ihr Abfall später und findet in geringerem Ausmaß statt (Cattell, 1963).

Dies griff Cattell (1987) später in seiner Investment-Theorie wieder auf. Großteils genetisch bedingt, ist g_f als erstes vorhanden. Die fluide Intelligenz wird über die Zeit hinweg in Lernerfahrungen „investiert“ und „kristallisiert“ sich dadurch zu g_c . Während der Kindheit und Schulzeit ist der Zusammenhang der beiden Konstrukte folglich am größten ($r = 0,32 - 0,47$; Cattell, 1987, p. 135).

Durch die vielseitige Beeinflussung von g_c durch andere Variablen, wie z.B. Motivation, Interesse, Lerngelegenheiten sowie Gedächtnis, zeigt sich bei Erwachsenen ein deutlich geringerer Zusammenhang ($r = 0,16$; Cattell, 1987, p. 135). Die beiden Konstrukte g_c und g_f können ab dem Erwachsenenalter als voneinander unabhängige Faktoren angesehen werden (Cattell, 1987).

2.1.2. Reasoning

Betrachtet man die gängigen, umfassenden Intelligenzmodelle (vgl. z.B. Stemmler et al., 2010), so findet sich in den meisten eine Komponente für logisch-schlussfolgerndes Denken bzw. *Reasoning*. Beispielsweise beschrieb Thurstone in seiner Theorie der primary abilities (1938), Induction / Reasoning als „Fähigkeit zum induktiven und deduktiven Denken“ (Süß, 2003b, S. 219).

Nach Cattell (1963) kann sowohl kristallisierte, als auch fluide Intelligenz dazu genutzt werden, Aufgaben, welche einen logischen Schluss erfordern, zu lösen. Allgemein zeigt sich die Fähigkeit zum logisch-schlussfolgernden Denken als bedeutsam. Definiert wird Reasoning nach Kubinger (2009) wie folgt: “Reasoning ist die Fähigkeit, Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge zu erkennen und zweckentsprechend verwerten zu können” (S.206).

Demnach könnten alle Aufgaben und Tests, welche die beschriebene Fähigkeit benötigen, Reasoning zugeordnet werden. Unterschieden wird dann lediglich noch das dazu verwendete Material, z.B. numerisch, figural, lexikalisch.

2.2. Das dichotome logistische Modell von Rasch

Im Mittelpunkt der Item Response Theory (IRT) steht das dichotome logistische Modell von Rasch (1960, 1980), kurz Rasch-Modell (vgl. Kubinger, 2003).

Wie in Gleichung 1 dargestellt, beschreibt es „die Wahrscheinlichkeit, dass Testperson (Tp) ν Item i löst („+“), in Abhängigkeit eines Personenparameters ξ_ν , d.i. die (wahre) Fähigkeit von ν , und eines Itemparameter σ_i , d.i. die (wahre) Schwierigkeit von i “ (Kubinger, 2003, S. 416).

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

Wie jedes andere, besitzt auch dieses Modell Voraussetzungen, wobei hier direkt auf Rasch (1960, 1980) verwiesen sei. Vollständige Darstellungen finden sich z.B. bei Fischer (1974), sowie bei Kubinger (2003, 2009) bzw. Kubinger, Rasch & Yanagida (2011a).

Von Relevanz für die vorliegende Arbeit ist, dass bei Gültigkeit des Modells, Personenfähigkeit und Itemschwierigkeit eindimensional gemessen werden können. Beabsichtigt man Personen anhand der „Anzahl gelöster Aufgaben“ miteinander zu vergleichen, müssen die Items eines Tests mit dem Modell konform sein. Nur dann enthält dieser Testwert „die gesamte relevante Information in Bezug auf die fragliche Fähigkeit der Tp“ (Kubinger, 2003, S.417) und ermöglicht den fairen Vergleich von Personen (vgl. z.B. Kubinger et al., 2011a). Die mathematischen Beweisführungen hierzu finden sich bei Fischer (1974, 1995).

Des Weiteren macht die Prüfbarkeit des Modells es für die Testkonstruktion interessant. Den bekanntesten, globalen Modelltest stellt dabei der „Likelihood-Ratio-Test (LRT) von Andersen“ (1973) dar. Er baut auf einer weiteren zentralen Eigenschaft des Rasch-Modells auf, der „Stichprobenunabhängigkeit“ (vgl. z.B. Kubinger, 2003). Demnach ist der Vergleich zweier Itemparameterschätzungen von der verwendeten Stichprobe statistisch unabhängig.

Dies prüft nun der LRT von Andersen, dessen Nullhypothese annimmt, „dass in jeder der beiden Teilgruppen dieselben Schwierigkeitsparameter gelten“ (Kubinger et al., 2011a, S. 557). Übersteigt die Testgröße den kritischen Wert, so sind die Items des Tests nicht mit dem Rasch-Modell konform.

Dies kann ebenfalls grafisch illustriert werden („Grafische Modellkontrolle“), indem die geschätzten Itemparameter für die zwei Teilstichproben in ein rechtwinkliges Koordinatensystem eingetragen werden. Sofern die Schätzungen gleich sind, müssen sie auf bzw. nahe einer 45° Geraden liegen (vgl. z.B. Kubinger, 2003).

Eine weitere Möglichkeit stellt der „ z -Test“ von Fischer und Scheiblechner (1970) mit Varianzschätzung nach Wald (1943) dar. Allerdings bringt das Testen auf Itemebene eine deutliche Überhöhung des Risikos 1. Art mit sich, sodass sich der „ z -Test“ nicht als Signifikanztest empfiehlt (Kubinger, 2003). Wenn der LRT nach Andersen einen signifikanten Wert liefert, kann der z -Test hypothesengenerierend verwendet werden, inwiefern welche/s Item/s ausgeschlossen werden sollte/n. Dies lässt sich aber auch anhand der grafischen Modellkontrolle bewerkstelligen.

Mit dieser Vorgehensweise werden dann sukzessiv so lange nicht konforme Items ausgeschlossen, bis der LRT nach Andersen für sämtliche Teilungskriterien nicht signifikante Werte ergibt. A posteriori kann Modellgültigkeit angenommen werden.

2.3. Verfahren

2.3.1. Gleichungen (GL; Kubinger & Gamsjäger, in Vorb.)

Das Verfahren Gleichungen wurde von Gamsjäger (2012) entwickelt und soll Reasoning anhand numerischen Aufgabenmaterials erfassen. Dies setzt Wissen um Rechenoperationen und –zeichen voraus, sodass „Reasoning sensu Cattell nicht im Bereich der fluiden, sondern der kristallisierten Intelligenz erfasst wird“ (Gamsjäger, 2012, S. 15).

Die einzelnen Aufgaben bestehen aus jeweils einer mathematischen Gleichung, für die aus einem Multiple-Choice-Antwortformat die richtige Kombination für x und y gewählt werden muss (Abbildung 1). Bei jeweils vier Antwortmöglichkeiten ergibt sich dadurch eine a priori Ratewahrscheinlichkeit von $\frac{1}{4} \times \frac{1}{4} = \frac{1}{16} = 6,25 \%$ (Gamsjäger, 2012).

$x + y = 10$

x	2	8	5	10
y	9	5	1	3

Abbildung 1: Instruktionsitem *Gleichungen* (Gamsjäger, 2012, S. 31)

Die Konformität des Tests mit dem dichotom logistischen Modell von Rasch wurde mehrfach untersucht und konnte sowohl bei der Erstellung, a posteriori (Gamsjäger, 2012), als auch später a priori (Schnait, 2015) gezeigt werden.

In der vorliegenden Arbeit wurde eine Version aus 15 Items und einer Zeitbeschränkung von zehn Minuten verwendet.

2.3.2. *Figurales Zweidimensionales Reasoning (F2R; Kubinger & Bartok, in Vorb.)*

Bartok (2015) erstellte einen neuartigen figuralen Reasoningtest, welcher eine nicht vollständig ausgefüllte 5x5 Matrix sowie ein offenes Antwortformat verwendet (Abbildung 2). Der Matrizentest erfasst die fluide Intelligenz.

○	◯	○	.	○
.	.	○○	.	.
.	.	.	?	.
.	.	○○○○	.	.
.	.	○○○○○	.	.

Abbildung 2: Instruktionsitem *F2R* (Bartok, 2015, S. 31)

Die Testerstellerin konnte nach Ausschluss von 15% der Items a posteriori Rasch-Modell-Konformität erreichen. Im Rahmen dieser Arbeit sollen daher die Analysen nochmals anhand einer neuen Stichprobe wiederholt werden. Eigens dafür wurde von Bartok eine Version aus 15 Items und einer Zeitbeschränkung von 20 Minuten zusammen gestellt.

2.3.3. Figurale Analogien (FA; Ünal & Kubinger, in Vorb.)

Das Verfahren Figurale Analogien (Ünal, 2012) wurde dazu konzipiert, kristallisiertes schlussfolgerndes Denken mittels figuralen Materials zu erfassen. Verwendet werden dazu Analogien, also Vergleiche zweier Objekte bezüglich ihrer Ähnlichkeit, welche einen logischen Schluss erfordern (Abbildung 3).

Die Aufgabeninhalte erfordern Wissen aus unterschiedlichen Bereichen:

- „1. Menschen: Geographie, Geschichte, Kultur;
2. Naturwissenschaften: Physik, Chemie, Biologie, Mathematik;
3. Alltag: Wissen über alltägliche Symbole und Gegenstände, Sport, Musik“

(Ünal, 2012, S. 46).

Die dafür notwendigen vorangegangenen Lernerfahrungen bedingen die Zuordnung des Verfahrens zur kristallisierten Intelligenz.

Anzumerken ist die relativ hohe Ratewahrscheinlichkeit, welche beim „1 aus 5“ Multiple-Choice-Antwortformat des Tests 20% beträgt.

Die Rasch-Modell-Analysen bei der Erstellung des Verfahrens ergaben a priori Modellgültigkeit, sodass die Anzahl gelöster Aufgaben als fairer Testkennwert verwendet werden kann.

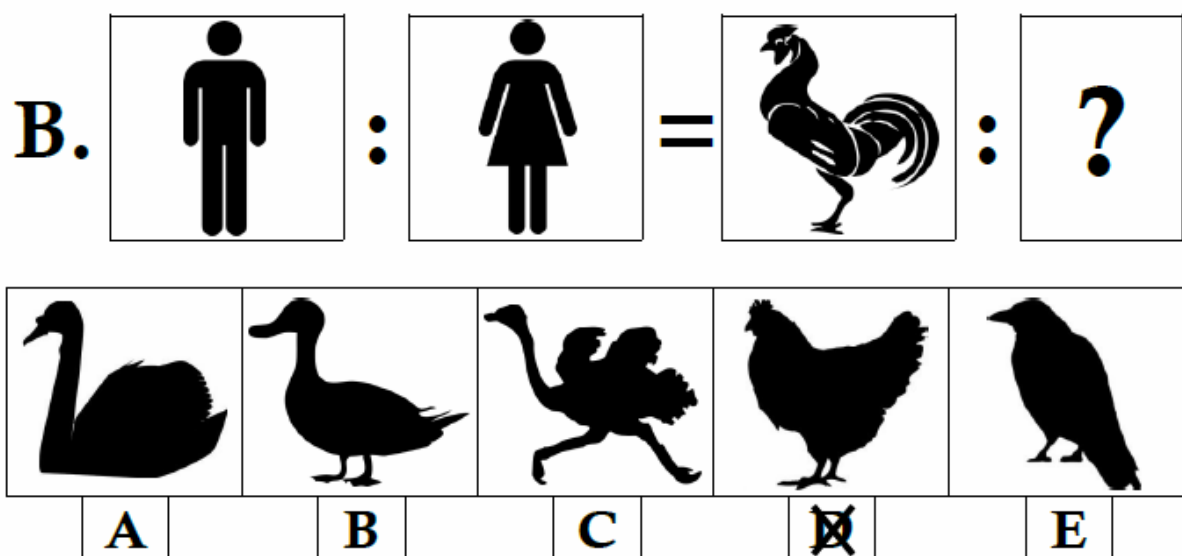


Abbildung 3: Instruktionsitem *Figurale Analogien* (Ünal, 2012, S. 44)

2.3.4. Anstrengungsvermeidungstest (AVT; Rollett & Rollett, 2000)

Der Fragebogen „Anstrengungsvermeidungstest“ (AVT; Rollett & Bartram, 1998) wurde dazu konzipiert, die schulische Anstrengungsvermeidung sowie den schulischen Pflichteifer zu erfassen. Definiert wird Anstrengungsvermeidung nach Rollett & Bartram als „Tendenz, sich ... emotionell negativ erlebten Anstrengungen durch den Einsatz geeigneter Vermeidungsstrategien zu entziehen“ (S. 7ff, zit. nach Gashi, 2011).

Verwendet wurde eine, für ältere Schüler adaptierte, Version aus dem Jahre 2000 (Rollett & Rollett). In dieser soll zu 32 Aussagen, betreffend Fühlen und Handeln beim Lernen und Arbeiten, Stellung bezogen werden. Das Antwortformat ist vierstufig; von „stimmt nicht“, „stimmt eher nicht“, „stimmt eher“ bis „stimmt“.

Für die modifizierte Version wird ein Gesamtwert für die Anstrengungsvermeidung verrechnet, welcher als externes Teilungskriterium für die Rasch-Modell-Analysen verwendet wurde.

3. Empirischer Teil

3.1. Hypothesen

Nachfolgend sind die zu prüfenden Hypothesen angeführt:

H₁ (1):

Die Items des Verfahren *Figurales Zweidimensionales Reasoning* von Kubinger & Bartok (in Vorb.) sind mit dem dichotom logistischen Modell von Rasch konform.

Zusammenhänge zwischen den einzusetzenden Verfahren:

Die hierzu aufgestellten Hypothesen basieren auf den Konstrukten der fluiden und kristallisierten Intelligenz sensu Cattell (1963), sowie dem Aufgabenmaterial der Verfahren (s. Tabelle 1).

Tabelle 1: *Überblick über die Konstrukte und Darbietungsform der Verfahren*

Aufgabenmaterial	Zuordnung sensu Cattell	
	Kristallisierte Intelligenz	Fluide Intelligenz
figural	Figurale Analogien (FA)	Figurales Zweidimensionales Reasoning (F2R)
numerisch	Gleichungen (GL)	

H₁ (2):

Der Zusammenhang zwischen den Verfahren Gleichungen und Figurale Analogien ist stärker als zwischen F2R und Gleichungen.

Gleichungen und Figurale Analogien sollen beide kristallisierte Intelligenz sensu Cattell erfassen, F2R hingegen fluide Intelligenz.

H₁ (3):

Der Zusammenhang zwischen den Verfahren Gleichungen und Figurale Analogien ist stärker als zwischen F2R und Figurale Analogien.

Diese Erwartung ergibt sich ebenfalls wie **H₁ (2)** auf Basis der zu erfassenden Konstrukte.

H₁ (4):

Der Zusammenhang zwischen den Verfahren F2R und Figurale Analogien ist stärker als zwischen F2R und Gleichungen.

Aufgrund des unterschiedlichen Aufgabenmaterials wird erwartet, dass der Zusammenhang zwischen den beiden Verfahren mit figuralem Inhalt stärker ist als mit dem Verfahren mit numerischem Material.

Insgesamt ergibt sich, basierend auf der Theorie der fluiden und kristallisierten Intelligenz, die zusammenfassende Darstellung in Tabelle 2.

Tabelle 2: *Theoriegeleitete, erwartete Rangfolge der Korrelationen zwischen den Verfahren*

$$r_s(\text{GL, FA}) > r_s(\text{F2R, FA}) > r_s(\text{F2R, GL})$$

Betrachtet man bereits gefundene Zusammenhänge anderer Studien, zeigt sich ein Widerspruch zu den rein theoriegeleiteten Hypothesen. So fand Bartok (2015) einen Zusammenhang zwischen Gleichungen und F2R von $r_s = 0,415$. Bei Schnait (2015) betrug der Korrelationskoeffizient zwischen Gleichungen und Figurale Analogien $r = 0,33$.

Die beiden Korrelationskoeffizienten unterscheiden sich auf Basis der verwendeten Stichprobengrößen nicht signifikant ($p = 0,300$; $r_1 = 0,415$; $n_1 = 192$; $r_2 = 0,33$; $n_2 = 265$) voneinander. Dennoch sind diese Befunde für die vorliegende Studie von Interesse. Ob sich dieser Widerspruch erneut zeigt, an welcher Stelle sich die Korrelation zwischen F2R und Figurale Analogien einordnet, oder, ob sich die Zusammenhänge den Hypothesen entsprechend darstellen werden.

3.2. Methode

3.2.1. Stichproben- und Untersuchungsplanung

Die Stichprobenplanung für Rasch-Modell-Analysen ist grundsätzlich möglich (siehe Kubinger, Rasch & Yanagida, 2009). Für verschiedene Szenarien wurden die notwendigen Stichprobengrößen berechnet (Kubinger, Rasch & Yanagida, 2011b), allerdings findet sich darunter kein Richtwert für einen Test mit 15 Items. Folglich war lediglich eine Orientierung an ähnlichen Fällen bzw. allgemeinen Empfehlungen möglich.

So schreiben Kubinger et al. (2011a, S.557) für den Fall von 20 Items: „Bei $k = 20$ im Schwierigkeitsbereich von -3 bis 3 gleichmäßig verteilten Aufgaben und einem relevanten Effekt derart, dass zwei (eher mittelschwierige) Aufgaben zwischen zwei Teilstichproben jeweils eine Differenz in der Parameterschätzung im Ausmaß eines Sechstels der Spannweite aller Schwierigkeitsparameter aufweisen, braucht man für $\alpha = 0,05$ und $\beta = 0,20$ einen Stichprobenumfang von 2 mal 101 Testpersonen.“

Allgemein ergibt sich, dass die Stichprobe, mit Blick auf das Risiko 2. Art 2×200 bis 250, hinsichtlich Risiko 1. Art ($\alpha = 0,05$) allerdings nicht größer als 2×150 sein sollte. Mit 2×150 Testpersonen kann eine Abweichung von 0,6 (1/10 Parameterspannweite), betreffend 3×2 Items, bei insgesamt 30 Items, mit $\alpha = 0,01$ und $1-\beta$ von beinahe 0,8 entdeckt werden (Kubinger et al., 2011b, S.331f).

Daran orientiert wurde für den vorliegenden Fall von $k = 15$, sowie angestrebtem $\alpha = 0,05$ und $\beta = 0,20$ ein Stichprobenumfang von ca. 2×100 bis 150 Testpersonen abgeleitet.

Die Relevanz von Berufs- und Bildungsberatung ist vor allem für junge Erwachsene bzw. Schüler in Abschlussklassen gegeben, sodass eine Stichprobe im Alter von 16 bis 25, bzw. ab 16 Jahren angestrebt wurde. Des Weiteren können die Verfahren in Gruppen administriert werden, wie sie beispielsweise in Schulen oder ähnlichen Institutionen vorkommen.

Die Stichproben der Verfahren Gleichungen (Kubinger & Gamsjäger, in Vorb.) sowie Figurale Analogien (Ünal & Kubinger, in Vorb.) bestanden aus Auszubildenden der Sicherheitsakademien in Niederösterreich, sodass auch diesmal ein Ansuchen an das Bundesministerium für Inneres gestellt wurde (Anhang A).

Das Ansuchen wurde im November 2015 gestellt und enthielt eine Beschreibung des Untersuchungsablaufs, sowie Informationsblätter zur Testung allgemein und den Verfahren im Speziellen. Letztere sind nicht erneut im Anhang angeführt, siehe hierzu Kapitel 2.3.

Die Genehmigung für die Durchführung der Untersuchung wurde im Dezember 2015 erteilt und Termine am Bildungszentrum Wien vereinbart.

3.2.2. Durchführung der Untersuchung

An fünf Terminen im Januar und Februar 2016 (15., 18., 20., 22.01 und 04.02.16) wurden insgesamt 15 Klassen Auszubildender am Bildungszentrum Wien der Sicherheitsakademie (Marokkanergasse 4, 1030 Wien) getestet.

Administriert wurden die Untersuchungen vom Autor selbst, zusätzlich war jeweils ein/-e weitere/-r Student/-in des Arbeitsbereichs Psychologische Diagnostik der Fakultät für Psychologie zur Unterstützung anwesend.

Die Auszubildenden wurden halbstandardisiert über Freiwilligkeit der Teilnahme sowie Anonymisierung der Daten informiert. Die Möglichkeit zur Ergebnismeldung anhand eines fünfstelligen Probandencodes wurde erklärt und entsprechende Informations- / Rückmeldungsblätter verteilt. Des Weiteren wurde darauf hingewiesen, dass die Teilnahme die schulischen Bewertungen in der Sicherheitsakademie nicht beeinflussen.

Die Bearbeitungszeit der gesamten Testbatterie sowie eines Deckblattes zu demographischen Daten betrug pro Klasse 50 bis 55 Minuten. Die genaue Zusammensetzung, Reihenfolge und Dauer der einzelnen Teile der Testbatterie ist Tabelle 3 zu entnehmen. Es wurde darauf geachtet, dass die zwei verschiedenen Testheftversionen gleich häufig Verwendung fanden.

Bei den Instruktionen des Verfahrens F2R gab es in einigen Klassen Verständnisschwierigkeiten, welche durch ein erneutes Vorlesen geklärt werden konnten.

Tabelle 3: *Itemanzahl und Zeitbeschränkungen der Testbatterie*

Verfahren	Itemanzahl	Zeitbeschränkung
Gleichungen	15	10 Minuten
Figurales Zweidimensionales Reasoning	15	20 Minuten
Figurale Analogien	27	8 Minuten
Anstrengungsvermeidungstest	32	keine

3.2.3. Stichprobenbeschreibung

Die deskriptiven Analysen und zugehörigen Abbildungen wurden mit SPSS 22.0 (IBM, 2013) erstellt. Insgesamt besteht die Stichprobe aus $n = 327$ Personen, 70,2% Männer und 29,8 % Frauen. Die Geschlechtsverteilung ist demnach nicht repräsentativ für die österreichische Bevölkerung (ca. 50:50, vgl. Statistik Austria, 2016a).

Der Median des Alters beträgt 24 Jahre ($M_{\text{Alter}} = 25,47$; $SD_{\text{Alter}} = 5,11$) mit einer Spannweite von 19 bis 51 Jahren. 94,5 % der TeilnehmerInnen gaben Deutsch als Muttersprache an, 5,5% andere Sprachen. Dies ergibt sich teilweise aus dem Aufnahmekriterium der österreichischen Staatsbürgerschaft bei der Polizei. Somit ebenfalls nicht repräsentativ für die Gesamtbevölkerung (ca. 87% österreichische Staatsbürgerschaft, vgl. Statistik Austria, 2016b). Die Häufigkeitsverteilung der höchsten abgeschlossenen Ausbildung ist Tabelle 4 zu entnehmen.

Die Verteilung der Rohwerte in den Verfahren F2R und AVT sind in den Abbildungen 4 und 5 dargestellt. Der Median gelöster Aufgaben im F2R betrug 8 von 15 ($M_{\text{F2R}} = 8,2$; $SD_{\text{F2R}} = 2,68$). Die beiden Testformen wurden gleich häufig (49,6 : 50,4%) vorgegeben und bearbeitet, um sie später als Teilungskriterium nutzen zu können. Im Anstrengungsvermeidungstest war der Median der Rohwerte 6 von maximal 32 ($M_{\text{AVT}} = 7,19$; $SD_{\text{AVT}} = 4,72$).

Tabelle 4: Häufigkeiten der höchsten abgeschlossenen Ausbildung

	Absolute Häufigkeit	Prozent
Pflichtschule	14	4,3
Lehre	49	15,1
BHS	114	35,2
AHS	123	38,0
Hochschule	24	7,4
Gesamt	324	100,0

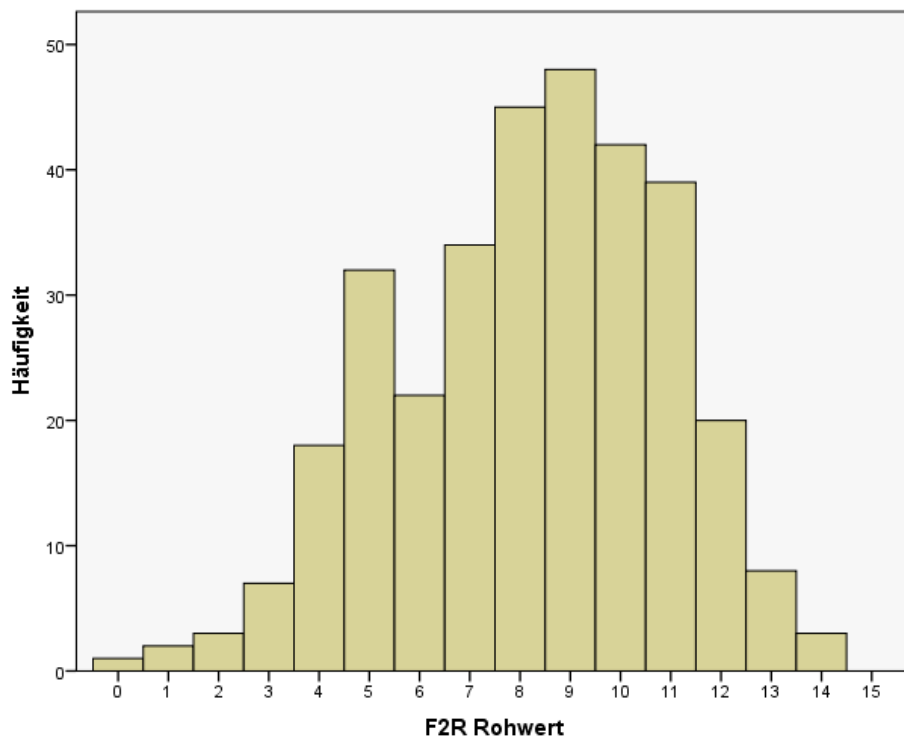


Abbildung 4: Verteilung der Rohwerte im F2R (vor den Rasch-Modell-Analysen)

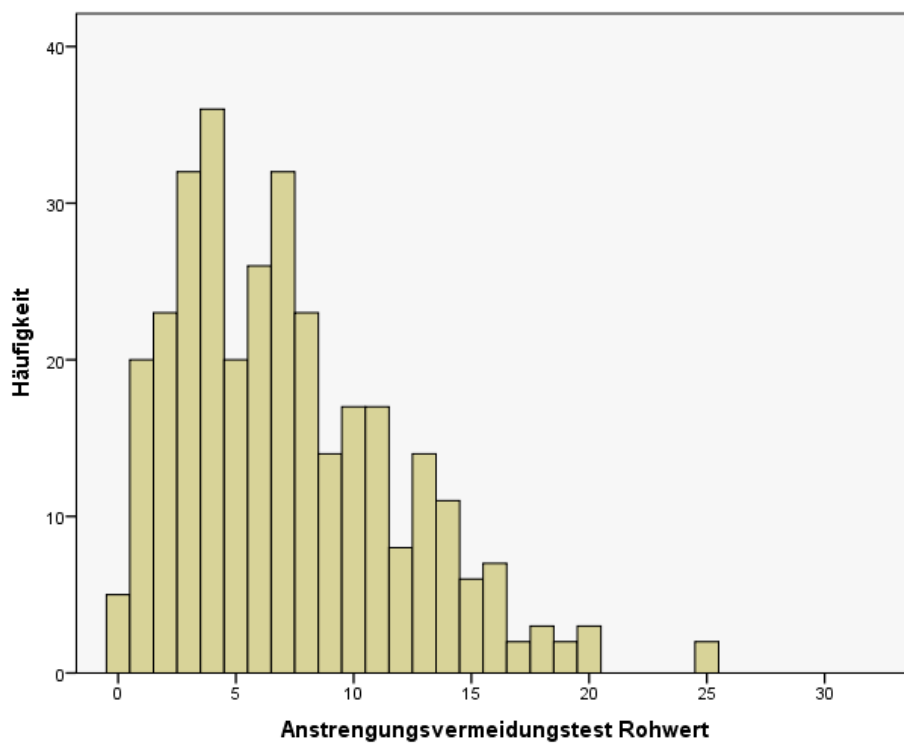


Abbildung 5: Verteilung der AVT Rohwerte

3.3. Ergebnisse

3.3.1. Rasch-Modell-Analysen F2R

Die nachfolgenden Analysen wurden in der freien Programmiersprache R (R Core Team, 2015a), unter Verwendung der Benutzeroberfläche „R Studio“ (RStudio Team, 2015) durchgeführt. Zum Laden der Daten aus SPSS wurde das R-Pakete foreign (R Core Team, 2015b) verwendet.

Das globale Risiko 1. Art wurde in der Untersuchungsplanung auf $\alpha = 0,05$ festgelegt, daraus ergibt sich für fünf Teilungskriterien ein vergleichsbezogenes Risiko von $\alpha = 0,01$ für jeden durchzuführenden LRT.

Teilungskriterien:

- Score F2R ($\leq$ Median vs. $>$ Median)
- Testform (A vs. B)
- Alter ($\leq$ Median vs. $>$ Median)
- Ausbildung (Pflichtschule, Lehre, BHS vs. AHS, Hochschule)
- Score AVT ($\leq$ Median vs. $>$ Median)

Die Teilungskriterien sollten miteinander nicht stark korreliert sein, was durch Kreuztabellen überprüft wurde. Die Ergebnisse sind in Tabelle 5 zusammengefasst. Es ergaben sich keine relevanten Zusammenhänge zwischen den Teilungskriterien, sodass die fünf genannten Kriterien für die Analysen zum Rasch-Modell beibehalten wurden.

Tabelle 5: Korrelationsmatrix der zu verwendenden Teilungskriterien

Teilungskriterium	Score F2R	Testform F2R	Alter	Ausbildung	Score AVT
Score F2R					
Testform F2R	$\chi^2 = 2,083$ $p = 0,149$ $\Phi = 0,080$				
Alter	$\chi^2 = 1,825$ $p = 0,177$ $\Phi = -0,076$	$\chi^2 = 1,532$ $p = 0,216$ $\Phi = 0,070$			
Ausbildung	$\chi^2 = 3,868$ $p = 0,049$ $\Phi = -0,110$	$\chi^2 = 0,030$ $p = 0,862$ $\Phi = 0,010$	$\chi^2 = 1,546$ $p = 0,214$ $\Phi = -0,070$		
Score AVT	$\chi^2 = 1,242$ $p = 0,265$ $\Phi = 0,062$	$\chi^2 = 4,969$ $p = 0,026$ $\Phi = 0,124$	$\chi^2 = 0,388$ $p = 0,534$ $\Phi = -0,035$	$\chi^2 = 1,285$ $p = 0,257$ $\Phi = 0,063$	

Anmerkung: $df = 1$; $\chi^2_{\text{krit}} = 3,841$

Als Erstes wurden mit Hilfe des R-Paketes eRm (Mair, Hatzinger & Maier, 2015) die Itemparameter geschätzt. Anschließend wurde der LRT nach Andersen berechnet und grafische Modellkontrollen für die einzelnen Teilungskriterien erstellt. Die Ergebnisse des LRT für alle Items sind in Tabelle 6 angeführt.

Es zeigten sich signifikante Werte für die Teilungskriterien Score F2R sowie die Testform. In den grafischen Modellkontrollen (s. Abbildungen 6 – 10) waren einerseits die Items 12 und 16 für den Score, andererseits Item 14 und 15 für die Testform auffällig. Zur Verbesserung der Übersichtlichkeit, wurden nur die Konfidenzelipsen, welche die 45° Gerade nicht schneiden, eingezeichnet. Gleiches gilt für die Beschriftung der Items, auch diese wurde nur für auffällige Items vorgenommen.

An dieser Stelle sei auf die abweichende Bezeichnung der Items hingewiesen, sie entspricht der Originalnummerierung von Bartok (2015). Ursprünglich wurden 20 Items erstellt, wovon die Items 3, 6 und 9 von der Testerstellerin im Zuge ihrer Rasch-Modell-Analysen ausgeschlossen wurden. Zusätzlich wurden die Items 19 und 20 entfernt um die, in der vorliegenden Arbeit verwendete, Version aus 15 Items zu erstellen.

Tabelle 6: *F2R – LRT für alle Items*

Teilungskriterium	χ^2	df	χ^2_{krit}	p	
Score F2R	31,967	14	29,141	0,004	sign.
Testform	64,796	14	29,141	0,000	sign.
Alter	12,209	14	29,141	0,590	nicht sign.
Ausbildung	10,830	14	29,141	0,699	nicht sign.
Score AVT	20,639	14	29,141	0,111	nicht sign.

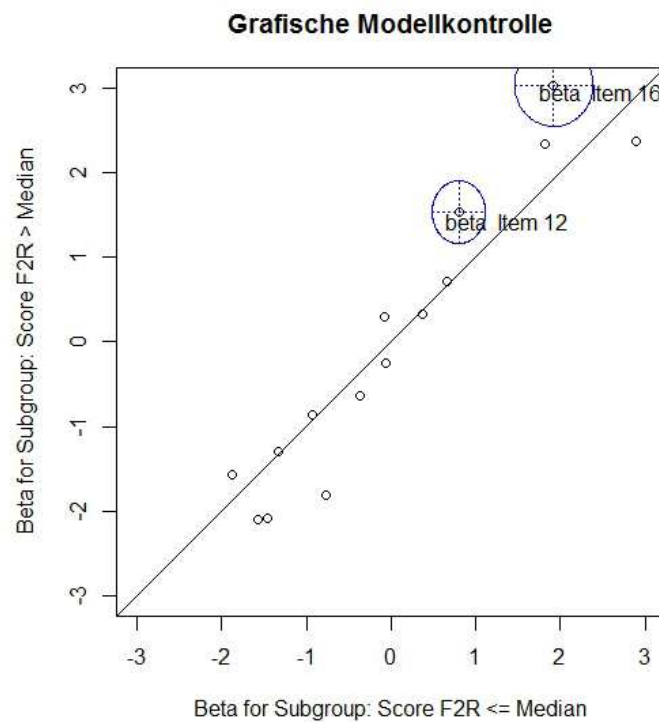


Abbildung 6: Grafische Modellkontrolle für das Teilungskriterium Score, alle Items

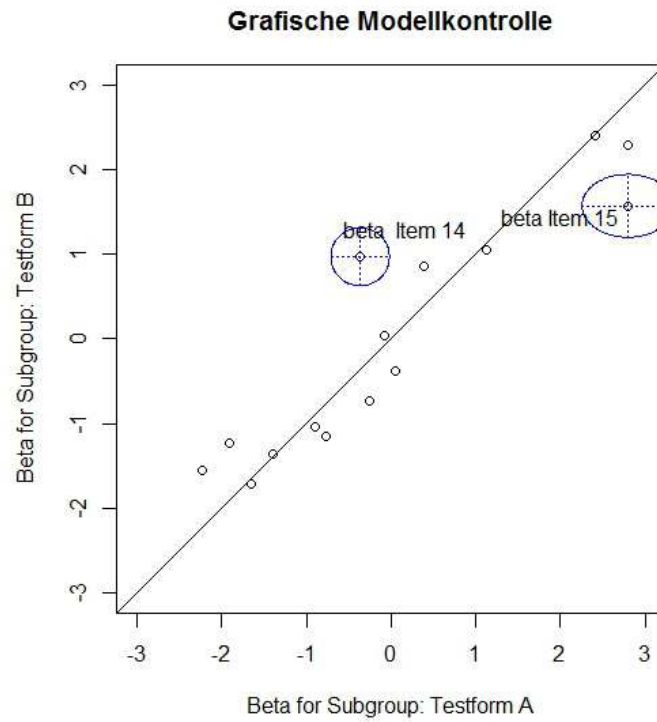


Abbildung 7: Grafische Modellkontrolle für das Teilungskriterium Testform, alle Items

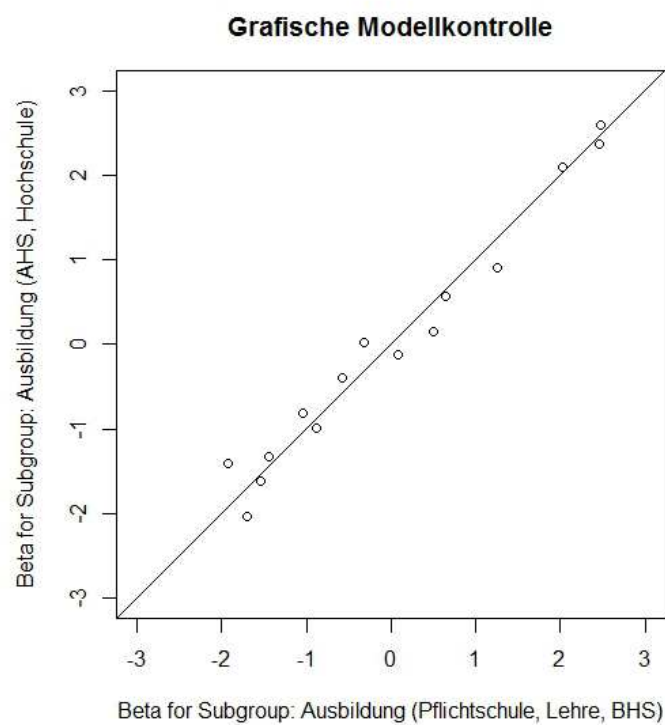


Abbildung 8: Grafische Modellkontrolle für das Teilungskriterium Ausbildung, alle Items

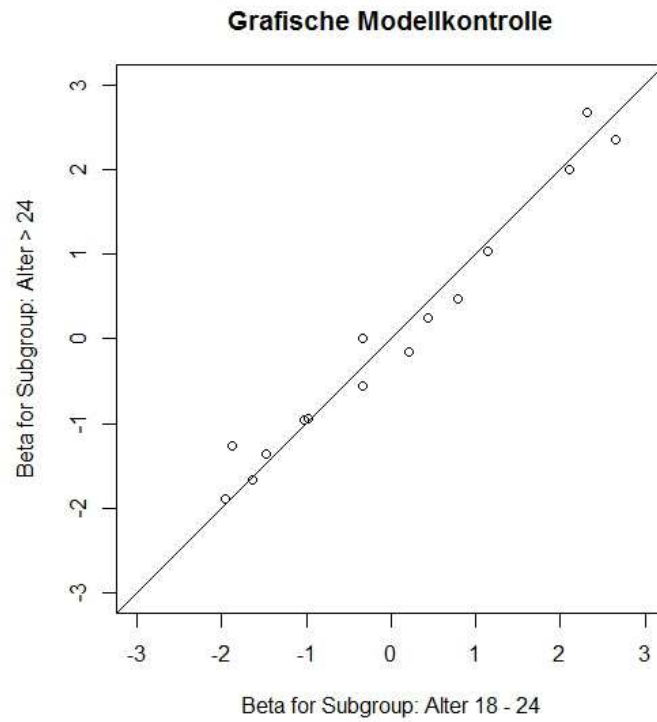


Abbildung 9: Grafische Modellkontrolle für das Teilungskriterium Alter, alle Items

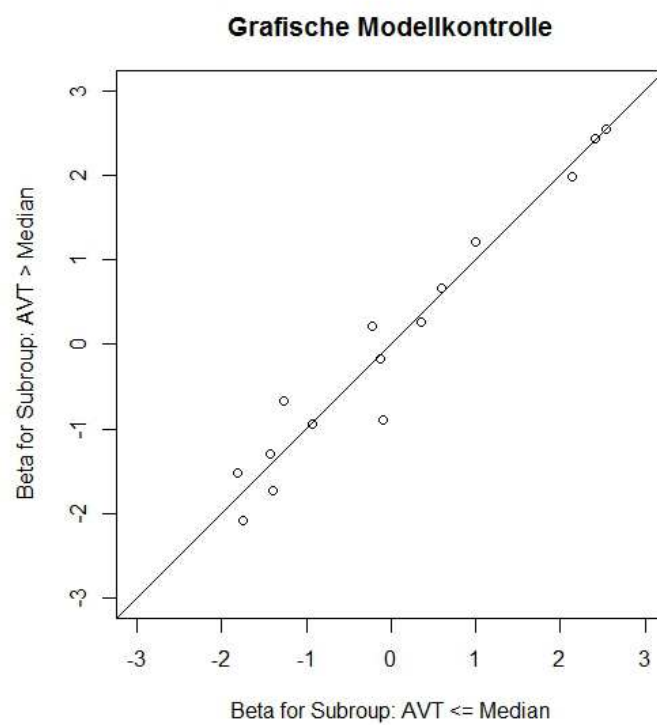


Abbildung 10: Grafische Modellkontrolle für das Teilungskriterium Score AVT, alle Items

Die größte Abweichung von der 45° Geraden wies Item 14 auf, sodass dieses als erstes ausgeschlossen wurde. Zusätzlich stützten inhaltliche Überlegungen dieses Vorgehen. So ließ sich bereits rein inhaltlich die Konformität des Items mit dem Rasch-Modell ausschließen. In den Abbildungen 11 und 12 ist das Item in den beiden Testformen dargestellt. Sie unterscheiden sich aufgrund eines Konstruktionsfehlers, welcher vor den Testungen nicht bemerkt wurde.

11.

			.	
.	.		.	.
.	.		.	.
.	.		?	.
.	.		.	.

Lösung:



☐ Ich kenne die Lösung nicht.

Abbildung 11: Item 14, Testform A, als „einfacher“ geschätzt

11.

	.			
.	.		.	.
.	.		.	.
.	?		.	.
.	.		.	.

Lösung:



☐ Ich kenne die Lösung nicht.



Abbildung 12: Item 14, Testform B, als „schwieriger“ geschätzt

Anders als die restlichen Items, lassen sich die beiden Testformen von Item 14 nicht zu einer gemeinsamen Matrix vereinen. Damit dies möglich wäre, müsste das Dreieck in Testform B (Zeile eins, Spalte vier) ebenfalls entsprechend gedreht sein (siehe Abbildung 12,

zusätzliches Dreieck oberhalb Spalte vier). Nur dann ergäbe sich logisch konsistent die erwartete Lösung (im Feld Lösung). Denkt man die tatsächlich vorgegebene Matrix durch, so ergibt sich eine andere Lösung – das Dreieck müsste gegen den Uhrzeigersinn gedreht werden (siehe zusätzlich eingezeichnete Lösung).

Die eigentlich logisch konsistente Lösung der Testform B wurde jedoch von keiner einzigen Testpersonen angegeben, sondern die ursprünglich erwartete Lösung (wie im Feld Lösung). Es verwundert daher nicht, dass sich die geschätzten Schwierigkeiten der Items für die beiden Testformen unterscheiden. Während in Form A sowohl Pfeil als auch Dreieck mit dem Uhrzeigersinn gedreht werden müssen, funktioniert Testform B anders. Hier wird zwar ebenfalls der Pfeil mit dem Uhrzeigersinn gedreht, das Dreieck allerdings gegen den Uhrzeigersinn (für eine konsistente Lösung) bzw. behält es in Spalte zwei und vier seine Form und verändert sich in den restlichen. Scheinbar wurde die letztere Regelkombination von den TeilnehmerInnen angewandt, sodass sich für Form B eine zusätzliche anzuwendende Regel und folglich höhere Schwierigkeit ergab. Item 14 wurde daher ausgeschlossen und die Itemparameter neu geschätzt.

Tabelle 7 zeigt die Ergebnisse des LRT nach Andersen nach Ausschluss von Item 14. In den grafischen Modellkontrollen verblieb Item 16 für das Teilungskriterium Score auffällig (Abbildung 13), der LRT ergab keinen signifikanten Wert mehr.

Die Teststatistik für das Teilungskriterium Testform blieb signifikant, ein weiterer Itemausschluss war notwendig. Für die Testform zeigte die grafische Kontrolle (Abbildung 14) Item 15 als abweichend, die beiden Versionen des Items sind in Abbildung 15 und 16 dargestellt. Inhaltlich konnte kein spezieller Grund für die Auffälligkeit des Items gefunden werden. Das Verändern der Position des Fragezeichens, sprich der gesuchten Zelle, scheint die Schwierigkeit jedoch zu verändern, obgleich es sich um die gleiche Matrix handelt. Das Item wurde ausgeschlossen und die Analysen erneut durchgeführt.

Tabelle 7: *F2R – LRT nach Ausschluss von Item 14*

Teilungskriterium	χ^2	df	χ^2_{krit}	p	
Score F2R	27, 510	13	27, 688	0, 011	nicht sign.
Testform	34, 791	13	27, 688	0, 001	sign.
Alter	11, 866	13	27, 688	0, 539	nicht sign.
Ausbildung	9, 022	13	27, 688	0, 771	nicht sign.
Score AVT	20, 247	13	27, 688	0, 089	nicht sign.

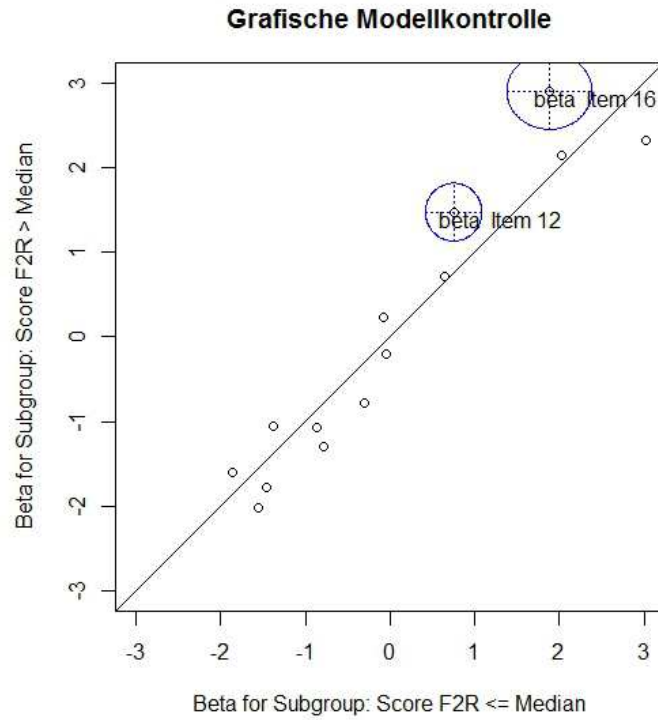


Abbildung 13: Grafische Modellkontrolle für das Teilungskriterium Score, nach
Ausschluss von Item 14

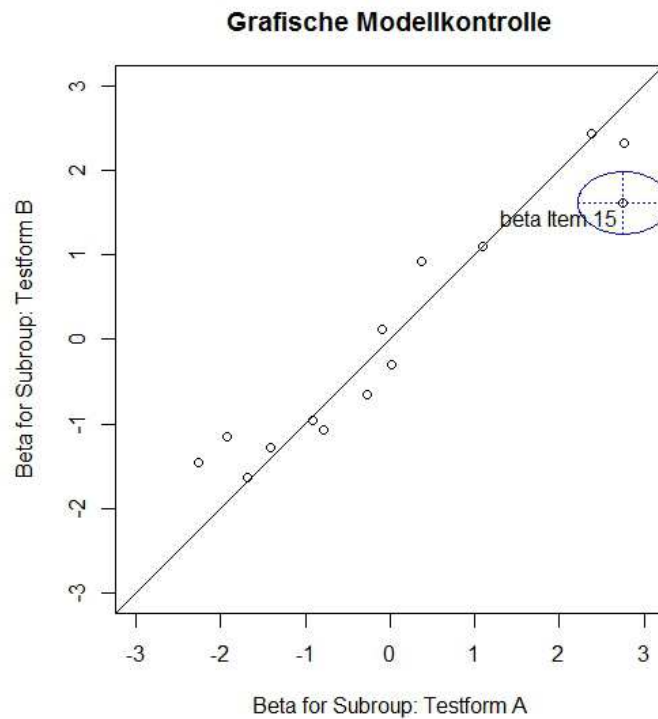


Abbildung 14: Grafische Modellkontrolle für das Teilungskriterium Testform, nach
Ausschluss von Item 14

12.

.	$\square\Delta\leftarrow$	$\leftarrow\square+$	$\square\leftarrow\bullet$	.
$\square\rightarrow+$	$\square\square\rightarrow$	$\rightarrow\square\bullet$	$\square\rightarrow\circ$	$\square\Delta\rightarrow$
.	$\square+\leftarrow$	$\leftarrow\square\square$	$\square\leftarrow\Delta$	.
?	$\square\bullet\rightarrow$	$\rightarrow\square\circ$	$\square\rightarrow+$	.
.	$\square\circ\leftarrow$	$\leftarrow\square\Delta$	$\square\leftarrow\square$	.

Lösung:

$\square\rightarrow\Delta$

☐ Ich kenne die Lösung nicht.

Abbildung 15: Item 15, Testform A, als „schwieriger“ geschätzt

12.

.	$\square\Delta\leftarrow$	$\leftarrow\square+$	$\square\leftarrow\bullet$	.
$\square\rightarrow+$	$\square\square\rightarrow$	$\rightarrow\square\bullet$	$\square\rightarrow\circ$	$\square\Delta\rightarrow$
.	$\square+\leftarrow$	$\leftarrow\square\square$	$\square\leftarrow\Delta$	.
.	$\square\bullet\rightarrow$	$\rightarrow\square\circ$	$\square\rightarrow+$	?
.	$\square\circ\leftarrow$	$\leftarrow\square\Delta$	$\square\leftarrow\square$	.

Lösung:

$\square\square\rightarrow$

☐ Ich kenne die Lösung nicht.

Abbildung 16: Item 15, Testform B, als „einfacher“ geschätzt

Tabelle 8 beinhaltet die Ergebnisse des LRT nach Ausschluss von Item 14 und 15. Während das Teilungskriterium Testform nicht mehr aufschien (Abbildung 17), zeigte sich jetzt wieder ein signifikanter Wert für den Score F2R. Auch grafisch wurde die Abweichung von Item 16 von der 45° Geraden deutlicher (Abbildung 18), die beiden Versionen des Items sind in den Abbildungen 19 und 20 angeführt.

Inhaltlich konnte keine Begründung gefunden werden, wieso das Item für Personen aus der Gruppe Score > Median als schwieriger geschätzt wurde. Item 16 wurde ausgeschlossen und die Parameterschätzungen abermals vorgenommen.

Tabelle 8: *F2R – LRT nach Ausschluss von Item 14 und Item 15*

Teilungskriterium	χ^2	df	χ^2_{krit}	p	
Score F2R	27,521	12	26,217	0,006	sign.
Testform	22,677	12	26,217	0,031	nicht sign.
Alter	11,946	12	26,217	0,450	nicht sign.
Ausbildung	8,965	12	26,217	0,706	nicht sign.
Score AVT	19,863	12	26,217	0,070	nicht sign.

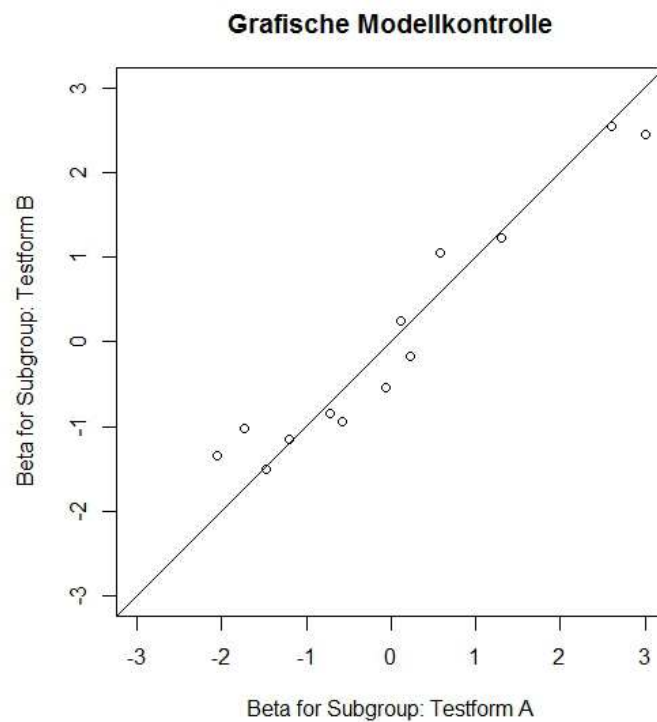


Abbildung 17: Grafische Modellkontrolle für das Teilungskriterium Testform, nach Ausschluss von Item 14 und 15

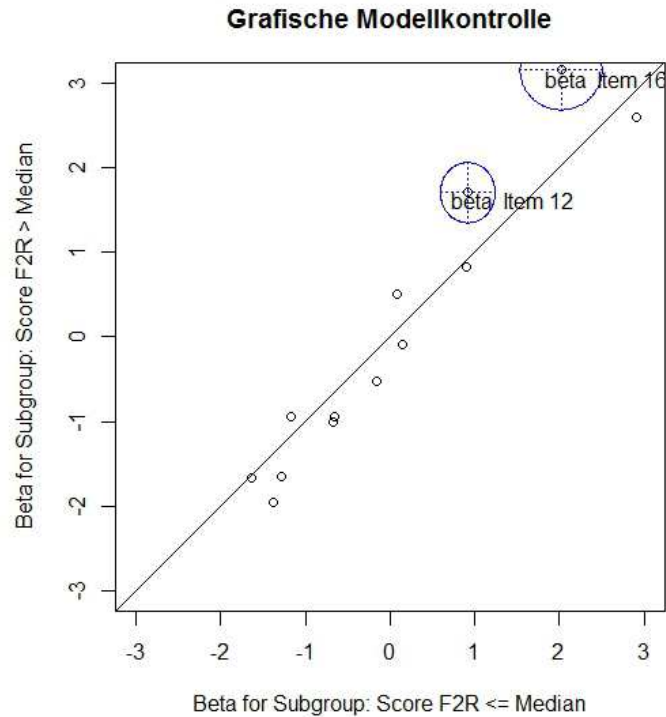


Abbildung 18: Grafische Modellkontrolle für das Teilungskriterium Score, nach Ausschluss von Item 14 und 15

13.

☐ ☐ ☒ ☐ ☐	☐ ☐ ☒ ☐ ☐	☐ ☒ ☐ ☐ ☐	☐ ☒ ☐ ☐ ☐	.
.	.	☐ ☐ ☒ ☐ ☐	.	.
.	.	.	.	?
.	.	☐ ☒ ☐ ☐ ☐	.	.
.	.	☐ ☒ ☐ ☐ ☐	.	.

Lösung:

☐ ☒ ☐ ☐ ☐

☐ Ich kenne die Lösung nicht.

Abbildung 19: Item 16, Testform A, als „schwieriger“ für Score F2R > Median geschätzt

13.

.	☐ ☐ ☒ ☐ ☐	☐ ☒ ☐ ☐ ☐	☐ ☒ ☐ ☐ ☐	☐ ☒ ☐ ☐ ☐
.	.	☐ ☐ ☒ ☐ ☐	.	.
?	.	.	.	.
.	.	☐ ☒ ☐ ☐ ☐	.	.
.	.	☐ ☒ ☐ ☐ ☐	.	.

Lösung:

☐☒☐☐☐
☐ Ich kenne die Lösung nicht.

Abbildung 20: Item 16, Testform B, als „schwieriger“ für Score F2R > Median geschätzt

In Tabelle 9 sind die LRT nach Ausschluss der drei Items 14, 15 sowie 16 aufgelistet. Für keines der fünf Teilungskriterien ergab sich ein signifikanter Wert, a posteriori Modellgültigkeit wurde erreicht. Grafisch verblieb Item 12 für das Teilungskriterium Score auffällig (Abbildung 21), der LRT ergab jedoch keinen signifikanten Testwert.

Tabelle 9: F2R – LRT nach Ausschluss von Item 14, Item 15 und Item 16

Teilungskriterium	χ^2	df	χ^2_{krit}	p	
Score F2R	23,527	11	24,725	0,015	nicht sign.
Testform	21,039	11	24,725	0,033	nicht sign.
Alter	11,187	11	24,725	0,428	nicht sign.
Ausbildung	8,760	11	24,725	0,644	nicht sign.
Score AVT	20,415	11	24,725	0,040	nicht sign.

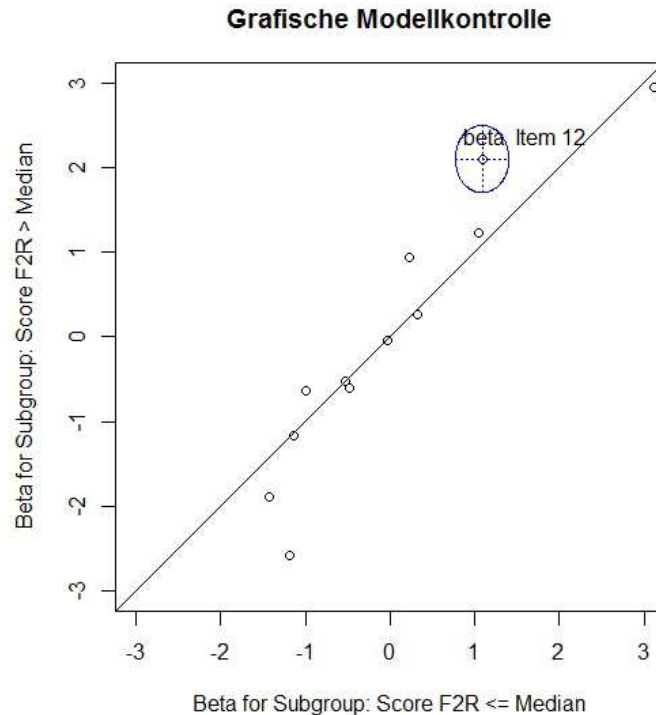


Abbildung 21: Grafische Modellkontrolle für das Teilungskriterium Score, nach
Ausschluss von Item 14, 15 und 16

Die Verwendung des Teilungskriteriums Testform war bereits bei Bartok (2015) speziell zur Überprüfung gedacht, ob die Erstellung verschiedener Parallelförmchen zu unterschiedlichen Schwierigkeitsparametern führen würde. Anhand gängiger Teilungskriterien (Score, Geschlecht, Schultyp, Muttersprache) hatte Bartok (2015) bei ihren Analysen zum Rasch-Modell kein Item ausschließen müssen. Daher soll auch in der vorliegenden Arbeit verglichen werden, wie viele Items ohne das Teilungskriterium Testform auffallen. Für den separaten Durchgang mit vier Teilungskriterien liegt das vergleichsbezogene Risiko 1. Art bei $\alpha = 0,0125$, um weiterhin das globale Risiko 1. Art von $\alpha = 0,05$ einzuhalten.

Wie in Tabelle 5 dargestellt, hatte sich bereits beim Durchgang mit allen Items ein signifikanter Wert für das Teilungskriterium Score ergeben. Item 16 war in der grafischen Modellkontrolle aufgefallen (Abbildung 6) und war demnach auszuschließen. Die Abweichungen für die Testform (Item 14 und 15) waren jedoch größer und so wurden zunächst diese Items ausgeschlossen. In der alternativen Analyse wurde das Teilungskriterium Testform nicht verwendet und direkt mit dem Ausschluss von Item 16 begonnen. Die Ergebnisse sind in Tabelle 10 angeführt, es ergaben sich keine signifikanten Werte. Bereits nach Ausschluss eines Items wurde a posteriori Modellgültigkeit erreicht. Die grafische Modellkontrolle für das Teilungskriterium Score ist in Abbildung 22 dargestellt, weiterhin verblieb Item 12 auffällig.

Tabelle 10: F2R – LRT mit 4 Teilungskriterien, nach Ausschluss von Item 16

Teilungskriterium	χ^2	df	χ^2_{krit}	p	
Score F2R	20,251	13	27,688	0,089	nicht sign.
Alter	10,709	13	27,688	0,635	nicht sign.
Ausbildung	11,631	13	27,688	0,558	nicht sign.
Score AVT	20,975	13	27,688	0,073	nicht sign.

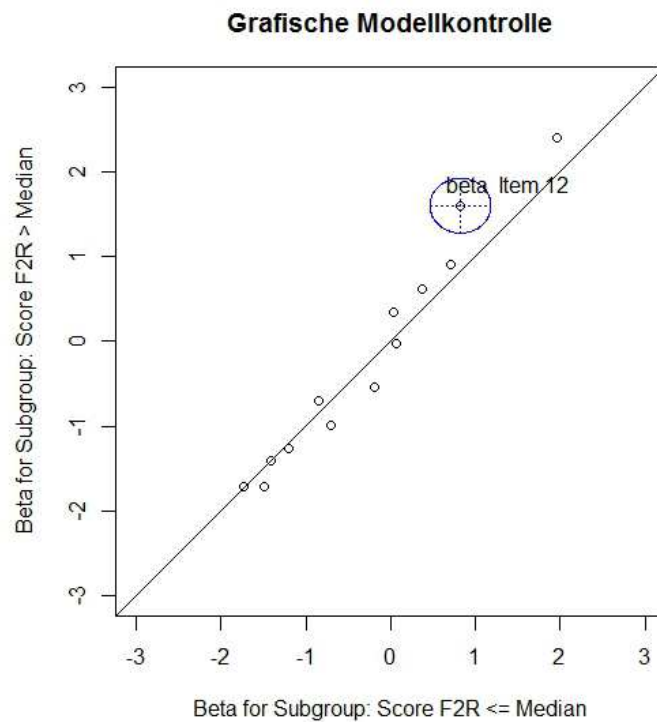


Abbildung 22: Grafische Modellkontrolle für das Teilungskriterium Score, nach Ausschluss von Item 16 bei vier Teilungskriterien

3.3.2. Zusammenhang zwischen den Verfahren

Zur Prüfung der diskriminanten Validität der Verfahren wurden jeweils Rangkorrelationskoeffizienten nach Spearman berechnet. Hierzu wurde die Anzahl gelöster Aufgaben in den einzelnen Verfahren verwendet, da sich alle drei Tests als mit dem Rasch-Modell konform darstellen. Bei F2R wurden nur die mit dem Rasch-Modell konformen Items zur Summenbildung verwendet (ohne Item 14, 15, und 16).

Basierend auf der Theorie der fluiden und kristallisierten Intelligenz von Cattell (1963) sollen die Verfahren unterschiedliche Konstrukte erfassen, es wurde daher eine bestimmte Reihung der Korrelationskoeffizienten erwartet (s. Tabelle 2, Kapitel 3.1)

Der Korrelationskoeffizient zwischen Gleichungen und Figurale Analogien beträgt $r_s = 0,222$, das Bestimmtheitsmaß $R^2 = 0,049$. Demnach werden ca. 5% der Varianz der Leistung in Gleichungen durch Figurale Analogien erklärt.

F2R und Figurale Analogien korrelieren zu $r_s = 0,171$, das Bestimmtheitsmaß ist hier $R^2 = 0,029$. Die Leistung in Figurale Analogien erklärt daher etwa 3% der Varianz der Leistung in F2R.

Zwischen F2R und Gleichungen beträgt der Rangkorrelationskoeffizient nach Spearman $r_s = 0,355$, das Bestimmtheitsmaß $R^2 = 0,126$. Es werden 12,6% der Varianz der Leistung in F2R durch die Leistung in Gleichungen erklärt.

Insgesamt ergibt sich die in Tabelle 11 angeführte Reihenfolge der Zusammenhänge zwischen den Verfahren.

Tabelle 11: Erwartete und empirisch gefundene Reihung der Korrelationen zwischen den Verfahren

Erwartet	r_s (GL, FA)	>	r_s (F2R, FA)	>	r_s (F2R, GL)
Beobachtet	r_s (F2R, GL) = 0,355	>	r_s (GL, FA) = 0,222		r_s (F2R, FA) = 0,171

Zusätzlich wurde untersucht, ob sich die Korrelationskoeffizienten signifikant voneinander unterscheiden. Dies wurde mit Hilfe des R Paketes psych (Revelle, 2015) bewerkstelligt, die durchgeführten Steiger Tests (1980) für abhängige Korrelationskoeffizienten ergaben die in Tabelle 12 dargestellten Ergebnisse. Es wird dabei jeweils die Korrelation r_{12} mit r_{13} verglichen, wobei gleichzeitig die Korrelation r_{23} berücksichtigt wird.

Tabelle 12: *Steiger Tests auf Unterschiede zwischen abhängigen Korrelationskoeffizienten*

r_{12}	r_{13}	r_{23}	t	df	t_{krit}	p (2-seitig)	
F2R x GL = 0, 355	F2R x FA = 0, 171	GL x FA = 0, 222	2, 810	321	1, 649	0, 005	sign.
GL x F2R = 0, 355	GL x FA = 0, 222	F2R x FA = 0, 171	1, 980	321	1, 649	0, 049	sign.
FA x GL = 0, 222	FA x F2R = 0, 171	GL x F2R = 0, 355	0, 820	321	1, 649	0, 410	nicht sign.

Demnach unterscheidet sich der Korrelationskoeffizient zwischen Figurale Analogien und Gleichungen nicht von dem zwischen Figurale Analogien und F2R, daher wurde in Tabelle 10 auf eine Reihung der beiden Korrelationen verzichtet. Die Hypothese H_1 (3): „Der Zusammenhang zwischen den Verfahren Gleichungen und Figurale Analogien ist stärker als zwischen F2R und Figurale Analogien“ ist folglich abzulehnen, die H_0 (3): r_s (GL, FA) = r_s (F2R, FA) beizubehalten.

Des Weiteren zeigt sich das Gegenteil der postulierten H_1 (2): „Der Zusammenhang zwischen den Verfahren Gleichungen und Figurale Analogien ist stärker als zwischen F2R und Gleichungen“, die Tests F2R und Gleichungen weisen einen signifikant stärkeren Zusammenhang auf als Gleichungen und Figurale Analogien. Gleiches gilt für die H_1 (4): „Der Zusammenhang zwischen den Verfahren F2R und Figurale Analogien ist stärker als zwischen F2R und Gleichungen“, auch hier widerspricht die beobachtete Korrelation der erwarteten.

Demnach korreliert die Leistung im Verfahren Gleichungen stärker mit der in F2R als mit der Leistung in Figurale Analogien, die Reihung der beiden Korrelationen in Tabelle 10 scheint berechtigt.

3.3.3. Berechnung Ergebnisbasiertes Risiko 2. Art

Für die Korrelationen wurde jeweils das ergebnisbasierte Risiko 2. Art (vgl. Kubinger et al., 2011a, S. 205f) berechnet. Dies ist in R mit dem Paket pwr (Champely, 2015) möglich.

Wie in Tabelle 13 dargestellt, wurde für die Korrelationen zwischen GL und FA, sowie zwischen F2R und GL praktisch kein Risiko 2. Art eingegangen. Für den Zusammenhang zwischen F2R und FA betrug β hingegen 7,3 %.

Tabelle 13: Ereignisbasiertes Risiko 2. Art für die Zusammenhänge zwischen den Verfahren

	r_s	n	α	Power
GL x FA	0,222	322	0,05	0,991
F2R x GL	0,355	323	0,05	0,999
F2R x FA	0,171	322	0,05	0,927

4. Diskussion und Ausblick

Das Ziel der vorliegenden Arbeit war die erneute Durchführung von Rasch-Modell-Analysen für das Verfahren Figurales Zweidimensionales Reasoning (Kubinger & Bartok, in Vorb.) sowie die Bestimmung der diskriminanten Validität zu zwei weiteren Tests zum logischen Denken.

Nach Ausschluss von zwei Items bezüglich des Teilungskriteriums Testform sowie eines Items aufgrund des Kriteriums Score konnte a posteriori Konformität mit dem Rasch-Modell für das Verfahren erreicht werden. Eines der beiden, für die Testformen auffälligen, Items musste wegen eines Konstruktionsfehlers ausgeschlossen werden. Für das zweite konnte keine plausible inhaltliche Begründung gefunden werden. Um eine parallele Testform zu erstellen, wurde innerhalb der gleichen Matrix die Position der gesuchten Zelle variiert. Dies scheint jedoch die Itemschwierigkeit in manchen Fällen unvorhersehbar zu verändern.

Auch für das Item, das anhand des Teilungskriteriums Score auszuschließen war, konnte keine inhaltliche Erklärung gefunden werden.

Von den insgesamt fünfzehn Items mussten also drei, relativ betrachtet 20% ausgeschlossen werden. Dies liegt deutlich über dem Richtwert von 10%, ab dem eine erneute Prüfung anhand einer weiteren Stichprobe notwendig ist (Kubinger & Draxler, 2007).

Zusätzlich wurde ein alternativer Durchgang der Analysen vorgenommen, bei dem die Testform nicht mit einbezogen wurde. Die Analyse ergab unter Ausschluss eines Items (6,67%) Konformität mit dem Rasch-Modell. Dies stützt die Annahme von Bartok, „dass das Material grundsätzlich eine Eignung bezüglich einer Testzusammenstellung aufweist (2015, S.70).“ Die Empfehlung (Bartok, 2015), sich bei neuen Studien auf eine Testform zu beschränken, kann ebenfalls bestärkt werden. Sofern mehrere Formen gewünscht sind, könnten diese durch Variation der Reihenfolge erstellt werden.

Abgesehen vom Verfahren F2R wurden zwei weitere Tests, Gleichungen und Figurale Analogien, vorgegeben. Zur Ermittlung der diskriminanten Validität wurden die Rangkorrelationskoeffizienten der Leistungen in den drei unterschiedlichen Verfahren zum logischen Denken berechnet.

Die beiden Tests zur kristallisierten Intelligenz sensu Cattell, Gleichungen sowie Figurale Analogien, korrelieren zu $r_s = 0,222$. Es soll die gleiche Fähigkeit erfasst werden, wobei sich jeweils das Aufgabenmaterial unterscheidet (numerisch vs. figural). Das dennoch

niedrige Bestimmtheitsmaß von $R^2 = 0,049$ spricht dafür, dass die Tests nicht austauschbar sind und das Konstrukt der kristallisierten Intelligenz verschieden erfassen.

Der Zusammenhang zwischen F2R und Figurale Analogien beträgt $r_s = 0,171$. Lediglich ca. 3 % der Varianz der Leistung in F2R werden durch die in Figurale Analogien erklärt. Da der F2R fluide, Figurale Analogien hingegen kristallisierte Intelligenz erfassen soll, ist dies als positiv zu sehen. Ein geringer Korrelationskoeffizient weist demnach darauf hin, dass die beiden Tests tatsächlich unterschiedliche Teilbereiche der Intelligenz erheben.

Am höchsten ist der Korrelationskoeffizient zwischen F2R und Gleichungen $r_s = 0,355$. Demnach können 12,6% der Varianz der Leistung in Gleichungen durch die in F2R erklärt werden. Insgesamt zeigt sich durch die Ergebnisse ein Widerspruch zu den rein theoriegeleiteten Hypothesen. Basierend auf der Theorie der fluiden und kristallisierten Intelligenz von Cattell (1963) wurde erwartet, dass F2R und Gleichungen den geringsten Zusammenhang aufweisen würden, der zwischen Figurale Analogien und Gleichungen sollte am stärksten sein und der zwischen F2R und FA dazwischen liegen.

Zur Prüfung der entstandenen Reihenfolge wurden die Korrelationskoeffizienten auf signifikante Unterschiede untersucht. Es ergab sich, dass sich die beiden Korrelationen FA mit F2R und FA mit GL nicht unterscheiden. Gleichungen und F2R korrelieren allerdings signifikant stärker miteinander als jeweils mit FA.

Wie bereits in Kapitel 3.1 beschrieben, hatten Bartok (2015) und Schnait (2015) ebenfalls den gleichen Widerspruch gefunden. Während sich die beiden Korrelationskoeffizienten (Bartok, 2015: $r_s(\text{F2R}, \text{GL}) = 0,415$; Schnait, 2015: $r(\text{GL}, \text{FA}) = 0,33$) aus unabhängigen Stichproben nicht signifikant unterschieden, zeigt sich unter Berücksichtigung von $r_{23} = r(\text{F2R}, \text{FA})$ und einer insgesamt größeren Stichprobe ein signifikanter Unterschied.

Die neue Erkenntnis der vorliegenden Studie ist folglich, dass F2R und Gleichungen einen signifikant stärkeren Zusammenhang aufweisen als auf den ersten Blick allein durch die theoretische Fundierung schlüssig scheint.

Bartok (2015) hatte bereits einen theoretischen Ansatz zur Erklärung aufgestellt, wonach das Lösen von Gleichungen ab einem gewissen Alter erlernt ist und sich nur noch wenig verändert. Daraus wurde gefolgert, dass die Leistung in Gleichungen stärker von fluider Intelligenz abhängen könnte.

An diese Überlegung soll hier angeknüpft werden. Gelegenheitsbeobachtungen bei den Testungen deuten darauf hin, dass die Aufgabenstellung im Verfahren Gleichungen durchaus auch verstärkt fluide Intelligenz benötigen könnte. Mehrere Personen versuchten die

Gleichung mathematisch aufzulösen, was bei zwei Unbekannten und einer Gleichung unmöglich ist. Die Verwendung von numerischem Material, sprich Gleichungen, ist normalerweise eine andere als im verwendeten Test. So lernt man im Laufe der Schulbildung Gleichungen umzuformen um zur Lösung zu gelangen, dies ist hier allerdings nicht möglich.

Die Lösung durch logische Schlussfolgerung aus mehreren vorgegebenen Möglichkeiten zu erschließen, stellt eine neuartige Aufgabenstellung mit Gleichungen dar, was sich mit der Definition des Einsatzgebietes fluider Intelligenz deckt. Dadurch bedingt könnten verstärkt fluide Anteile in die Leistung in Gleichungen einfließen und den höheren Korrelationskoeffizienten mit dem fluiden Reasoning-Test F2R erklären. Mit 12,6 % erklärter Varianz sind die Tests dennoch größtenteils unabhängig. Die restlichen Korrelationen zwischen den Verfahren sind allesamt schwächer, die Tests sind demnach nicht austauschbar und erfassen, wie beabsichtigt, unterschiedliche Konstrukte bzw. die gleichen Konstrukte auf unterschiedliche Weise.

Als Limitation der Studie ist die Stichprobe anzusprechen, welche nicht zufällig aus der Grundgesamtheit gezogen wurde. Sie genügt damit nicht den Anforderungen der wissenschaftlichen Psychologie und es besteht die Gefahr nicht nur zufälliger, sondern systematischer Verzerrungen und Fehler. Die Akquirierung von Stichproben stellt grundsätzlich ein Problem dar, welches nur schwer zu lösen ist.

Hinsichtlich des zeitlichen sowie ökonomischen Aufwandes der Testbatterie kam realistisch betrachtet nur die Gruppenvorgabe in Frage. Dadurch beschränkte sich die mögliche (Zufalls-)Auswahl auf Schulen oder ähnliche Einrichtungen. Solcher Art Institutionen besitzen jedoch durchgehend eigenständige Lehraufträge und Interessen. Die bürokratische Genehmigung von Testungen, insbesondere der Bereitstellung der notwendigen Zeit, ist nicht selbstverständlich. Zufälliges Ziehen von Einrichtungen würde dies noch zusätzlich erschweren.

Des Weiteren ist die verwendete Stichprobe, bestehend aus Auszubildenden der Sicherheitsakademie, nicht repräsentativ für die Grundgesamtheit. Geschlecht (~70% m, ~30% w) sowie Muttersprache (94,5% deutsch) entsprechen nicht den Quoten in der Gesamtbevölkerung.

Die Gruppenvorgabe der Testbatterie ist ebenfalls anzusprechen. Während mehrfach auf die Freiwilligkeit der Teilnahme hingewiesen wurde, nahmen lediglich 6 von 333 (1,8%) Personen nicht teil. Dies scheint verglichen mit anderen Studien mit freiwilliger Teilnahme

sehr gering. Begründet sein könnte dies durch hohe intrinsische Motivation und Interesse der StudienteilnehmerInnen. Dafür spricht, dass etwa ein Sechstel der Stichprobe (56 Personen) das Angebot bezüglich Rückmeldung ihrer Testergebnisse in Anspruch nahm. Zusätzlich stützt die Gelegenheitsbeobachtung der durchgehend konzentrierten Bearbeitung der Testbatterie diese Überlegung.

5. Zusammenfassung

Im Rahmen dieser Arbeit wurde die Konformität mit dem Rasch-Modell und damit die Eindimensionalität des Verfahrens Figurales Zweidimensionales Reasoning überprüft. A posteriori konnte Modellgültigkeit durch den Ausschluss von drei Items erreicht werden. Aufgrund der verschiedenen Testformen mussten zwei Items wegen Auffälligkeit in den grafischen Modellkontrollen entfernt werden. Die Erstellung paralleler Testformen durch Veränderung der gesuchten Zelle scheint in bestimmten Fällen die Schwierigkeit zu beeinflussen.

Unter Verwendung normaler Teilungskriterien reichte der Ausschluss eines Items um Modellgültigkeit zu erzielen. Die Einschätzung der Testautorin, „dass das Material grundsätzlich eine Eignung bezüglich einer Testzusammenstellung aufweist (Bartok, 2015, S.70)“ kann daher gestützt werden.

Neben dem Verfahren F2R, welches Reasoning im fluiden Bereich erfassen soll, wurden zwei weitere Tests zur kristallisierten Intelligenz eingesetzt, namentlich Gleichungen mit numerischem Material, und Figurale Analogien mit figuralem Material. Wider der theoriegeleiteten Erwartungen zeigte sich der Zusammenhang zwischen Gleichungen und F2R $r_s = 0,355$ signifikant stärker als die restlichen Korrelationen (F2R x FA: $r_s = 0,171$; $p = 0,001$; GL x FA: $r_s = 0,222$; $p = 0,000$).

Als Erklärung scheint denkbar, dass Gleichungen ebenfalls fluide Aspekte benötigt und dadurch der höhere Korrelationskoeffizient mit dem fluiden Reasoning-Test F2R resultiert. Bedingt sein könnte dies durch die ungewohnte Aufgabenstellung mit Gleichungen. Die Lösung aus gegebenen Möglichkeiten zu erschließen widerspricht dem erlernten und gewohnten Auflösen zum Erhalten der Unbekannten.

Insgesamt kann die diskriminante Validität der Verfahren als gegeben betrachtet werden. Selbst der Stärkste der drei Zusammenhänge erklärt lediglich knapp 13% der Varianz. Die Verfahren unterscheiden sich demnach, wie gewollt, in den zu erfassenden Konstrukten bzw. erfassen sie auf verschiedene Art und Weise, sodass sie nicht austauschbar sind.

5.1. Literaturverzeichnis

- AMS. (2016). Berufsinformationssystem. Zugriff am 20.05.2016 unter <http://www.ams.at/bis/bis/Kompetenzen.php?noteid=19>
- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123–140.
- Bartok, L. (2015). Figurales Zweidimensionales Reasoning - Erstellung eines Tests zur Messung des fluiden schlussfolgernden Denkens mit Hilfe figuralen Materials. Diplomarbeit, Universität Wien.
- Cattell, R. B. (1963). Theory of fluid and cristallized intelligence: a critical experiment. *Journal of educational psychology*, 54(1), 1–22.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Elsevier.
- Champely, S. (2015). *pwr: Basic Functions for Power Analysis*. R package version 1.1-3. <http://CRAN.R-project.org/package=pwr>
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Hans Huber.
- Fischer, G. H. (1995). Derivations of the Rasch Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models* (p. 15 – 38). New York: Springer.
- Fischer, G. H. & Scheiblechner, H. H. (1970). Algorithmen und Programmen für das probabilistische Testmodel von Rasch. *Psychologische Beiträge*, 12, 23–51.
- Gamsjäger, C. (2012). Entwicklung des Tests „Gleichungen“ zur Erfassung von Reasoning. Diplomarbeit, Universität Wien.
- Gashi, A. (2011). Untersuchung über Beziehungen zwischen Anstrengungsvermeidung, Selbststeuerung und Lernemotionen bei Jugendlichen aus psychologischer Sicht. Diplomarbeit, Universität Wien.

-
- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of educational psychology*, 57(5), 253–270.
- IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.
- Kubinger, K. D. (Hrsg.) (1989). *Moderne Testtheorie. Ein Abriß samt neuesten Beiträgen*. München: PVU.
- Kubinger, K. D. (2003). Probabilistische Testtheorie. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415–423). Weinheim: Beltz.
- Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2. Aufl.). Göttingen: Hogrefe.
- Kubinger, K. D. & Bartok, L. (in Vorb.) *Figurales Zweidimensionales Reasoning*.
- Kubinger, K. D. & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell. *Diagnostica*, 53 (3), 131 - 143.
- Kubinger, K. D. & Gamsjäger, C. (in Vorb.) *Gleichungen*.
- Kubinger, K. D., Rasch, D. & Yanagida, T. (2009). On designing data-sampling for Rasch model calibrating an achievement test. *Psychology Science Quarterly*, 51(4), 370–384.
- 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), 321–333.
- Mair, P., Hatzinger, R. & Maier M. J. (2015). *eRm: Extended Rasch Modeling*. 0.15-6. <http://erm.r-forge.r-project.org/>

-
- MedUni Wien. (2016). Neue Aufnahmetests für Medizinstudium ab 2013. Zugriff am 20.05.16 unter http://www.meduniwien.ac.at/homepage/3/news-und-topstories/?tx_ttnews%5Btt_news%5D=3142&cHash=b54a257cb40eba4700051d72cac02265
- Moosbrugger, H. & Hartig, J. (2003). Klassische Testtheorie. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 408–415). Weinheim: Beltz.
- RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. URL <http://www.rstudio.com/>.
- R Core Team (2015a). *foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase,* R package version 0.8-66. <http://CRAN.R-project.org/package=foreign>
- R Core Team (2015b). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: Paedagogiske Institut.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Exp. ed.). University of Chicago, Chicago.
- Revelle, W. (2015) *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, Illinois, USA, <http://CRAN.R-project.org/package=psych>, Version = 1.5.8.
- Rollett, B. & Bartram, M. (1998). *Anstrengungsvermeidungstest (AVT)*. (3.Auflage). Göttingen: Hogrefe Verlag.

-
- Schnait, B. (2015). Zur Dimensionalität von Reasoning-Tests: Ein Beitrag zur Validierung einer neuen Testbatterie zur Erfassung des logisch-schlussfolgernden Denkens. Diplomarbeit, Universität Wien.
- Statistik Austria. (2016a). Bevölkerung nach Alter und Geschlecht. Zugriff am 07.05.2016 unter http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/bevoelkerungsstruktur/bevoelkerung_nach_alter_geschlecht/index.html
- Statistik Austria. (2016b). Bevölkerung nach Staatsangehörigkeit und Geburtsland. Zugriff am 07.05.2016 unter http://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/bevoelkerungsstruktur/bevoelkerung_nach_staatsangehoerigkeit_geburtsland/031396.html
- Steiger, J. H. (1980). Tests for Comparing Elements of a Correlation Matrix. *Psychological Bulletin*, 87(2), 245–251.
- Stemmler, G., Hagemann, D., Amelang, M. & Bartussek, D. (2010). *Differentielle Psychologie und Persönlichkeitsforschung* (7., neu bearbeitete Aufl.). Stuttgart: Kohlhammer.
- Süß, H.-M. (2003a). Culture fair. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 82–86). Weinheim: Beltz.
- Süß, H.-M. (2003b). Intelligenztheorien. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 217–224). Weinheim: Beltz.
- Thurstone, L. L. (1938). *Primary mental abilities*. University of Chicago, Chicago.
- Ünal, S. (2014). Figurale Analogien: ein Test zur Messung des kristallisierten schlussfolgernden Denkens mit Hilfe figuralen Materials. Diplomarbeit, Universität Wien.

Ünal, S. & Kubinger, K. D. (in Vorb.) *Figurale Analogien*.

Vanecek, E. (2003). Geschichte der Psychologischen Diagnostik. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 175–181). Weinheim: Beltz.

Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, 54(3), 426–482. <http://doi.org/10.2307/1990256>

Zimbardo, P. G. & Gerrig, R. J. (2008). *Psychologie* (18. Aufl.). München: Pearson Studium.

5.2. Abbildungsverzeichnis

Abbildung 1: Instruktionsitem <i>Gleichungen</i> (Gamsjäger, 2012, S. 31).....	19
Abbildung 2: Instruktionsitem <i>F2R</i> (Bartok, 2015, S. 31).....	20
Abbildung 3: Instruktionsitem <i>Figurale Analogien</i> (Ünal, 2012, S. 44)	21
Abbildung 4: Verteilung der Rohwerte im F2R (vor den Rasch-Modell-Analysen).....	28
Abbildung 5: Verteilung der AVT Rohwerte.....	28
Abbildung 6: Grafische Modellkontrolle für das Teilungskriterium Score, alle Items	31
Abbildung 7: Grafische Modellkontrolle für das Teilungskriterium Testform, alle Items	32
Abbildung 8: Grafische Modellkontrolle für das Teilungskriterium Ausbildung, alle Items ..	32
Abbildung 9: Grafische Modellkontrolle für das Teilungskriterium Alter, alle Items	33
Abbildung 10: Grafische Modellkontrolle für das Teilungskriterium Score AVT, alle Items	33
Abbildung 11: Item 14, Testform A, als „einfacher“ geschätzt	34
Abbildung 12: Item 14, Testform B, als „schwieriger“ geschätzt	34
Abbildung 13: Grafische Modellkontrolle für das Teilungskriterium Score, nach Ausschluss von Item 14.....	36
Abbildung 14: Grafische Modellkontrolle für das Teilungskriterium Testform, nach Ausschluss von Item 14.....	36
Abbildung 15: Item 15, Testform A, als „schwieriger“ geschätzt	37
Abbildung 16: Item 15, Testform B, als „einfacher“ geschätzt	37
Abbildung 17: Grafische Modellkontrolle für das Teilungskriterium Testform, nach Ausschluss von Item 14 und 15.....	38
Abbildung 18: Grafische Modellkontrolle für das Teilungskriterium Score, nach Ausschluss von Item 14 und 15.....	39
Abbildung 19: Item 16, Testform A, als „schwieriger“ für Score F2R > Median geschätzt ...	39
Abbildung 20: Item 16, Testform B, als „schwieriger“ für Score F2R > Median geschätzt ...	40
Abbildung 21: Grafische Modellkontrolle für das Teilungskriterium Score, nach Ausschluss von Item 14, 15 und 16.....	41
Abbildung 22: Grafische Modellkontrolle für das Teilungskriterium Score, nach Ausschluss von Item 16 bei vier Teilungskriterien.....	42

5.3. Tabellenverzeichnis

Tabelle 1: Überblick über die Konstrukte und Darbietungsform der Verfahren	23
Tabelle 2: Theoriegeleitete, erwartete Rangfolge der Korrelationen zwischen den Verfahren	24
Tabelle 3: Itemanzahl und Zeitbeschränkungen der Testbatterie	26
Tabelle 4: Häufigkeiten der höchsten abgeschlossenen Ausbildung	27
Tabelle 5: Korrelationsmatrix der zu verwendenden Teilungskriterien	30
Tabelle 6: F2R – LRT für alle Items	31
Tabelle 7: F2R – LRT nach Ausschluss von Item 14	35
Tabelle 8: F2R – LRT nach Ausschluss von Item 14 und Item 15	38
Tabelle 9: F2R – LRT nach Ausschluss von Item 14, Item 15 und Item 16	40
Tabelle 10: F2R – LRT mit 4 Teilungskriterien, nach Ausschluss von Item 16	42
Tabelle 11: Erwartete und empirisch gefundene Reihung der Korrelationen zwischen den Verfahren	43
Tabelle 12: Steiger Tests auf Unterschiede zwischen abhängigen Korrelationskoeffizienten.	44
Tabelle 13: Ereignisbasiertes Risiko 2. Art für die Zusammenhänge zwischen den Verfahren	45

A. Ansuchen an das Bundesministerium für Inneres

Universität Wien
Fakultät für Psychologie
Institut für Angewandte Psychologie:
Gesundheit, Entwicklung und Förderung
Liebiggasse 5
1010 Wien

Bundesministerium für Inneres
Sicherheitsakademie
MinRat Mag. Dr. Norbert Leitner
Bildungszentrum Wien
Marokkanergasse 4
1030 Wien

Wien, am 24.11.15

Ansuchen um Genehmigung einer Datenerhebung im Rahmen einer Masterarbeit:
Arbeitstitel: Tripelweiser Dimensionalitätsvergleich der Reasoning-Tests „Gleichungen“,
„Figurale Analogien“ und „Figurales Zweidimensionales Reasoning“

Sehr geehrter Herr Dr. Leitner,

Ich studiere Psychologie an der Universität Wien und schreibe meine Masterarbeit am Institut für Angewandte Psychologie: Gesundheit, Entwicklung und Förderung, unter der Leitung von Univ.-Prof. Dr. Mag. Klaus D. Kubinger. In meiner Masterarbeit befasse ich mich mit einem Dimensionalitätsvergleich mehrerer Tests zur Erfassung von schlussfolgerndem Denken. Die Erhebung umfasst drei verschiedene Verfahren:

1. „Gleichungen“ (Kubinger & Gamsjäger, in Vorb.),
 2. „Figurale Analogien“ (Ünal & Kubinger, in Vorb.),
 3. „Figurales Zweidimensionales Reasoning“ (F2R: Kubinger & Bartok, in Vorb.),
- sowie einen kurzen Fragebogen zur Motivation (AVT: Rollett & Rollett, 2000).

Ziel der Untersuchung ist es zu prüfen wie gut die jeweiligen Verfahren die zu erfassenden Fähigkeiten messen und inwiefern sie sich ähneln bzw. unterscheiden. Dadurch wird ein Beitrag zur Prüfung der Validität der Verfahren geleistet. In Folge soll damit die Qualität der psychologischen Leistungsdiagnostik gesichert werden, welche bei Schullaufbahn-, Bildungs- und Studienwahlberatung sowie bei der Eignungsdiagnostik wesentlicher Bestandteil ist.

Um entsprechende statistische Analysen durchführen zu können, sollen insgesamt Daten von 250 – 300 Personen ab 16 Jahren erhoben werden. 2012 hat Frau Ünal am Bildungszentrum Traiskirchen Personen getestet und damit das Verfahren „Figurale Analogien“ erstellt. Ich wende mich daher wieder an die Sicherheitsakademie, da sich die Studien thematisch überschneiden.

Auch diesmal besteht die Möglichkeit für die TeilnehmerInnen **Rückmeldung über ihre Fähigkeiten im logisch-schlussfolgernden Denken** zu erhalten.

Sofern Sie eine offizielle Genehmigung für die Testungen geben, würde man sich bei der Festlegung von Testterminen nach den jeweiligen Bildungszentren richten, die teilnehmen möchten. Durchgeführt werden soll die Testung (50 min) in Gruppen, in Räumlichkeiten der Bildungszentren. Ich bin bemüht, den Aufwand für die Standorte gering zu halten und den Unterricht möglichst wenig zu stören.

Die teilnehmenden Standorte leisten einen wesentlichen Betrag zur Sicherung der Qualität psychologischer Diagnostik, welche für die Beratung von SchülerInnen und Auszubildenden hinsichtlich leistungsbezogener Fähigkeiten eingesetzt werden kann. Dies wirkt sich auch positiv auf den Bildungsbereich im Allgemeinen aus.

Im Anhang finden Sie das vollständige Ansuchen sowie eine genaue Beschreibung der Verfahren, jeweils mit Beispielen. Dabei sind aus Copyright-Gründen nicht alle Aufgaben angeführt. Es ist aber zu betonen, dass sich die restlichen Aufgaben konzeptionell nicht von den Beispielitems unterscheiden.

Mit höflicher Bitte um Unterstützung der Studie,

Benedikt Winter
Masterand an der Fakultät für Psychologie
Universität Wien

Anlagen:

Ansprechpartner
Untersuchungsablauf
Informationen zu den Verfahren
Informationsblatt zur Testung