

# MASTERARBEIT / MASTERTHESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Dimensionalitätsvergleich der Reasoning-Tests Figurale Analogien,  
Family Relation Reasoning Test und Extended Syllogismen“

verfasst von / submitted by

Kerstin Grafl, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

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

UA 066 840

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

Masterstudium Psychologie

Betreut von / Supervisor:

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

# **Danksagung**

Ich möchte mich auf diesem Weg bei all jenen Menschen bedanken, die mich in den letzten Jahren begleitet, unterstützt und inspiriert haben.

Danken möchte ich in erster Linie meinem Betreuer, Herrn Univ.-Prof. Klaus Kubinger, für seine fachliche und organisatorische Unterstützung, sowie seine Geduld, Wissens- und Erfahrungsschatz, wovon ich sehr viel lernen konnte.

Ich bedanke mich auch bei Frau Fuchs-Linka MSc, HCM-Generaldirektion der Unternehmung Wiener Krankenanstalten-Verbund, für die Genehmigung zur Durchführung der Testungen in den Schulen für Gesundheits- und Krankenpflege, sowie bei Frau Mag<sup>a</sup>. Monika Lukas, Direktorin Wilhelminenspital Wien und Frau Michaela Dorfmeister, MBA-Schuldirektorin der Gesundheits- und Krankenpflegeschule des Sozialmedizinischen Zentrum Ost Wien.

Ebenfalls bedanke ich mich bei den zahlreichen Schülerinnen und Schülern der Gesundheits- und Krankenpflegeschulen welche, trotz laufenden Unterrichts, so motiviert und freiwillig an der Bearbeitung der Testhefte arbeiteten.

Besonderer Dank gilt meiner Familie, meiner Mutter Ingeborg Grafl, meiner Tante Rita Neubauer und meinen beiden Geschwistern Gregor Grafl und Jasmin Grafl für ihre Unterstützung und Liebe.

Ein herzliches Dankeschön gilt auch meinen Kolleginnen und Kollegen sowie FreundInnen Isabella Vogh, BSc., Sanja Djordevic Msc., Benedikt Winter, MSc., Michaela Jelinek, Verena Szelgrad, Brigitte Meitz und Mag<sup>a</sup>. Bianca Silianof, für ihre Unterstützung, Freundschaft und Inspiration.

## Abstract (Deutsch)

In der vorliegenden Arbeit wurde die Konformität des kristallinen Family-Relation-Reasoning-Tests (FRRT) und des fluid-lexikalischen Extended Syllogismen Tests (ES) mit dem Rasch-Modell überprüft. Die Daten wurden im Zeitraum März bis Juni 2015 anhand einer zusammengestellten Testbatterie in Gesundheits- und Krankenpflegeschulen des Wiener Krankenanstalten-Verbunds im Gruppensetting erhoben. Das Ziel war einerseits die Eindimensionalität der Items einer Prüfung zu unterziehen und andererseits die diskriminante Validität, unter Berücksichtigung des Figuralen-Analogie-Tests (FA), darzulegen. Zusätzlich wurde der Anstrengungs-Vermeidungs-Test (AVT) für Schüler als externes Teilungskriterium herangezogen. Als weitere Teilungskriterien fungierten Alter, Ausbildung, Geschlecht und Erstsprache. Die Ergebnisse wurden einerseits deskriptivstatistisch und andererseits inferenzstatistisch anhand von grafischen Modellkontrollen sowie mittels Korrelationskoeffizienten dargestellt. Während die Modellgültigkeit im FRRT ohne Itemselektion gezeigt werden konnte, waren im Rahmen des ES vier Items auszuschließen. Der ES besitzt danach die Eigenschaft der Rasch-Modell-Konformität. Es wurden, neben der Erhebung der fluiden Intelligenz durch den ES (lexikalisches Aufgabenmaterial), auch die kristalline Intelligenz über den FRRT (lexikalisch) und die „Figurale Analogien“ (figural) erfasst.

Die Korrelation zwischen ES und FRRT  $r_s = .375$  erwies sich etwas stärker als zu erwarten war. Demnach kann ein fluider Lösungsprozess-Anteil im FRRT angenommen werden. Allerdings erklärt selbst der stärkste dieser Zusammenhänge nur 14% der Varianz der Leistungen im FRRT, somit konnte gezeigt werden, dass die drei Verfahren eine diskriminante Validität aufweisen: Der zuvor festgelegte Grenzwert der Korrelationsstärken von  $r_s = .40$  wurde nur ansatzweise erreicht. Anhand der festgestellten Reihung  $r_s$  (FRRT, ES)  $.375$ ;  $r_s$  (FA, ES)  $.350$ ;  $r_s$  (FA, FRRT)  $.298$  konnte die zu erwartende Hierarchie der Korrelationen zwischen den Verfahren wie vorab postuliert  $r_s$  (FA, FRRT)  $>$   $r_s$  (FRRT, ES)  $>$   $r_s$  (FA, ES), nicht beobachtet werden.

*Schlüsselwörter:* Schlussfolgerndes Denken, Reasoning, fluide Intelligenz, kristallisierte Intelligenz, Rasch-Modell, diskriminante Validität, *Figurale Analogien*, *Family-Relation-Reasoning-Test*, *Extended Syllogisms*, *Anstrengungs-Vermeidungs-Test*

## Abstract (English)

In the present work the conformity of the crystalline Family-Relation-Reasoning-Test (FRRT) and the fluid-lexical Extended Syllogism Test (ES) with the Rasch model was evaluated. The data were collected from March to June 2015 using a test battery in health and nursing schools of the Wiener Krankenanstaltenverbund in a group setting. The aim was to determine the one-dimensionality of the scale items on the one hand and to demonstrate the discriminant validity, taking into account the figurative analogy test (FA), on the other hand. In addition, the AVT for students was used as an external split criterion. Further criteria for splitting were age, education, gender and first language. The results were presented descriptively statistically and inferentially statistically using graphical model tests and correlation coefficients. While the validity of the model could be shown in the FRRT without item selection, four items had to be excluded in the ES in consideration of the division criteria. Thus, the ES has the property of Rasch model conformity.

In addition to the collection of fluid intelligence by the ES (lexical task material), crystalline intelligence was also collected via the FRRT (lexical) and the FA (figural).

The correlation between ES and FRRT  $r_s = .375$  proved to be somewhat stronger than expected. Thus, a fluid solution process part can be assumed in the FRRT. However, even the strongest of these correlations explains only 14% of the variance of the performances in the FRRT, thus it could be shown that the three methods have a discriminant validity. The previously defined limit of correlation strengths of  $r_s = .40$  is only rudimentarily reached. On the basis of the established order  $r_s$  (FRRT, ES) .375;  $r_s$  (FA, ES) .350;  $r_s$  (FA, FRRT) .298, the expected hierarchy of correlations between the procedures, as previously postulated  $r_s$  (FA, FRRT) >  $r_s$  (FRRT, ES) >  $r_s$  (FA, ES), could not be observed.

Keywords: reasoning, reasoning, fluid intelligence, crystallized intelligence, Rasch modelling, discriminant validity, *Figural Analogies*, *Family Relation Reasoning Test*, *Extended Syllogisms*, *Anstrengungs-Vermeidungs-Test*

# Inhaltsverzeichnis

<b>Danksagung .....</b>	<b>1</b>
<b>Abstract (Deutsch).....</b>	<b>2</b>
<b>Abstract (English) .....</b>	<b>3</b>
<b>Inhaltsverzeichnis .....</b>	<b>4</b>
<b>Einleitung.....</b>	<b>5</b>
<b>1      Theoretischer Hintergrund .....</b>	<b>6</b>
1.1 Intelligenztheorien.....	8
1.1.1 Theorie der fluiden und kristallisierten Intelligenz.....	9
1.1.2 Logisch-schlussfolgerndes Denken (Reasoning).....	10
1.2 Das dichotom-logistische Modell von Rasch .....	11
<b>2      Verfahren.....</b>	<b>16</b>
2.1 Family-Relation-Reasoning-Test.....	16
2.2 Figurale Analogien.....	17
2.3 Extended Syllogismen.....	17
2.3.1 Anstrengungsvermeidungstest .....	18
<b>3      Empirischer Teil.....</b>	<b>20</b>
3.1 Hypothesen .....	20
3.2 Methode .....	20
3.2.1 Planung und Durchführung der Untersuchung .....	21
3.2.2 Stichprobenbeschreibung.....	22
<b>4      Ergebnisse.....</b>	<b>25</b>
4.1 Analysen zur Überprüfung der Gültigkeit des Raschmodells .....	25
4.1.1 FRRT.....	25
4.1.2 Ergebnisse der Rasch-Modell-Analysen des Extended Syllogismen Test .....	30
4.2 Korrelative Zusammenhänge der Reasoning-Tests.....	42
<b>5      Diskussion und Ausblick .....</b>	<b>44</b>
<b>6      Zusammenfassung.....</b>	<b>46</b>
<b>7      Literaturverzeichnis.....</b>	<b>47</b>
<b>8      Abbildungsverzeichnis.....</b>	<b>50</b>
<b>9      Tabellenverzeichnis.....</b>	<b>51</b>
<b>10     Abkürzungsverzeichnis.....</b>	<b>53</b>
<b>11     Anhang.....</b>	<b>54</b>

## Einleitung

Die vorliegende Arbeit umfasst den Vergleich psychologisch-diagnostischer Verfahren, die schlussfolgerndes Denken erfassen. Diese Fähigkeit – in der englischen Literatur auch als *Reasoning* bezeichnet – ist ein bedeutender Faktor der menschlichen Intelligenz und wird als eine fundamentale Denkfunktion erachtet. Reasoning wird in zahlreichen Intelligenztestbatterien und spezifischen psychologischen Tests ermittelt (Jäger, 1973). Diese Verfahren finden häufig ihren Einsatz im Rahmen von Auswahlverfahren, Berufs- und Bildungsberatungen oder im diagnostischen Bereich bei Leistungs- und Lernproblemen.

Mit zunehmender Informationsvielfalt fluten viele Reize und Informationen auf den menschlichen Geist ein. Eine hohe Fähigkeit im logisch-schlussfolgernden Denken, im kristallisierten Bereich der Intelligenz, kann helfen, die Vielfalt an Informationen und Wissen ins Bewusstsein zu integrieren, Zusammenhänge schneller zu erkennen, komplexe Problemstellungen zu lösen und aus Erfahrungen zu lernen.

In den letzten Jahren wurden am Institut für Psychologie der Universität Wien unter der Leitung von Herrn Univ.-Prof. Dr. Mag. Klaus Kubinger neuartige Verfahren zur Messung von logisch-schlussfolgerndem Denken, sowie der kristallisierten Intelligenz entwickelt. Durch entsprechende Studien sollen die in Entwicklung stehenden Verfahren auf Eindimensionalität überprüft und ein Beitrag zur Validierung geleistet werden.

In der gegenständlichen Arbeit wird die Dimensionalität von neu entwickelten Verfahren, die das logisch-schlussfolgernde Denken im kristallisierten und fluiden Bereich messen, verglichen. Die Verfahren *Family-Relation-Reasoning* und *Extended Syllogismen* werden auf Konformität mit dem dichotom-logistischen Testmodell nach Rasch überprüft. Im Anschluss wird die konvergente Validität dieser Verfahren, unter Berücksichtigung des Verfahrens *Figurale Analogien*, analysiert.

Der Theorieteil bietet einen Überblick zu den Intelligenztheorien und Modellen, die dem logisch-schlussfolgernden Denken zugrunde liegen. Testtheoretische Aspekte, welche für die Untersuchung der Verfahren relevant sind, werden ebenfalls dargestellt.

Im Methodenteil wird auf die Verfahren, welche zur Anwendung gelangten, differenziert eingegangen. Im Weiteren werden die Datenerhebung und Stichprobe beschrieben. Im empirischen Teil erfolgt die Darlegung zu den Ergebnissen der Rasch-Modell-Analysen sowie zu den Zusammenhängen der angewendeten Verfahren.

# 1 Theoretischer Hintergrund

Schlussfolgerndes Denken – Reasoning – ist ein Teilaspekt des theoretischen Konstrukts der Intelligenz. Da es verschiedene Blickwinkel auf Intelligenz gibt, werden im Folgenden verschiedene Intelligenzdefinitionen und -theorien erläutert und philosophische Aspekte beleuchtet. Intelligenz ist kein direkt beobachtbares Merkmal und kann nur indirekt in Leistungssituationen über das Verhalten gemessen werden (Süß, 2003). Alltagspsychologischer Auffassung nach ist Intelligenz eine relativ einheitliche Fähigkeit, jedoch ist im wissenschaftlichen Sinne differenziert darzulegen, was unter intellektuellen Leistungen zu verstehen ist.

Der Begründer der Differentiellen Psychologie, William Stern (1912), bezeichnete Intelligenz als eine allgemeine Fähigkeit des Individuums, sein Denken bewusst auf neue Forderungen einzustellen. Intelligenz wird demnach als allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens verstanden.

Wechsler (1956) definierte Intelligenz als die Fähigkeit, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen.

Gemäß Anastasi (1958) messen Intelligenztests nur die Fähigkeit, in der jeweiligen, speziellen Kultur erfolgreich zu sein. Intelligenz sei die Fähigkeit zur Auffindung von Ordnung in der Welt (Hofstätter, 1977). All diesen Definitionen ist gemeinsam, dass sie einen geringen Erklärungswert aufweisen, wenn die Begriffe vernünftig, zweckvoll usw. durch das Wort „intelligent“ ersetzt werden. Die Definition wird demnach zu einer sinnfreien Tautologie (Hofstätter, 1977). Erste wissenschaftliche Ansätze zum schlussfolgernden Denken können den Philosophen der griechischen Antike und den Formallogikern zugeschrieben werden (Schüling, 2002). Es wurde zwischen synonymen Begriffen, wie Logik (von altgriechisch *logos*), Vernunft und Geist, differenziert. Als Begründer der *Syllogistik* kann Aristoteles mit seiner Lehre der formalen Folgerichtigkeit des Denkens angeführt werden.

*Kant* wiederum beschrieb die transzendente Logik, die Lehre von den erfahrungsunabhängigen, vernunftgemäßen Bestimmungen des Verstandes. Die transzendentalen Bestimmungen des Verstandes mit den gegebenen Bedingungen des Seienden versuchte Hegel mit der dialektischen Logik in Einklang zu bringen.

Ein Teilgebiet der Philosophie, die formale Logik, beschäftigt sich mit Schlussfolgerungsprozessen. Sie untersucht die streng logischen Regeln des schlussfolgernden Denkens und blendet die menschlichen Komponenten dabei aus. Im Allgemeinen wird unter Logik die formale Logik verstanden, in gewissem Maße als nachträgliche Reflexion. Logik

gründet sich nicht auf der Empirie und daher müssen Sätze der Logik keinen sinnvollen Inhalt haben. In diesem Sinne wird mit jener Form des Denkens, als „negativer Proberstein der Wirklichkeit“, umgegangen (Anzenbacher, 1985).

Die wissenschaftliche Psychologie integriert neben der formalen Logik auch die menschliche Informationsverarbeitung (Beckmann & Guthke, 1999). Aus dem verfügbaren Testinventar leitete Kubinger (2009, S.206) die Definition von Reasoning als die „Fähigkeit, Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge zu erkennen und zweckentsprechend verwerten zu können“ ab. Schlussfolgerndes Denken lässt sich demnach in induktives und deduktives Denken unterteilen. Die Bedeutung von schlussfolgerndem Denken liegt darin, dass aus einer gegebenen Information neues Wissen abgeleitet werden kann. Dies kann durch induktives Vorgehen geschehen, indem man vom Besonderen auf das Allgemeine schließt, oder auf deduktive Weise, wo vom Allgemeinen auf das Besondere geschlossen wird. Beim induktiven Vorgehen führt die Kenntnis einzelner Fakten zu einer Verallgemeinerung, jedoch muss solch eine Verallgemeinerung nicht unbedingt für alle möglichen Bedingungen gelten. Kant benannte den Induktionsschluss auch als Wahrscheinlichkeitsschluss (Beckmann & Guthke, 1999).

Induktives Denken wird auch als fundamentale Leistung der menschlichen Intelligenz gesehen, aus vergangen Erfahrungen zu lernen, indem aus bestimmten Erlebnissen Gesetzmäßigkeiten, Muster und allgemeine Regeln abgeleitet werden können (Waldmann & Weinert, 1990). Prinzipiell können zwei Pfade beim schlussfolgernden Denken unterschieden werden. Die Abbildung 1 zeigt schematisch die beiden Wege, Induktion und Deduktion.

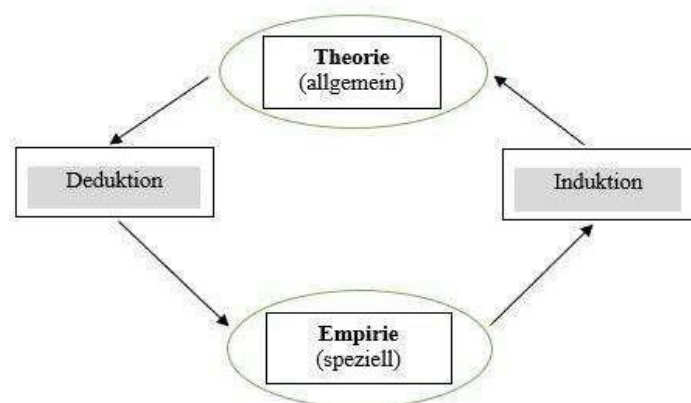


Abbildung 1. Wege der Erkenntnisgewinnung (modifizierte Darstellung)



Beim deduktiven Schließen des schlussfolgernden Denkens steht die logische Gültigkeit, auch als logisches Schließen bezeichnet, im Vordergrund. Logisches Schließen umfasst demnach, dass sich aus vorgegeben Fakten (Prämissen) eine Schlussfolgerung (Konklusion) zwingend (notwendig) ableiten lässt. Brentano (1956) bezeichnet deduktive Schlüsse als „Gewissheitsschlüsse“, da bei der Deduktion keine Erweiterung der Informationsmenge stattfindet und bei korrekter Anwendung der Regeln und Wahrheit der Prämissen auch bei der Schlussfolgerung von der Wahrheit ausgegangen werden kann (Beckmann & Guthke, 1999).

Nach Beckmann und Guthke (1999) ist jedoch die Schlussrichtung zwischen Induktion und Deduktion kritisch zu betrachten. Wenn im Zuge einer vollständigen Induktion, die bereits alle relevanten Einzelfälle in Form von Prämissen für die Schlussfolgerung enthält, keine neuen Informationen generiert werden, lassen sich die daraus gezogenen Schlüsse auch als Deduktion bezeichnen. Die Schlussrichtung sollte demnach als Oberflächenmerkmal gesehen werden und der Charakter der Konklusion (Informationen der Prämissen werden erweitert / nicht erweitert) als Differenzierungsmerkmal verwendet werden. Die Prüfung der Fähigkeit zum induktiven Denken im Rahmen der psychologischen Diagnostik erfolgt nicht selten über Deduktion. Eine Unterscheidung von Aufgabentypen, die schlussfolgerndes Denken erfassen, erscheint daher nicht immer sinnvoll (Beckmann & Guthke, 1999).

## **1.1 Intelligenztheorien**

Psychologische Tests, die zur Messung von Intelligenz eingesetzt werden, müssen hinsichtlich ihrer zugrunde liegenden theoretischen Konzepte und Annahmen differenziert werden, da intelligenz-diagnostische Verfahren, je nach Intelligenzmodell, zumeist nur Teilaspekte der Intelligenz erfassen (Stemmler, Hagemann, Amelang & Bartussek, 2011).

Seit Beginn der Intelligenzforschung ist es umstritten, dass sich die generelle geistige Fähigkeit eines Menschen durch einen übergeordneten Intelligenzwert beurteilen lässt, wie es im Generalfaktorenmodell nach Spearman beschrieben wird. Spearman (1904, 1914, 1927) postulierte nach Ergebnissen von Korrelationsberechnungen und darauffolgenden Faktorenanalysen in seinem Modell, der sogenannten Zwei-Faktoren-Theorie, dass ein Faktor (engl. *g-factor*) allen Leistungsbereichen für allgemeine Intelligenz zugrunde liegt. Dies ist jener Varianzanteil, der alle Fähigkeiten im Bereich der Intelligenz gemeinsam ist. Es wurden weitere, dem *g-factor* untergeordnete, Faktoren extrahiert, die sogenannten spezifischen Faktoren (*s-factors*). Diese Faktoren sind bereichsspezifisch, voneinander unabhängig und

dem g-factor hierarchisch untergeordnet. Der Generalfaktor (g-factor) beeinflusst die spezifischen Faktoren, wie z.B. bei räumlichen, mathematischen oder visuellen Aufgabenstellungen. Charles Spearman war kein Vertreter des Generalfaktorenmodells; er war der Annahme, dass anstelle eines einzigen übergeordneten Faktors mehrere voneinander unabhängige Generalfaktoren existieren, die jeweils einer primären, kognitiven Fähigkeit entsprechen (1932, zitiert nach Thurstone, 1961, S. 351-367; siehe dazu Jäger, 1973; Stemmler et al., 2011; Süß, 2003).

Der Psychologe Louis Leon Thurstone extrahierte, aufbauend auf den Arbeiten von Spearman (Jäger, 1973), sieben Primärfaktoren der Intelligenz (*primary mental abilities*) und fasste die sieben voneinander unabhängigen generellen Faktoren der Intelligenz unter dem Überbegriff der allgemeinen Intelligenz zusammen. Basierend auf Ergebnissen wiederholter Faktorenanalysen postulierte er folgende sieben Primärfaktoren, (Guilford & Hoepfner, 1976; Stemmler et al., 2011; Süß, 2003):

- V (*verbal comprehension*): Wortschatz, verbale Beziehungen verstehen und interpretieren
- R (*reasoning*): Schlussfolgerndes Denken, Fähigkeit zum induktiven und deduktiven Denken
- W (*word fluency*): Wortflüssigkeit
- N (*number*): Fähigkeit zum Umgang mit Zahlen, Rechenfähigkeit
- S (*space*): Räumlich-visuelles Vorstellungsvermögen
- M (*memory*): Gedächtnisleistung, Merkfähigkeit, Kurzzeitgedächtnis
- P (*perceptual speed*): Wahrnehmungsgeschwindigkeit

### **1.1.1 Theorie der fluiden und kristallisierten Intelligenz**

Raymond B. Cattell wandte sich von einer Generalfaktor-Modellvorstellung der Intelligenz ebenfalls ab und veröffentlichte 1957 seine Theorie der fluiden und kristallinen Intelligenz. Das Modell von Cattell führt Aspekte der Zwei-Faktoren-Theorie nach Spearman und des Primärfaktorenmodells gemäß Thurstone zusammen (Cattell, 1957, 1963). Gemäß Cattell (1957, 1963) setzt sich der g-factor nach Spearman aus zwei voneinander abhängigen Faktoren zusammen – der fluiden allgemeinen Intelligenz (engl. *gf-factor*) und der kristallisierten Intelligenz (engl. *gc-factor*).

Cattell entwickelte die Theorie mit seinem Kollegen Horn weiter, mit der Annahme, dass die fluide Intelligenz ihren Höhepunkt bereits im frühen Erwachsenenalter erreicht und sich dann

ein Abbau dieser Fähigkeiten einstellt (Horn & Cattell, 1966). Fluide Intelligenz wird benötigt, um sich in neuen Situationen anpassen zu können und abstrakte Problemstellungen zu lösen. Hierbei helfen vorangegangene Lernerfahrungen nicht. Beeinflusst wird die fluide Intelligenz durch Vererbung oder durch Verletzungen im zentralen Nervensystem. Um fluide Intelligenz erfassen zu können, ist Aufgabenmaterial wie semantische Beziehungen, figurales Schlussfolgern oder Induktion von Nöten (Horn & Cattell, 1966). Dabei erkannte Cattell (1963) bereits die Notwendigkeit von *culture fair Tests* zur Erfassung der fluiden Intelligenz.

Die kristallisierte Intelligenz dagegen entwickelt sich langsamer über die gesamte Lebensspanne, baut nur gering bis gar nicht ab und ist das Produkt vorangegangener Lernerfahrungen in bestimmten Wissensbereichen. Somit ist die kristallisierte Intelligenz auch abhängig von der Umwelt, der Kultur und der Bildung einer Person. Gemäß Cattell (1963) wird die kristallisierte Intelligenz hauptsächlich durch kulturabhängige Gewohnheiten wie auch kulturspezifische Bildungssysteme beeinflusst und kann daher am besten mit kulturspezifischem Aufgabenmaterial erfasst werden wie zum Beispiel mit Sprachverständnis, Rechen-, Wortschatz- und Allgemeinwissenstests (Zimbardo & Gerig, 2008).

In einer Längsschnittstudie konnten die Annahmen Horns und Cattell bezüglich der Entwicklungsverläufe der Faktoren *gf* und *gc* bestätigt werden (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002).

### **1.1.2 Logisch-schlussfolgerndes Denken (Reasoning)**

Die vorliegende Arbeit orientiert sich, wie auch die gängige deutschsprachige Literatur, an den beiden synonym verwendeten Begriffen *logisch-schlussfolgerndes Denken* und *Reasoning*. Das bis heute einflussreich, replizierte Primärfaktoren-Modell stellt eine bedeutende Kernkomponente der „Primärfaktoren“ des Konstrukts Intelligenz gemäß Thurstone (1938) dar und findet sich in zahlreichen Intelligenztests wieder. Reasoning bezeichnet die Fähigkeit, aus vorgegebenem Aufgabenmaterial, Folgerungen bezüglich logischer Zusammenhänge zu erfassen, wie unter Anderem erkennen von Beziehungen zwischen verschiedenen Objekten oder sinnvolles Ergänzen von Mustern (Treiber, 2010).

Reasoning-Aufgaben wird zudem die höchste Augenschein-Validität (*face-validity*) beigemessen und sollen gemäß Beckmann und Guthke (1999, S.19) weder zu bekannt noch allzu neu sein, um die Möglichkeit zu reduzieren, auf Erfahrungen zurückzugreifen. Zudem sollen jene Aufgaben, die induktives und deduktives Schlussfolgern erfordern, die Anwendung der entsprechenden Regeln zur Lösung implizieren. Die Iteminhalte können aus

drei unterschiedlichen Materialbereichen (figural, lexikalisch, numerisch) bestehen, sollen jedoch nicht andere Konstrukte erfassen und demnach hohe Interkorrelationen aufweisen (Beckmann & Guthke, 1999). Gängige Intelligenzmodelle (vgl. z.B. Stemmler et al., 2011) umfassen zumeist eine Dimension für logisch-schlussfolgerndes Denken. Thurstone (1938) beschrieb in seiner Theorie zu den *primary abilities* Induktion bzw. Reasoning als „Fähigkeit zum induktiven und deduktiven Denken“ (Süß, 2003, S. 219). Gemäß Cattell (1963) dienen sowohl kristallisierte als auch fluide Intelligenz dazu, Aufgabenstellungen, die einen logischen Schluss verlangen, erfolgreich zu bearbeiten. Alle Aufgaben und Tests, welche die beschriebene Fähigkeit erfordern, können folglich dem Reasoning zugeordnet werden. Zu unterscheiden ist hierbei, wie bereits angeführt, dass das verwendete Aufgabenmaterial numerisch, figural oder lexikalisch gestaltet sein kann.

In der Psychologischen Diagnostik steht eine große Anzahl an Verfahren zur Messung der Reasoning-Fähigkeit zur Verfügung, wie zum Beispiel der bekannte Wiener Matrizen-Test (WMT; Formann & Piswanger, 1979). Die Anzahl der zur Verfügung stehenden, einschlägigen Verfahren zur Messung des schlussfolgernden Denkens im sprachlichen Bereich ist jedoch begrenzt. Diese kommen vor allem in Intelligenztestbatterien als Untertests vor (Treiber, 2010).

## 1.2 Das dichotom-logistische Modell von Rasch

Die Item Response Theory (IRT) geht auf das dichotome logistische Modell von Rasch (1960, 1980), zumeist auch als Rasch-Modell bezeichnet, zurück (vgl. Kubinger, 2003). Wie der Ausdruck

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

darstellt, beschreibt es die Wahrscheinlichkeit, dass eine Testperson (Tp)  $v$  ein Item  $i$  löst („1“), in Abhängigkeit des Personenparameters  $\xi_v$ , d.h. die (wahre) Fähigkeit von  $v$ , und des Itemparameters  $\sigma_i$ , d.h. die (wahre) Schwierigkeit von  $i$  (Kubinger, 2003). Der Itemparameter (Leichtigkeit, logarithmiert  $\sigma$  bzw.  $\varepsilon$ ) sowie der Personenparameter (Fähigkeit, logarithmiert  $\xi$  bzw.  $\theta$ ) können voneinander unabhängig geschätzt werden.

Dieses Modell mit seinen spezifischen Voraussetzungen ist nach seinem Begründer, dem dänischen Mathematiker Georg Rasch (1901–1980), benannt und den *Latent-Trait*-Modellen zuzurechnen. Es erhebt den Anspruch, spezifisch objektive Testresultate, item- und personenunabhängig, abzubilden. Im Rahmen des Rasch-Modells wird davon ausgegangen, dass unabhängig davon, welche Items aus einem inhaltlich festgelegten Itempool herangezogen werden, sich für eine Testperson vergleichbare Fähigkeitsparameter ergeben, und dass sich – vice versa – vergleichbare Itemparameter zeigen. Ebenso unabhängig ist dies davon, welche Personengruppen aus einer vorher festgelegten Population getestet wurden. Demnach können bei spezifisch objektiven Messungen zwei Personen bezüglich der Stärke der Ausprägung eines Merkmals verglichen werden – unabhängig von den Testleistungen anderer Personen. Ebenso sind Aussagen über intraindividuelle Veränderungen zu einem Persönlichkeitsmerkmal im Zeitverlauf exklusiv mit diesem Testmodell möglich. Zudem liefert es Testwerte, die auf dem (logarithmierten) Rationalskalenniveau, dem statistisch höchsten Skalenniveau, liegen, womit Verhältnisse abgebildet werden können. Schließlich erlaubt es auch eine eindeutige Charakterisierung gegenüber dem Testmodell der "Klassischen Testtheorie": Das zugrundeliegende "verhaltenstheoretischen" Konzept beinhaltet, dass, je stärker ein zu erfassendes Merkmal bei einer Person ausgeprägt ist, umso höher auch die Wahrscheinlichkeit (*probability*) ist, Items – die dieses Merkmal repräsentieren – positiv zu beantworten bzw. zu lösen.

Die Item-Charakteristik (IC)-Funktionskurve zwischen individuellen Fähigkeitsparametern und der Wahrscheinlichkeit, solche Items zu lösen, zeigt eine logistische, sigmoide Verlaufsform. Dieses Testmodell erweist sich als direkt falsifizierbar. Items oder Personen, die diesem Funktionsverlauf nicht entsprechen, gelten demnach als nicht Rasch-skalierbar. Umfassende Abhandlungen werden bei Fischer (1974) sowie bei Kubinger (2003, 2009) und Kubinger, Rasch & Yanagida (2011a) dargestellt. Die Reliabilität im Sinne der internen Skalenkonsistenz (Rost, 2004) kann als Resultat Rasch-konformer Testentwicklungen als gegeben angenommen werden, da im Rahmen der probabilistischen Testtheorie sämtliche Skalen-Items eine ideale Trennschärfe aufweisen.

Das ursprünglich speziell nur für dichotome Items entwickelte einparametrische Testmodell nach Rasch berücksichtigt einzig die Schwierigkeit der Items, womit die IC-Kurven parallel zueinander verlaufen. Die Items unterscheiden sich im Rasch-Modell nur durch ihren Abstand auf der latenten Dimension, der durch die Differenz der Itemschwierigkeiten  $\sigma_i - \sigma_j$  ausgedrückt ist. Die vorliegende Arbeit basiert auf den Grundannahmen dieses speziellen Modells, während Weiterentwicklungen z.B. das „Zwei-Parameter-Modell“ mit nicht-

parallelen IC-Kurven von Birnbaum (1968) zusätzlich auch die Diskriminations-Fähigkeit der Items berücksichtigen. Dieses Modell weist daher unterschiedliche Anstiegskoeffizienten auf und die IC-Kurven können einander schneiden. Zusätzlich kann auch die Rate-wahrscheinlichkeit im sogenannten „Drei-Parameter-Modell“ Berücksichtigung finden. Das spezielle Rasch-Modell wurde in späterer Folge auch für polytome, ordinale Items mit mehr als zwei Merkmalsausprägungen weiterentwickelt. Das Partial-Credit (PC) - Modell ist eine Verallgemeinerung des Rasch-Modells bei mehrkategorialen, ordinalen Itemantworten und bei teilrichtigem Beantworten von Items mit unterschiedlichen Stufen adäquat und daher für *performance*, beispielsweise bei Erfassung von Intelligenz oder Fähigkeiten, geeignet (Masters, 1982). Der Grundgedanke ist, dass ein einfaches zweikategorielles Beantworten im Sinne des Lösen einer Aufgabe (richtig vs. falsch) weniger an Information bietet und mit Hilfe des PC-Modells differenziertere Schätzungen zur Fähigkeit einer Person ermöglicht werden sollen. Mit diesem Modellansatz könnte zudem die Zahl der Abstufungen pro Item und die relative Schwierigkeit zwischen den Stufen von Item zu Item berücksichtigt werden. Diese Modelleigenschaft kann sich für Weiterentwicklungen in der Testkonstruktion als sehr nützlich erweisen.

Die in der „Klassischen Testtheorie“ gängigen intervallskalierten Ratingskalen, zumeist als Likert-Skalen bezeichnet, sind dadurch charakterisiert, dass per Konvention über alle Items von identischen Abständen der Merkmalsausprägungen auszugehen ist. Im Rahmen der probabilistischen Testtheorie sind beim Ratingskalen-Modell (*Rating Scale Model*, RSM) demnach ordinalskalierte Kategorie-Daten über ein vorab festgelegtes Set von Antwortalternativen je Item vorgesehen, wobei das RSM nicht für die Erfassung von Fähigkeiten vorgesehen ist. Typisch für dieses Modell ist demnach die Vorgabe einer einzigen Art von Antwortformat für jedes Item. Während im PCM für jede Kategorie pro Item ein eigener Parameter festzulegen ist, wird im Ratingskalenmodell für sämtliche Items ein identes Antwortformat (gleiche Anzahl an Kategorien, gleiche Benennung) angewendet, wobei die Schwellenabstände zwischen den Kategorien für sämtliche Items jeweils gleich groß sind. Daher ist, im Gegensatz zum PCM, pro Item ein Schwierigkeitsparameter und je ein Attraktivitätsparameter (Schwellenparameter) für die einzelnen Kategorien vorgesehen (Andrich, 1978). Der Einsatz mehrkategorialer Verallgemeinerungen des einparametrischen Rasch-Modells unterliegt in der Praxis jedoch Limitationen, da die Anzahl der dann zu schätzenden Parameter vergleichsweise hohe Stichprobenumfänge erfordern würde (Kubinger, 2003 in Kubinger & Jäger, S.421).

Das den sog. *Latent-Trait-Modellen* zuzurechnende, nach dem dänischen Mathematiker Georg Rasch (1901–1980) benannte Testmodell, erhebt den Anspruch, spezifisch objektive Testresultate zu gewinnen. Mit diesem item- und personenunabhängigen Ansatz werden die Kritikpunkte, die klassisch konstruierte Verfahren aufweisen, prinzipiell vermieden. Demnach ergeben sich für eine getestete Person vergleichbare Testresultate (Personenparameter), gleichgültig welche Items aus einem vorab inhaltlich definierten Itempool verwendet werden; vice versa ergeben sich stets die gleichen Itemparameter (Wahrscheinlichkeit, ein Item zu lösen), unabhängig davon, welche Personengruppen aus einer definierten Population untersucht werden. In Abhängigkeit der verwendeten Stichprobe und Itemauswahl können sich allerdings die konkreten Parameterwerte unterscheiden, nur das Verhältnis zwischen den Personen und Items bleibt bestehen.

Der Vorteil dieses Testmodells ist darüber hinaus die Untersuchbarkeit über graduelle Veränderungen eines Persönlichkeitsmerkmals bei ein- und derselben Person im Zeitverlauf, d.h. bei längsschnittlichen Studien. Zudem liefert es Testresultate, die das höchstmögliche Skalenniveau, das logarithmierte Rationalskalenniveau gewährleisten, d.h. es können Testresultate von Personen einem direkten Vergleich unterzogen werden. Schließlich kann mit der probabilistischen Testtheorie auch eine eindeutige Charakterisierung des ihm zugrundeliegenden "verhaltenstheoretischen" Konzeptes vorgenommen werden: Je stärker ein Leistungsmerkmal bei einer Testperson ausgeprägt ist, umso höher ist die Wahrscheinlichkeit, dass sie Items, die dieses Merkmal repräsentieren, lösen kann. Die für das Item charakteristische Funktionskurve zwischen individuellem Testresultat (Personenparametern) und der Wahrscheinlichkeit, ein Item zu lösen, zeigt eine logistische, sigmoide Form und ist streng monoton wachsend. Demnach kann für dieses Testmodell auch eine Falsifizierbarkeit angenommen werden: Personen oder Items, die diesen Funktionsverlauf nicht aufweisen, gelten als nicht "nach dem Rasch-Modell skalierbar" (Rasch, 1961). Die Prüfbarkeit des Modells macht es für die Konstruktion psychologischer Tests attraktiv. Der *Likelihood-Ratio-Test* nach Andersen (1973) stellt dabei die bekannteste Modellprüfung dar. Dieser baut auf der Eigenschaft der Stichprobenunabhängigkeit des Rasch-Modells auf (Kubinger, 2003). Der Vergleich zweier Itemparameterschätzungen jener verwendeten Stichprobe ist daher statistisch unabhängig. Ebenfalls kann dies durch eine graphische Modellkontrolle gezeigt werden, indem die geschätzten Itemparameter der zwei Teilstichproben in ein rechtwinkeliges Koordinatensystem eingetragen werden. Sind die Schätzungen gleich, liegen sie nahe bzw. auf der 45°-Geraden (Kubinger, 2003). Nicht konforme Items werden nach und nach so lange

ausgeschlossen, bis der *Likelihood-Ratio*-Test für jegliche Teilungskriterien nicht-signifikante Werte aufweist. Somit kann nachträglich die Gültigkeit des Rasch-Modells angenommen werden.



## 2 Verfahren

Im Rahmen der vorliegenden Arbeit wurden drei Tests hinsichtlich ihrer Konformität zum Rasch-Modell untersucht, wie der anschließende Überblick zeigt. Anzumerken ist, dass die Vorgabereihenfolge der Items in den jeweiligen Tests systematisch variiert war und in diesem Zusammenhang nachfolgend als „Variante“ bezeichnet wird.

Der Family-Relation-Reasoning-Test (FRRT) wurde in einer Version mit 20 Items in drei Varianten vorgelegt und erfasst mit seinem lexikalischen Aufgabenmaterial kristallisierte Intelligenz. Der Figurale-Analogien-Test (FA) wurde in einer Version mit 27 Items in zwei Varianten mit figuralem Material vorgegeben und erfasst ebenfalls kristallisierte Intelligenz. Der Extended-Syllogismen (ES) erfasst mit seinen 20 Items in zwei Varianten fluide Intelligenz mit lexikalischem Aufgabenmaterial.

### 2.1 Family-Relation-Reasoning-Test (Kubinger, Poinstingl, Weidinger & Skoda, in Vorbereitung)

Nach einer Idee von Kubinger, entwickelte Skoda (2005) in ihrer Diplomarbeit den Verwandtschaften-Reasoning-Test (VRT), welcher Reasoning im sprachlichen Bereich anhand von Aussagen zu Verwandtschaftsbeziehungen messen soll. Dieser sollte als Alternative zu den Syllogismen (Srp, 1994) und aufgrund der geringen Anzahl an Verfahren zur Messung des schlussfolgernden Denkens im sprachlichen Bereich, erstellt werden.

Die kristalline Intelligenz wird anhand lexikalischen Aufgabenmaterials erfasst. Die Aufgabe der Testperson ist es, die fehlende Verwandtschaftsbeziehung aus Sätzen über Familienbeziehungen durch logisches Kombinieren zu finden. Das Verfahren umfasst 20 Items (aus insgesamt 29 Items wurden 3 Testformen mit systematisch variiertem Vorgabereihenfolge zusammengestellt) mit einer vorgegebenen Bearbeitungsdauer von 15 Minuten. Wie im nachfolgenden Beispielitem des FRRT (Abbildung 2) ersichtlich, soll der korrekte Verwandtschaftsgrad als Antwort eingetragen werden.

<p>Karin hat die Tochter Tina. Peter ist der Bruder von Tina. Karin ist Peters</p> <p>_____</p>
---

Abbildung 2. Beispielitem des Family-Relation-Reasoning-Test (FRRT)

## 2.2 Figurale Analogien (Ünal & Kubinger, in Vorbereitung)

Kristallisierte Intelligenz wird anhand figuralen Aufgabenmaterials erfasst. Dies erfordert ein Wissen zu verschiedenen Bereichen, wie Geographie, Geschichte, Naturwissenschaften sowie Wissen über alltägliche Symbole und Gegenstände. Es werden dazu Analogien verwendet, welche einen logischen Schluss in Bezug auf ihre Ähnlichkeit erfordern. 27 Items werden in 2 Testformen bei einer Bearbeitungszeit von 8 Minuten vorgegeben.

Hier sei noch anzumerken, dass die a-priori Ratewahrscheinlichkeit 1/5 beträgt. Bei der Erstellung des Verfahrens ergaben die Rasch-Modell-Analysen Modellgültigkeit, somit kann die Anzahl gelöster Aufgaben als fairer Test-Kennwert verrechnet werden.

Die Abbildung 3 zeigt als Beispielimitem die Relation von männlich zu weiblich wie Hahn zu Henne. Aus den vorgegebenen Antwortmöglichkeiten soll die korrekte durch Ankreuzen kenntlich gemacht werden.

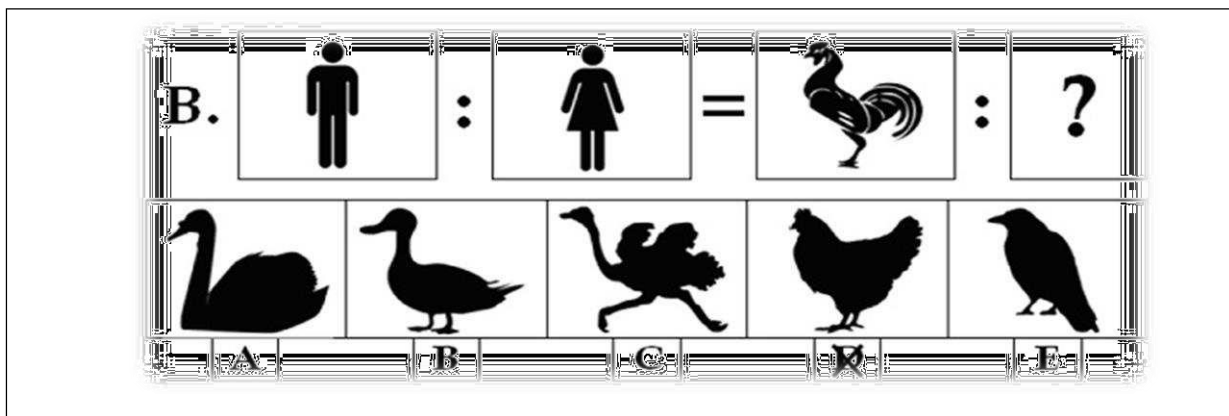


Abbildung 3. Beispielimitem für Figurale Analogien

## 2.3 Extended Syllogismen (Treiber & Kubinger, in Vorbereitung)

Srp (1993) entwickelte im Rahmen ihrer Diplomarbeit einen Syllogismen-Test mit insgesamt 74 Items, für welche Rasch-Modell-Konformität angenommen werden kann. Dabei handelt es sich um einen computergestützten Test, der *tailored testing* (vgl. Kubinger, 1996) ermöglicht. Im Rahmen ihrer Abschlussarbeit entwickelte Treiber (2010) ebenfalls 55 Items, ausgehend von den 19 schlüssigen Syllogismen der klassischen Logik, für welche Rasch-Modell-Konformität angenommen werden kann.

Der formalen Logik entnommen, besteht ein „Syllogismus“ aus drei Aussagen, und zwar aus zwei Prämissen und einer Konklusion. Aufgabe ist es, aus den beiden Prämissen zur richtigen Konklusion zu gelangen (Kubinger, 2009, S. 140).

In der vorliegenden Arbeit wurden insgesamt 20 Items in zwei Testformen vorgegeben, wobei die Reihenfolge der Items systematisch variiert war. Als Bearbeitungszeit waren 13 Minuten vorgesehen. Die Abbildung 4 zeigt Itembeispiele für sinnvolle und sinnfreie Syllogismen.

Jeder Baum ist eine Pflanze. Jede Birke ist ein Baum. (sinnvoll)	Jeder Nasder ist glänzend. Einige Polgen sind nicht glänzend. (sinnfrei)
a) Einige Birken sind keine Pflanzen. b) Jede Birke ist eine Pflanze. c) Keine Birke ist eine Pflanze. d) Einige Birken sind Pflanzen. e) Alle Birken sind keine Pflanzen.	f) Jeder Polge ist kein Nasder. g) Einige Polgen sind keine Nasder. h) Einige Polgen sind Nasder. i) Jeder Polge ist ein Nasder. j) Kein Polge ist ein Nasder.

Abbildung 4. Beispielitem für Extended Syllogismen

### 2.3.1 Anstrengungsvermeidungstest (AVT, Rollett & Rollett, 2000)

Die Konzeption des ursprünglichen Fragebogens (AVT; Rollett & Bartram, 1998) dient dazu, schulische Anstrengungsvermeidung sowie Pflichteifer zu erfassen. Anstrengungsvermeidung gemäß Rollett & Bartram wird nach Gashi (2011, S.7ff) als „Tendenz, sich ... emotionell negativ erlebten Anstrengungen durch den Einsatz geeigneter Vermeidungsstrategien zu entziehen“ definiert.

Untersuchungen zu motivationalen Bedingungen im Rahmen des schulischen Lernens zeigten, dass mangelnde Leistungsbereitschaft von Schülerinnen und Schülern nicht nur über die Abwesenheit einer entsprechenden Leistungsmotivation zu erklären ist. Wenn in einem bestimmten Leistungsbereich frustrierende Erfahrungen gemacht wurden, können diese dazu führen, dass sich eine entsprechend neue Motivation entwickelt, mit dem Ziel, Anforderungen in diesem Handlungsfeld aktiv zu vermeiden und zu umgehen. Diese wurde von den Autoren unter dem Begriff Anstrengungsvermeidungsmotivation zusammengefasst.

Der AVT ist ein Verfahren zur Erhebung von Anstrengungsvermeidung in Bezug auf Anforderungen der sozialen Umwelt und insbesondere von aktivem Vermeiden leistungsbezogener Aktivitäten. Das Konstrukt Pflichteifer wird mit einer weiteren Skala erfasst. Der

Test wurde nach den Kriterien der Probabilistischen Testtheorie gemäß Rasch konstruiert. Demnach kann die Eindimensionalität der beiden Skalen vorausgesetzt werden.

Herangezogen wurde eine adaptierte Version für ältere Schüler (Rollett & Rollett, 2000). Da das Inventar ursprünglich für ältere Schülerinnen und Schüler konzipiert war, wurden für die gegenständliche Erhebung einzelne Aussagen für den Erwachsenenbereich modifiziert und das Aufgabenmaterial gekürzt. Anhand von 32 Items wurde eine Stellungnahme zu Fühlen und Handeln beim Lernen und Arbeiten verlangt. Die Skalierung des Antwortformats wurde von sechs auf vier Stufen reduziert und reichte von (1) *stimmt nicht* bis (4) *stimmt*.

Die Teilnehmenden wurden ersucht, das Inventar möglichst zügig zu bearbeiten. Ein explizites Zeitlimit zum Ausfüllen war jedoch nicht vorgesehen. Für die adaptierte Version wurde der Gesamtscore für die Anstrengungsvermeidung verrechnet und als externes Teilkriterium im Rahmen der Rasch-Modell-Analysen herangezogen.

Die Abbildung 5 zeigt AVT-Beispielsitems für die Anstrengungsvermeidungsmotivation und den Pflichteifer.

„Andere wollen immer gerade dann, dass ich etwas für die Schule/Arbeit tue, ich mich erhole/entspanne“ (AV)

„Wenn ich für die Schule/Arbeit etwas tue, merke ich oft gar nicht, wie schnell die Zeit vergeht“ (PE)

Abbildung 5. Beispielsitems für Anstrengungsvermeidung (AV) und Pflichteifer (PE)

## 3 Empirischer Teil

### 3.1 Hypothesen

Nachfolgend sind die zu prüfenden Hypothesen angeführt:

$H_0$  (1):

Die Items des Verfahrens Family-Relation-Reasoning-Test (FRRT, Kubinger, Poinstingl, Weidinger & Skoda, in Vorb.) sind mit dem dichotom logistischen Modell von Rasch konform.

$H_0$  (2):

Die Items des Verfahrens Extended Syllogisms (ES, Treiber & Kubinger, in Vorb.) sind mit dem dichotom logistischen Modell von Rasch konform.

### 3.2 Methode

Die deskriptiv- und inferenzstatistischen Auswertungen wurden mit dem Statistikprogramm IBM SPSS® 20 für Windows und die testtheoretischen Analysen mit Hilfe des R-Paketes eRm (Mair, Hatzinger & Maier, 2015) durchgeführt. Das Signifikanzniveau wurde vorab, entsprechend der Irrtumswahrscheinlichkeit, mit  $\alpha = 5\%$  festgelegt, sodass ein Ergebnis im Rahmen der Inferenzstatistik mit  $p \leq .05$  als signifikant bezeichnet wird.

Zur Charakterisierung metrischer Parameter wurden die Kennwerte arithmetisches Mittel ( $M$ ), Standardabweichung ( $SD$ ), Minimum (min), Maximum (max) sowie bei schiefen Messwerteverteilungen das alternative Lagemaß Median ( $Md$ ) mit dem Innerquartilsabstand (IQR) angeführt. Zur Veranschaulichung der Verteilung von Messwerten wurden auch Histogramme erstellt; für die Modellprüfungen wurden entsprechende bivariate Streudiagramme angefertigt. Zur Beschreibung der Kategorien nominalskaliertter Variablen wurden die Häufigkeiten und ihre entsprechenden Anteilswerte (%) ermittelt und angegeben. Gegebenenfalls wurden hierzu Kreisdiagramme erstellt.

Im Rahmen der Inferenzstatistik wurden die Zusammenhänge zwischen den Reasoning-Verfahren mit dem Koeffizienten der Rangkorrelation nach Spearman  $r_s$  untersucht. Zudem wurden auch die entsprechenden 95% - Konfidenzintervalle (KI) ermittelt, um den Bereich, in dem der wahre Wert der Korrelation liegen kann, darzustellen. Für den Zusammenhang zwischen zwei nominalskalierten Variablen wurde auf Grundlage von Kreuztabellen der Assoziationskoeffizient  $r_\phi$  herangezogen.

### 3.2.1 Planung und Durchführung der Untersuchung

Der Erhebungszeitraum erstreckte sich von März bis Juni 2015. Als Bearbeitungszeit waren für die Figuralen Analogien 8 Minuten, für den Family-Relation-Reasoning-Test 15 Minuten und für die Extended Syllogismen 13 Minuten vorgesehen, wie in Tabelle 1 angeführt. Für diese drei Verfahren waren die angeführten Zeiten verbindlich. Für den Anstrengungsvermeidungstest war kein explizites Zeitlimit vorgesehen; hierzu waren im Durchschnitt 5 bis 10 Minuten erforderlich. Hinzuzurechnen sind noch die Instruktionsphasen und die Erhebung der soziodemographischen Eigenschaften, so dass in Summe etwa 60 Minuten als Gesamtdauer im Gruppensetting zu veranschlagen waren.

Die Testleiteranweisungen waren in der Grundstruktur gleich gehalten, wobei sich jedoch die Informationen zu den jeweiligen Verfahren am Aufgabenmaterial orientierten. Die entsprechende Testleiteranweisung der vorliegenden Arbeit, die Instruktion für die Teilnehmenden sowie die Testinhalte können dem Anhang entnommen werden.

Tabelle 1 *Überblick zu den Konstrukten und Aufgabenmaterialien der angewendeten Verfahren*

Verfahren	Zuordnung (sensu Cattell)	Aufgabenmaterial
<i>Figurale Analogien (FA)</i>	kristallisiert	figural
<i>Family-Relation-Reasoning (FRRT)</i>	kristallisiert	lexikalisch
<i>Extended Syllogismen (ES)</i>	fluid	lexikalisch

Die Vorgabereihenfolge der Testbatterie war ident, jedoch unterlag die Itemabfolge innerhalb der Reasoningtests einer systematischen Variation, so dass sechs Testheftversionen resultierten. Die Tabelle 2 zeigt die erzielten Anordnungen anhand der Testhefte.

Tabelle 2 *Vorgabereihenfolge der Verfahren gemäß Testheftversion*

Testheftversion	Deckblatt	Ext. Syllogismen	Fig. Analogien	FRRT	AVT-Schüler
1	A	A	A	A	A
2	A	B	B	B	A
3	A	A	A	C	A
4	A	B	B	A	A
5	A	A	A	B	A
6	A	B	B	C	A

### 3.2.2 Stichprobenbeschreibung

Für die vorliegende Studie konnten die Datenprotokolle von insgesamt 304 Teilnehmenden herangezogen werden. Die Stichprobe setzte sich aus 166 (54,4%) Schülerinnen und Schülern der Gesundheits- und Krankenpflege im Wilhelminenspital und 139 (45,6%) Schülerinnen und Schülern des Sozialmedizinischen Zentrums Ost in Wien zusammen.

Die teilnehmenden Personen waren zu 71,4% weiblich und 28,6% männlich. Der Altersbereich lag zwischen 16,8 und 47,6 Jahren. Der Altersmedian lag bei 22,6 Jahren. Für weibliche Teilnehmerinnen lag der Median bei 22,1 und für männliche Teilnehmer bei 24,3 Jahren. Die Abbildung 6 zeigt die Verteilung der Altersstruktur der Stichprobe.

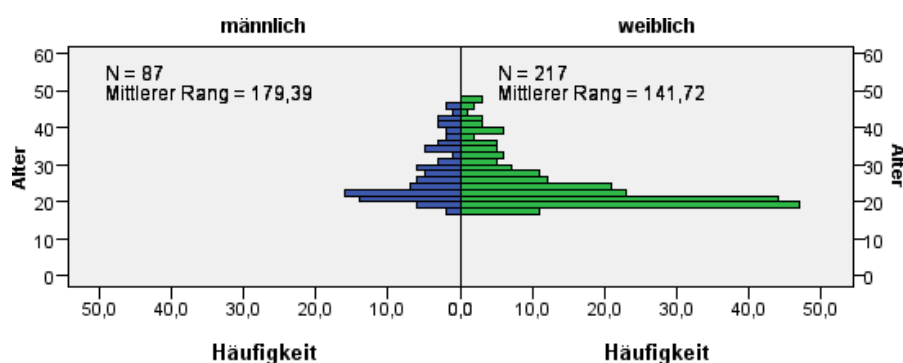


Abbildung 6. Verteilung des Lebensalters in Jahren bezüglich des Geschlechts der Teilnehmenden

Von den Teilnehmenden gaben 243 (79,9%) Deutsch und 61 (21,1%) eine andere Sprache als Erstsprache an.

Die Verteilung der Stichprobe anhand der höchsten abgeschlossenen Ausbildung war für 18 (5,9%) Pflichtschule, für 60 (19,7%) Lehre, für 90 mit dem vergleichsweise höchsten Anteilswert (29,5%) Matura an einer allgemeinbildenden höheren Schule, für 54 (17,7%) berufsbildende höhere Schulen, für 15 (4,9%) akademisches Niveau und 68 (22,3%) führten sonstige Abschlüsse an.

Die Verteilung der AVT-Testwerte in Rohwertpunkten ist in Abbildung 7 zusammen mit dem Teilungskriterium dargestellt. Dieses wurde anhand des Medians festgelegt. Die durchschnittliche Ausprägung im Anstrengungsvermeidungstest lag bei  $M_{AVT} = 8.81$ ;  $SD_{AVT} = 5.36$  ( $Md = 8.0$ ; IQR 5 – 12) bei einem Score von maximal 32.

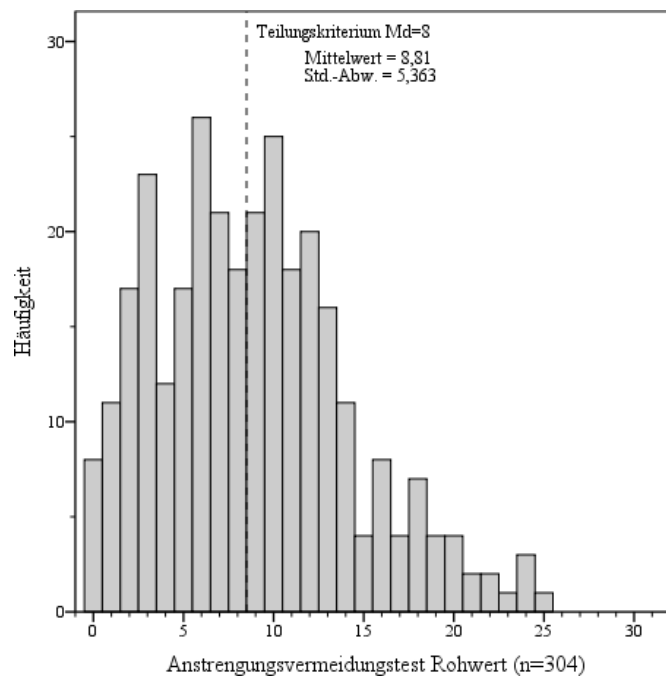


Abbildung 7. Verteilung der AVT-Rohwerte

Die Testrohwerte des FRRT (Wertebereich 0 - 20) sind in Abbildung 8 zusammen mit dem Teilungskriterium ersichtlich. Dieses wurde anhand des Medians festgelegt. Die durchschnittliche Ausprägung im Family-Relation-Reasoning-Test lag bei  $M_{\text{FRRT}} = 9.03$ ;  $SD_{\text{FRRT}} = 3.86$  ( $Md = 9$ , IQR 6 – 12) bei einem Score von maximal 20.

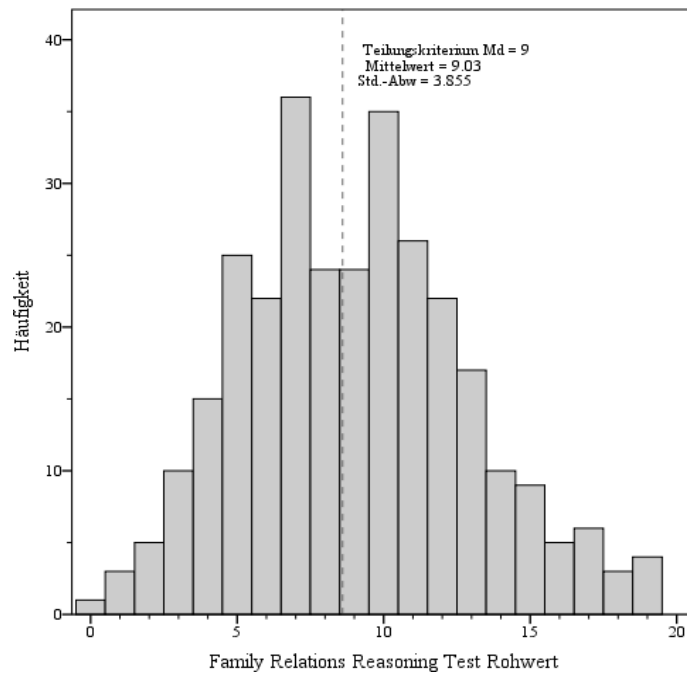


Abbildung 8. Verteilung der FRRT-Rohwerte



Die Verteilung der Testrohwerte der Extended Syllogismen (Wertebereich 0 - 20) ist in Abbildung 9 ersichtlich. Die durchschnittliche Ausprägung lag bei  $M_{ES} = 9.55$ ;  $SD_{FRRT} = 3.69$  ( $Md = 9.5$ ; IQR 7 – 12) bei einem erreichbaren Score von maximal 20.

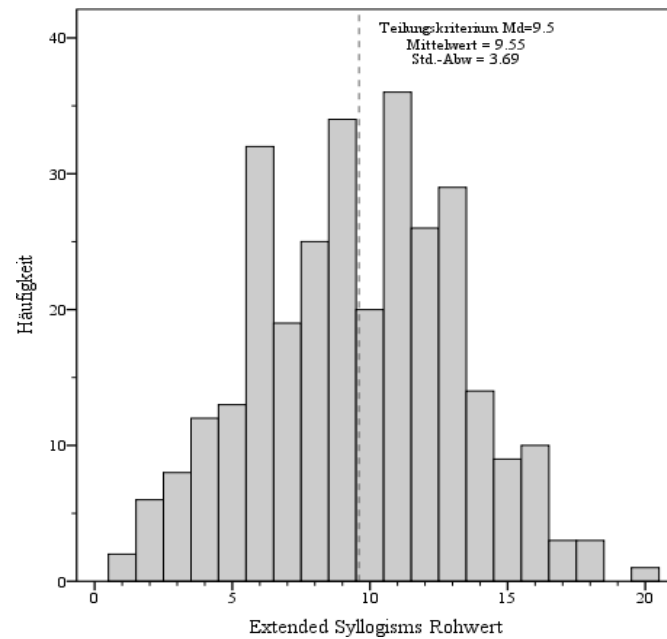


Abbildung 9. Verteilung der Extended Syllogisms-Rohwerte

## 4 Ergebnisse

### 4.1 Analysen zur Überprüfung der Gültigkeit des Rasch-Modells

Unter Verwendung der Benutzeroberfläche „R Studio“ (R Studio Team, 2015) wurden die nachfolgenden Analysen in der freien Programmiersprache R (R Core Team, 2015a) durchgeführt. Zum Laden der Daten aus der SPSS-Datenmatrix wurde das R-Pakete `foreign` (R Core Team, 2015b) herangezogen. Das globale Risiko 1. Art wurde, entsprechend der Irrtumswahrscheinlichkeit, in der Untersuchungsplanung auf  $\alpha = .05$  vorab festgelegt. Daraus ergab sich für sechs Teilungskriterien ein vergleichsbezogenes Risiko, entsprechend einer Bonferroni-Korrektur von  $\alpha = .0083$ , für jeden durchzuführenden Likelihood-Ratio-Test (LRT). Der bedingte Likelihood-Quotiententest (LQT bzw. LRT) gemäß Andersen (1973) prüft, ob die zu analysierenden Daten konform zum Rasch-Modell sind. Hierzu ist zunächst die Stichprobe nach bestimmten Teilungskriterien in zwei Gruppen aufzuteilen und die Likelihoods der beiden Subgruppen zu errechnen.

Im zweiten Schritt ist sodann der Chi-Quadrat-Wert zu berechnen. Für den LRT gilt dementsprechend: Ist der errechnete  $\chi^2$ -Wert kleiner als der kritische Wert, dann kann die  $H_0$ : „Gültigkeit des Rasch-Modells“ beibehalten werden. Ist der ermittelte  $\chi^2$ -Wert größer als der kritische Wert, dann ist die  $H_0$ : „Gültigkeit des Rasch-Modells“ zu verwerfen. Mit jedem nicht signifikanten LRT nimmt dementsprechend die Wahrscheinlichkeit zu, dass die  $H_0$  gilt, gemäß Poppers „Grad der Bewährung“ (Poinstingl, Mair & Hatzinger, 2007).

#### 4.1.1 FRRT

Als internes Teilungskriterium wurden der FRRT-Score und als externe Teilungskriterien das Alter (Mediansplit), die Ausbildung (Pflichtschule, Lehre, BHS vs. AHS, Hochschule), AVT-Score, Geschlecht und Erstsprache (Deutsch vs. Nicht-Deutsch) herangezogen. Die Teilungskriterien sollten miteinander nicht stark korreliert sein, was durch Kreuztabellen überprüft wurde. Die Ergebnisse sind in Tabelle 3 zusammengefasst. Es ergaben sich keine relevanten Zusammenhänge zwischen den externen Teilungskriterien mit dem internen, so dass die sechs angeführten Kriterien für die Analysen zum Rasch-Modell beibehalten werden konnten.

Tabelle 3 *Matrix der rphi-Koeffizienten mit Signifikanzbeurteilungen (p) zu den Kreuztabellen der dichotomen Teilungskriterien (FRRT)*

	$r_\phi$					
	1	2	3	4	5	6
1 FRRT	-	-.004	.084	.017	-.016	-.065
2 Alter	.951	-	-.045	-.254	-.167	.039
3 Ausbildung	.199	.490	-	-.026	.067	-.081
4 AVT	.763	<.001	.694	-	.068	.010
5 Geschlecht	.783	.004	.304	.233	-	.121
6 Erstsprache	.260	.492	.210	.863	.036	-

Anmerkung:  $df = 1$ ;  $\chi^2_{krit} = 3.841$ ,  $p$ -Werte sind grau unterlegt

Die Zusammenstellung der globalen Prüfgrößen gemäß Andersen's Likelihood-Ratio-Test, wie die Tabelle 4 zeigt, erbrachte für alle Teilungskriterien jeweils nicht signifikante Ergebnisse. Die Geltung des Rasch Modells konnte daher angenommen werden.

Tabelle 4 *Internes und externe Teilungskriterien für den FRRT*

Teilungskriterium	$\chi^2$	$df$	$\chi^2_{krit} (0.0083)$	$p$	Signifikanz
Score FRRT	28.734	28	49.166	.426	nicht sign.
Alter	33.545	28	49.166	.216	nicht sign.
Ausbildung	27.656	28	49.166	.483	nicht sign.
Score AVT	28.803	27	47.840	.371	nicht sign.
Geschlecht	36.253	27	47.840	.110	nicht sign.
Erstsprache	27.947	27	47.840	.414	nicht sign.

Anmerkung: AVT, Geschlecht und Erstsprache ohne Item 64

Die Rasch-Modell-Konformität wurde unter Berücksichtigung der Bonferroni-Korrektur gemäß der Anzahl der angelegten Teilungskriterien, wie nachfolgende Zusammenfassung zeigt, geprüft. Aufgrund der nicht signifikanten Ergebnisse konnte die Rasch-Modell-Geltung jeweils angenommen werden. Die nachfolgenden Modellkontrollen anhand von bivariaten Streudiagrammen zeigen zunächst jenes für das interne Teilungskriterium, den Median des FRRT-Scores, sodann für die der externen Teilungskriterien. Dargestellt sind die Parameterschätzungen der Items in den beiden, nach den Teilungskriterien getrennten

Gruppen, wie die Abbildung 10 bis Abbildung 15 zeigen. Sofern das Rasch-Modell gilt, dürfen diese Schätzungen keine unterschiedlichen Werte ergeben und sollen demnach auf der 45-Grad-Gerade durch den Ursprung liegen.

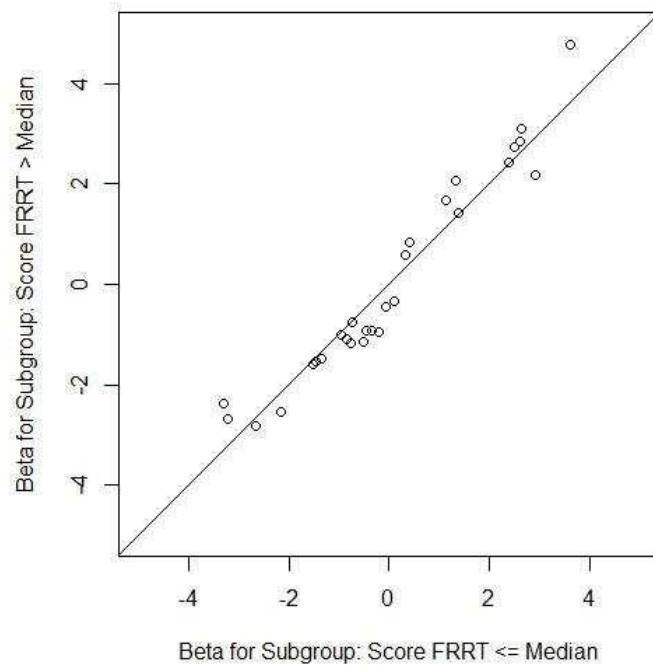


Abbildung 10. Graphische Modellkontrolle, internes Teilungskriterium FRRT-Score

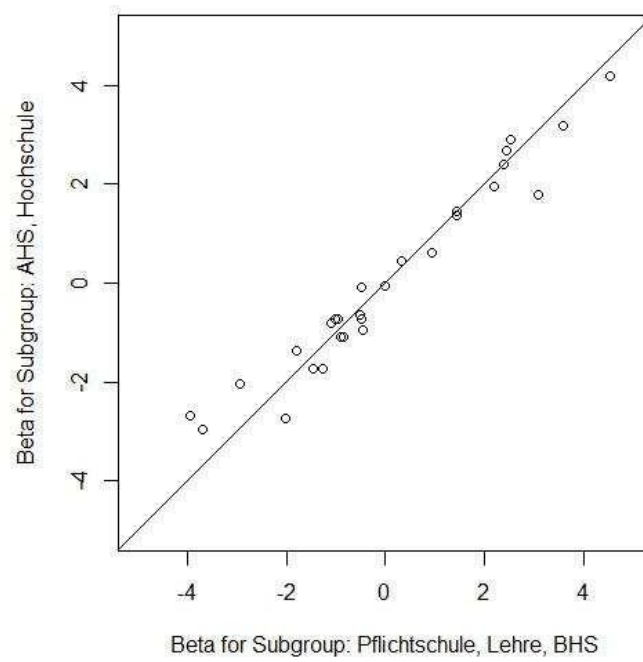


Abbildung 11. Graphische Modellkontrolle, externes Teilungskriterium Bildung

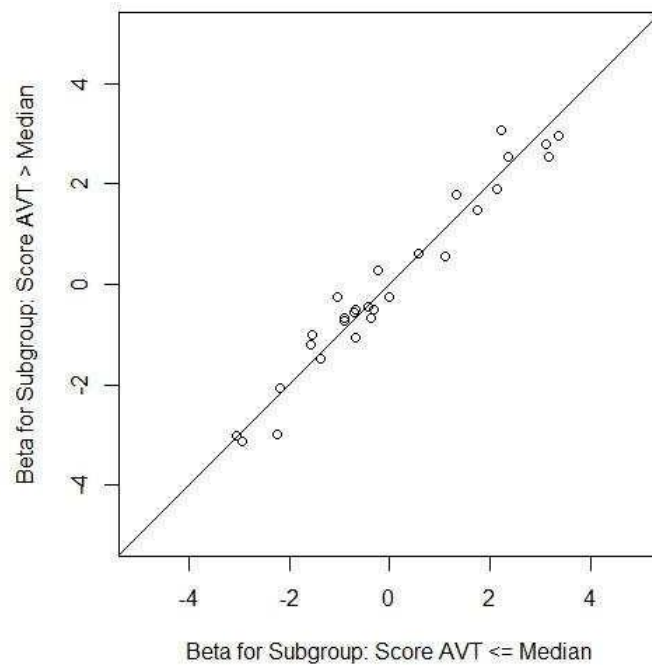


Abbildung 12. Graphische Modellkontrolle, externes Teilungskriterium AVT-Median

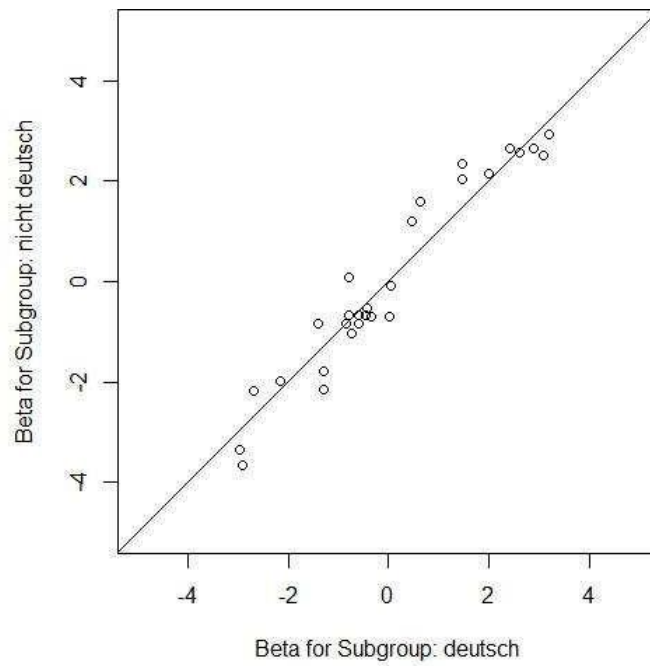


Abbildung 13. Graphische Modellkontrolle, externes Teilungskriterium Erstsprache

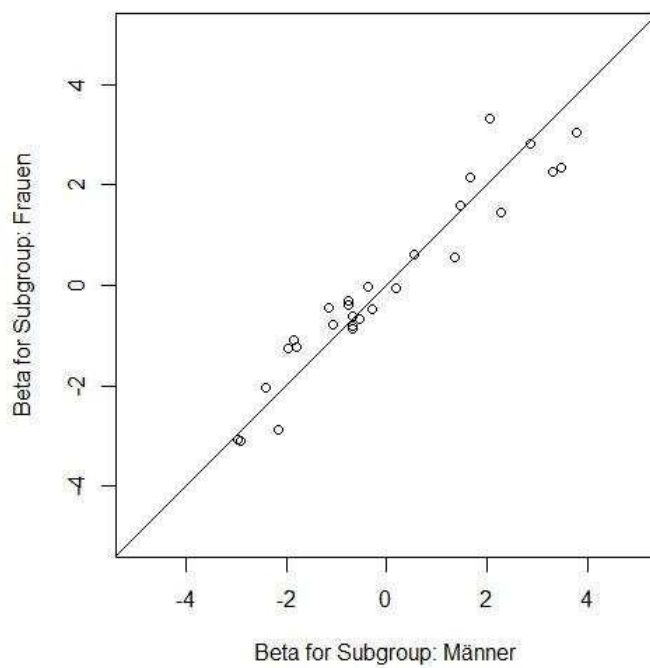


Abbildung 14. Graphische Modellkontrolle, externes Teilungskriterium Geschlecht

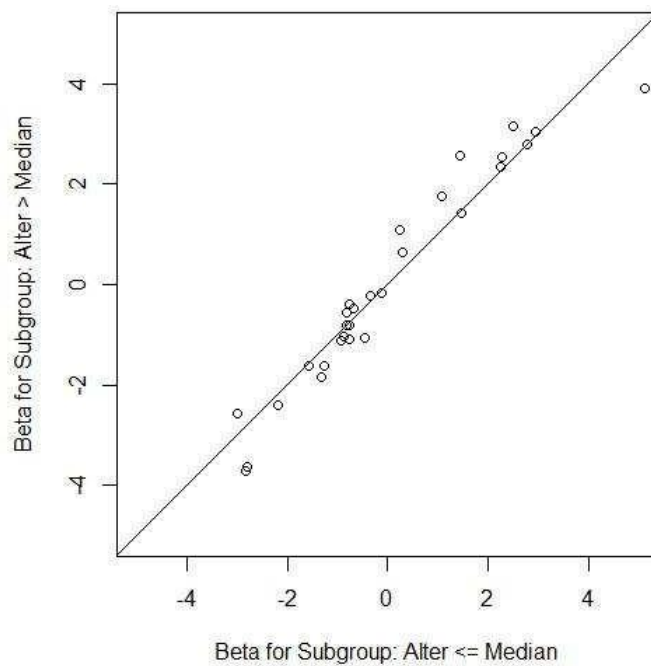


Abbildung 15. Graphische Modellkontrolle, externes Teilungskriterium Altersmedian

#### 4.1.2 Ergebnisse der Rasch-Modell-Analysen des Extended Syllogismen Test

Als internes Teilungskriterien wurden der ES-Score (Mediansplit), als externe Teilungskriterien das Alter (Mediansplit), die Ausbildung (Pflichtschule, Lehre, BHS vs. AHS, Hochschule), der AVT-Score, Geschlecht und Erstsprache (Deutsch vs. Nicht-Deutsch) herangezogen. Die Korrelation der Teilungskriterien mit dem ES-Score wurden mittels Chi-Quadrat-Testung auf Basis von Kreuztabellen überprüft. Die Ergebnisse sind in Tabelle 5 zusammengefasst.

Es zeigte sich, bis auf die Korrelation der Ausbildungsstufen mit dem ES-Score (Mediansplit), keine relevanten Zusammenhänge.

Tabelle 5 *rphi-Koeffizienten mit Signifikanzbeurteilungen (p) zu den Kreuztabellen der dichotomen Teilungskriterien (ES)*

	$r_\phi$	$p$
	1 ES	
1 ES	-	-
2 Alter	-.050	.386
3 Ausbildung	.290	<.001
4 AVT	-.012	.840
5 Geschlecht	-.080	.163
6 Erstsprache	-.059	.304

Anmerkung:  $df = 1$ ;  $\chi^2_{\text{krit}} = 3.841$ ,  $p$ -Werte sind grau unterlegt

Als internes Teilungskriterium wurde der Median des Extended Syllogismen Scores herangezogen. Als externe Teilungskriterien fungierten der Altersmedian, die Ausbildung (Pflichtschule, Lehre & BHS vs. AHS & Volkshochschule), der AVT-Median, Geschlecht, Erstsprache (Deutsch vs. nicht Deutsch). Die Zusammenstellung der globalen Prüfgrößen gemäß Andersen's Likelihood-Ratio- Test, wie die Tabelle 6 zeigt, erbrachte für das interne Teilungskriterium ES und für Ausbildung ein signifikantes Ergebnis.

Tabelle 6 *Internes und externe Teilungskriterien für ES*

Teilungskriterium	$\chi^2$	$df$	$\chi^2_{\text{krit}} (0.008)$	$p$	
Score Ext. Syll.	91.919	19	36.972	< .001	sign.
Alter	22.501	19	36.972	.260	nicht sign.
Ausbildung	38.122	19	36.972	.006	sign.
Score AVT	19.317	19	36.972	.437	nicht sign.
Geschlecht	18.000	19	36.972	.522	nicht sign.
Muttersprache	27.605	19	36.972	.091	nicht sign.

Die nachfolgenden Abbildungen zeigen die graphischen Modellkontrollen ohne Ausschluss von Items. Sofern das Rasch-Modell gilt, dürfen diese Schätzungen keine unterschiedlichen Werte ergeben und sollen demnach auf der 45-Grad-Gerade durch den Ursprung liegen.



Sofern die Konfidenzellipsen diese nicht schneiden, ist von einer signifikanten Abweichung der jeweiligen Items auszugehen. Der Übersichtlichkeit halber wurden die Ellipsen nur bei abweichenden Items eingezeichnet. Wie in Abbildung 16 zu erkennen ist, schneiden die Konfidenzellipsen der Items 2, 4, 8 und 12 die 45-Grad-Gerade durch den Ursprung nicht. Demnach waren diese Items sukzessive auszuschließen und die Analysen zu wiederholen, um die Rasch-Modell-Konformität zu erreichen.

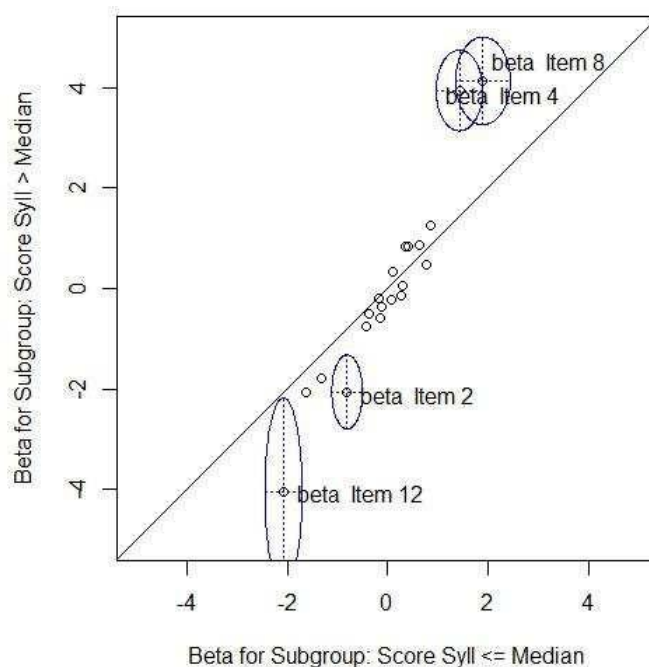


Abbildung 16. Graphische Modellkontrolle, internes Teilungskriterium ES-Score

Für die Items 4 und 8 mussten geringere Schwierigkeitsparameterschätzungen für Personen mit einem niedrigem Syllogismen-Score festgestellt werden. Für die Items 2 und 12 bestand demgegenüber eine zunächst höhere Schwierigkeitsparameterschätzung für Personen mit einem höheren Syllogismen-Score.

Die Abbildung 16 bis Abbildung 21 zeigen die graphischen Modellkontrollen für die Teilungskriterien des Extended Syllogismen Test.

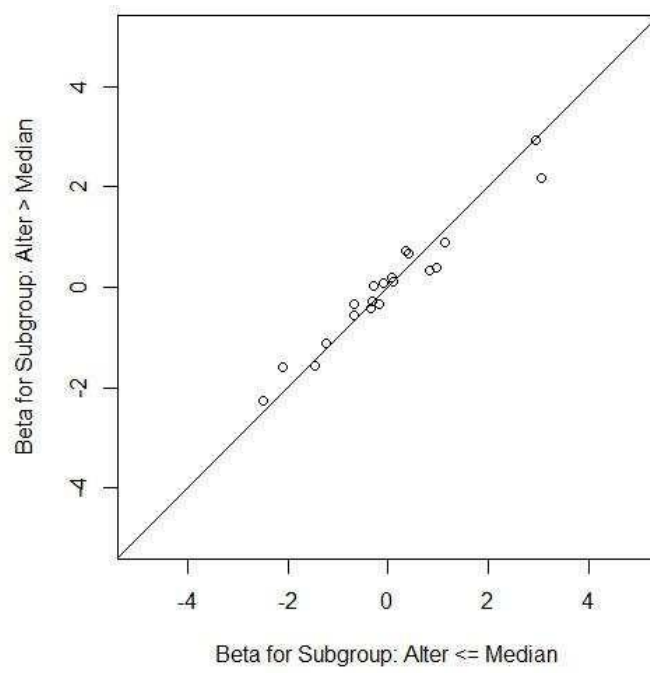


Abbildung 17. Graphische Modellkontrolle, internes Teilungskriterium Altersmedian

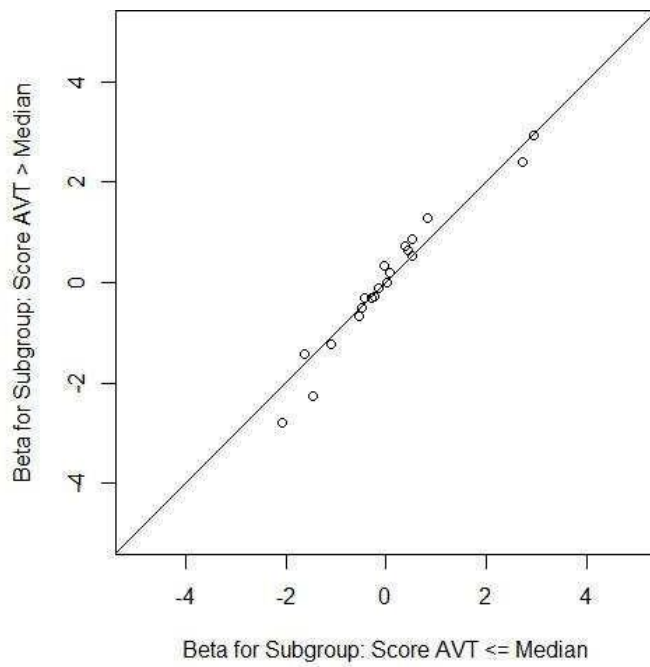


Abbildung 18. Graphische Modellkontrolle, externes Teilungskriterium AVT-Median

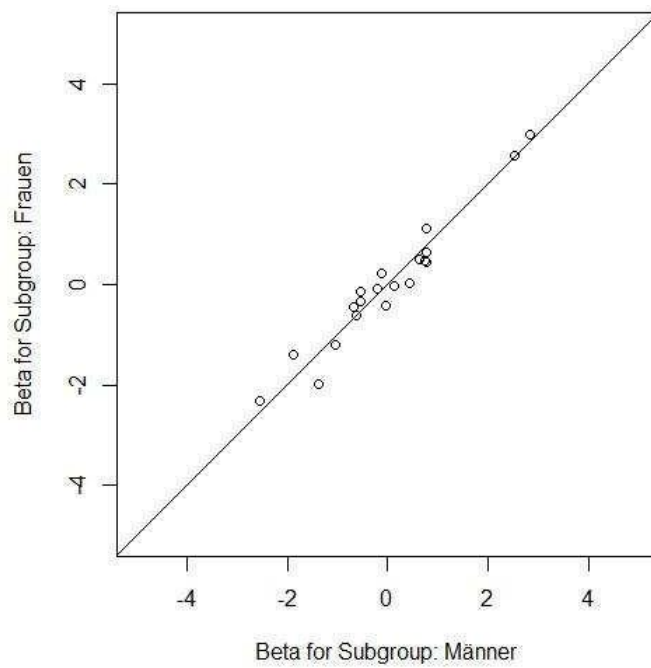


Abbildung 19. Graphische Modellkontrolle, externes Teilungskriterium Geschlecht

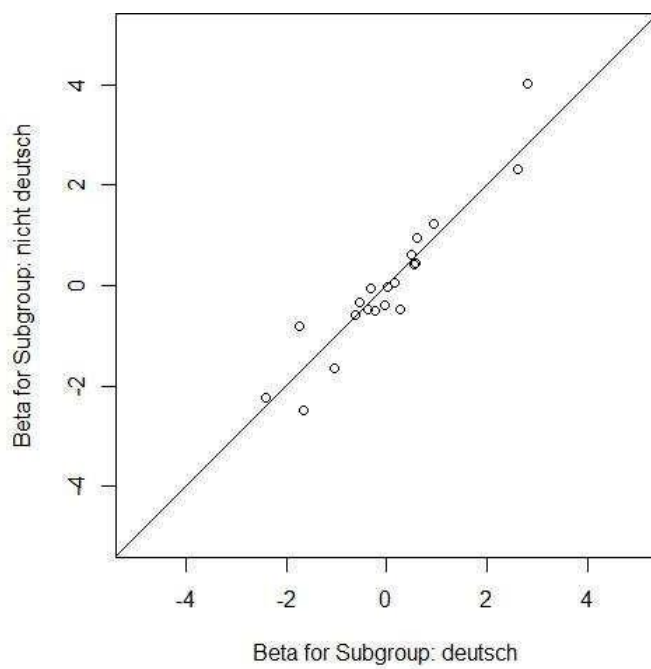


Abbildung 20. Graphische Modellkontrolle, externes Teilungskriterium Erstsprache

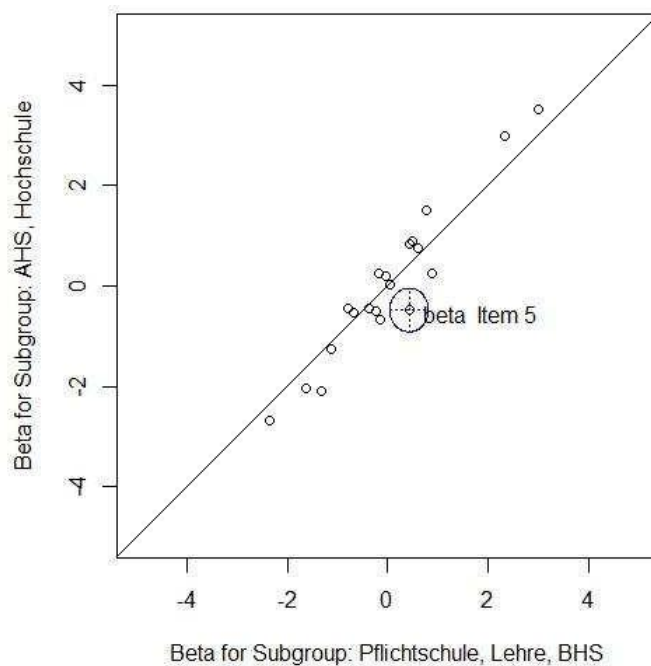


Abbildung 21. Graphische Modellkontrolle, externes Teilungskriterium Ausbildung

Das Item 5 fiel Personen mit höherer Ausbildung relativ leichter als Personen mit niedriger Ausbildung. Im Rahmen der Modellprüfungen für Extended Syllogismen wurde Item 4 als erstes ausgeschlossen und die Analysen wiederholt, die Ergebnisse finden sich zusammengefasst in Tabelle 7.

Tabelle 7 Internes und externe Teilungskriterien für ES nach Ausschluss von Item 4

Teilungskriterium	$\chi^2$	df	$\chi^2_{\text{krit}} (0.0083)$	p	
Score Ext. Syll.	59.907	18	35.573	< .001	sign.
Alter	18.312	18	35.573	.435	nicht sign.
Ausbildung	36.880	18	35.573	.005	sign.
Score AVT	19.078	18	35.573	.387	nicht sign.
Geschlecht	18.343	18	35.573	.433	nicht sign.
Muttersprache	27.393	18	35.573	.072	nicht sign.

Demnach, unter Ausschluss des Items 4, waren weiterhin signifikante, nicht Rasch-Modell-konforme Ergebnisse zu beobachten, wie die Abbildung 22 Abbildung 23 veranschaulichen.

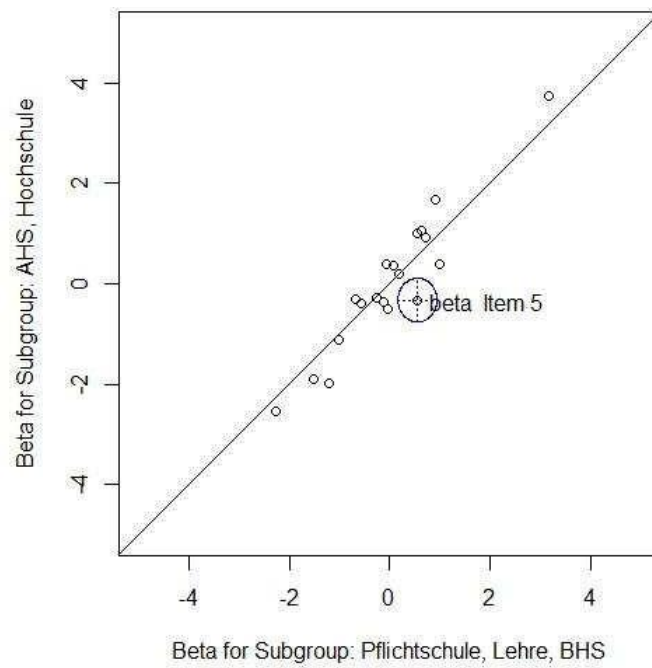


Abbildung 22. Graphische Modellkontrolle, externes Teilungskriterium Ausbildung

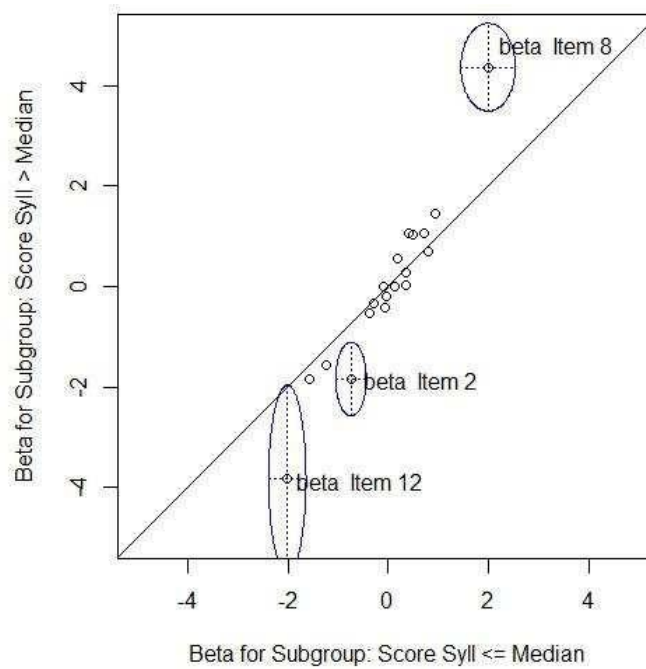


Abbildung 23. Graphische Modellkontrolle, internes Teilungskriterium ES-Median

Die Tabelle 8 zeigt die Teilungskriterien für Extended Syllogismen ohne Item 4 und 8. Im Teilungskriterium Ausbildung zeigt sich weiterhin ein signifikanter Wert.

Tabelle 8 *Internes und externe Teilungskriterien für ES nach Ausschluss der Items 4 und 8*

Teilungskriterium	$\chi^2$	$df$	$\chi^2_{\text{krit}} (0.0083)$	$p$	
Score Ext. Syll.	37.662	17	34.162	0.003	sign.
Alter	18.471	17	34.162	0.360	nicht sign.
Ausbildung	36.031	17	34.162	0.005	sign.
Score AVT	19.028	17	34.162	0.327	nicht sign.
Geschlecht	18.326	17	34.162	0.369	nicht sign.
Muttersprache	25.460	17	34.162	0.085	nicht sign.

Demnach, unter Ausschluss der Items 4 und 8, waren weiterhin signifikante, nicht Rasch-Modell-konforme Ergebnisse zu beobachten, wie die Abbildung 24 und Abbildung 25 zeigen.

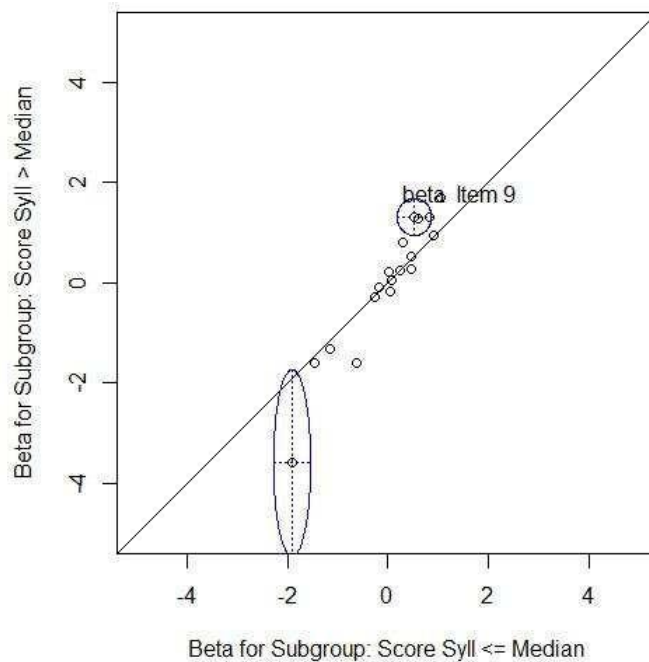


Abbildung 24. Graphische Modellkontrolle, internes Teilungskriterium ES-Median

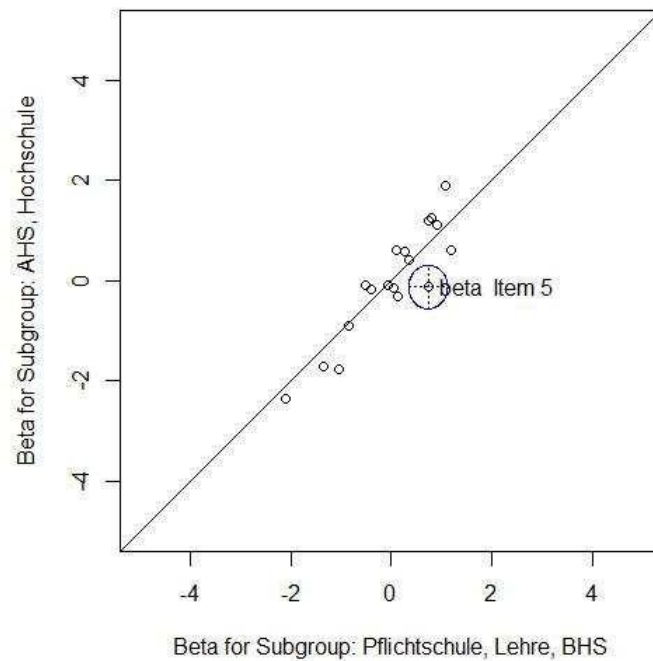


Abbildung 25. Graphische Modellkontrolle, externes Teilungskriterium Ausbildung

Item 9 wies für Personen mit einem höheren Extended Syllogismen-Score eine höhere Schwierigkeitsparameterschätzung auf und war für die weitere Prüfung ebenfalls auszuschließen.

Die Tabelle 9 zeigt die Teilungskriterien für Extended Syllogismen ohne die Items 4, 8 und 9. Im Teilungskriterium Ausbildung ist weiterhin ein signifikanter Wert zu verzeichnen.

Tabelle 9 Internes und externe Teilungskriterien für ES nach Ausschluss der Items 4, 8 und 9

Teilungskriterium	$\chi^2$	$df$	$\chi^2_{\text{krit}} (0.0083)$	$p$	
Score Ext. Syll.	17.937	15	31.303	.266	nicht sign.
Alter	18.426	16	32.739	.300	nicht sign.
Ausbildung	34.248	16	32.739	.005	sign.
Score AVT	18.571	16	32.739	.292	nicht sign.
Geschlecht	17.242	16	32.739	.370	nicht sign.
Muttersprache	25.438	16	32.739	.062	nicht sign.

Demnach, unter Ausschluss der Items 4, 8 und 9 waren weiterhin signifikante, nicht Rasch-Modell-konforme Ergebnisse zu beobachten, wie die Abbildung 26 und Abbildung 27 zeigen.

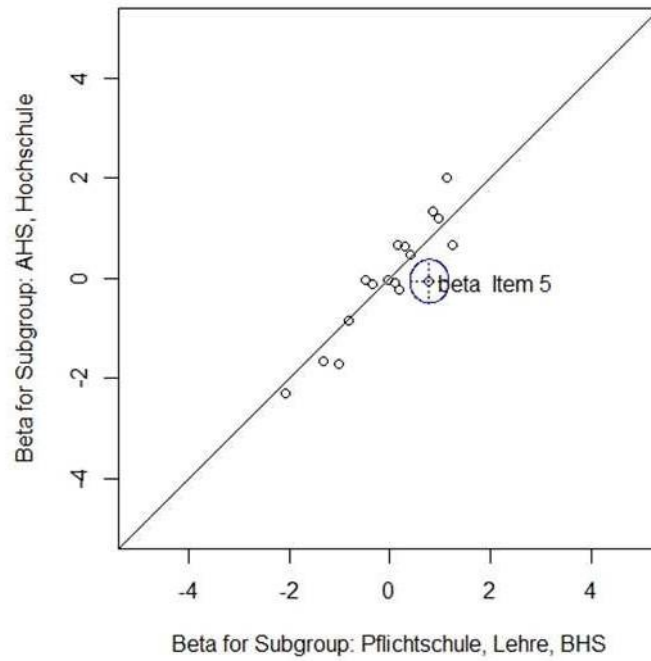


Abbildung 26. Graphische Modellkontrolle, externes Teilungskriterium Ausbildung (Item 5 Ausbildung,  $p = .005$ )

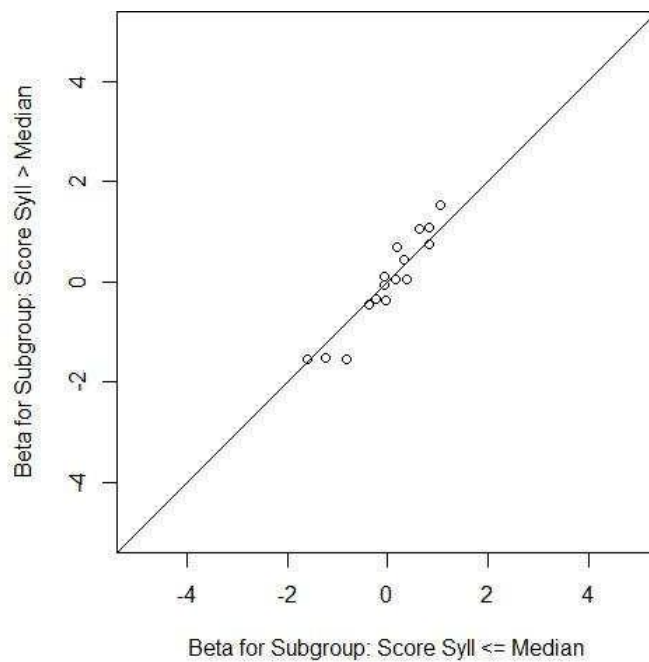


Abbildung 27. Graphische Modellkontrolle, internes Teilungskriterium ES-Median



Die Tabelle 10 zeigt in einer Zusammenfassung die Teilungskriterien für Extended Syllogismen ohne den Items 4, 5, 8 und 9. Nach sukzessivem Ausschluss der Items 4, 5, 8 und 9 zeigten sich keine signifikanten Werte und es konnte a posteriori Modellgültigkeit angenommen werden.

Tabelle 10 *Internes und externe Teilungskriterien für ES nach Ausschluss der Items 4, 5, 8 und 9*

Teilungskriterium	$\chi^2$	$df$	$\chi^2_{\text{krit}} (0.008)$	$p$	
Score Ext. Syll.	26.891	15	31.303	.030	nicht sign.
Alter	17.338	15	31.303	.299	nicht sign.
Ausbildung	25.561	15	31.303	.043	nicht sign.
Score AVT	18.372	15	31.303	.244	nicht sign.
Geschlecht	16.940	15	31.303	.322	nicht sign.
Muttersprache	24.415	15	31.303	.058	nicht sign.

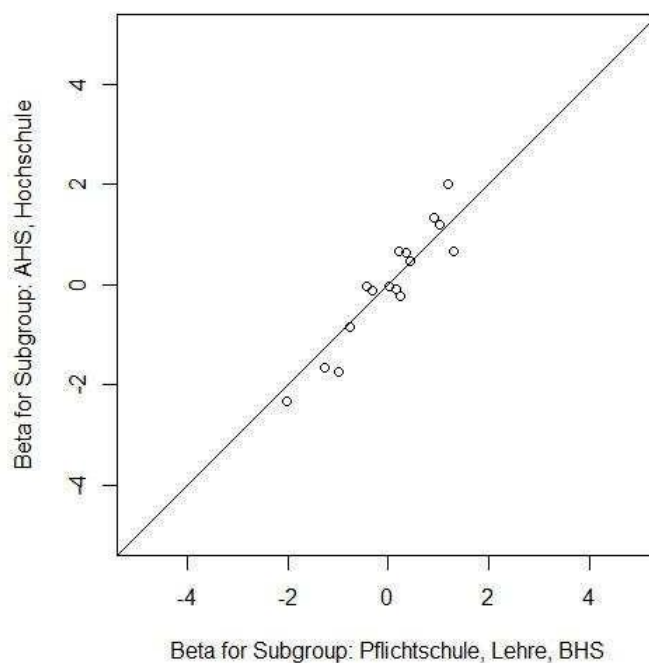


Abbildung 28. Graphische Modellkontrolle, externes Teilungskriterium Ausbildung

Die Tabelle 11 zeigt die ausgeschlossenen Items des Extended-Syllogismen-Tests mit den Antwortmöglichkeiten. Die korrekten Antworten sind jeweils hervorgehoben.

Tabelle 11 *Im Rahmen der Analyse nach dem Rasch-Modell als nicht-konform, resultierende und daher ausgeschlossene Items des Extended-Syllogismen-Tests*

ITEM 4	ITEM 5
Für alle Klavierstücke gilt, dass sie Kunstwerke sind. Für alle Klavierstücke gilt, dass sie Melodien sind.	Jedes Aliquod ist ein Substitut. Mindestens ein Aliquod ist ein Aliud.
a) Für alle Melodien gilt, dass sie keine Kunstwerke sind. b) Für alle Melodien gilt, dass sie Kunstwerke sind. c) Für keine Melodie gilt, dass sie ein Kunstwerk ist. d) Für einige Melodien gilt, dass sie keine Kunstwerke sind. e) <b>Für einige Melodien gilt, dass sie Kunstwerke sind.</b>	a) Mindestens ein Aliud ist kein Substitut. b) Kein Aliud ist ein Substitut. c) <b>Mindestens ein Aliud ist ein Substitut.</b> d) Jedes Aliud ist ein Substitut. e) Jedes Aliud ist kein Substitut.
ITEM 8	ITEM 9
Kein Quopp ist ein Wabb. Jeder Quopp ist ein Fif.	Kein Wobbler ist ein Spinner. Einige Twister sind Spinner.
a) Kein Fif ist ein Wabb. b) Alle Fifs sind Wabbs. c) Jeder Fif ist kein Wabb. d) <b>Einige Fifs sind keine Wabbs.</b> e) Einige Fifs sind Wabbs.	a) <b>Einige Twister sind keine Wobblers.</b> b) Kein Twister ist ein Wobbler. c) Jeder Twister ist ein Wobbler. d) Einige Twister sind Wobbler. e) Alle Twister sind keine Wobblers.

Zusammenfassend kann für den Ausschluss der Items 4, 8 und 9 das interne Teilungskriterium Extended Syllogismen Score angeführt werden, während für Item 5 das externe Teilungskriterium Ausbildung keine Rasch-Modell-Konformität ergab. Im Speziellen weist Item 4 sowohl in der ersten als auch in der zweiten Prämisse sinnvolle Begriffe auf, während Item 8 in beiden Prämissen vollständig aus frei erfundenen Worten ohne Realitätsbezug zusammengesetzt ist. Diese beiden Items wiesen einen höheren Schwierigkeitsschätzer im internen Teilungskriterium Extended Syllogismen Score für Personen mit einem niedrigeren Fähigkeitsniveau auf. Demgegenüber ist Item 9 zum Teil aus sinnvollen Begriffen konstruiert und zeigte bei Personen mit einem höheren Score im Teilungskriterium Extended Syllogismen Score einen höheren Schwierigkeitsparameter. Item 5 erfordert die Kenntnis von Fremdwörtern und wies für Personen mit niedrigerem Ausbildungsniveau eine höhere Schwierigkeitsparameterschätzung auf.

## 4.2 Korrelative Zusammenhänge der Reasoning-Tests

Die diskriminante Validität der drei neuen Tests wurde jeweils anhand des Koeffizienten der Rangkorrelation nach Spearman geprüft. Die untersuchten Verfahren FA, FRRT und Extended Syllogismen (ES) erheben den Anspruch, unterschiedliche Konstrukte des Reasoning-Faktors zu erheben. Da sich alle drei Tests als Rasch-Modell-konform herausstellten, konnte hierzu die Anzahl der gelösten Aufgaben, als erschöpfende Statistik für die Fähigkeit in den einzelnen Verfahren, verwendet werden.

Beim Verfahren Extended Syllogismen wurden nur jene 16 mit dem Rasch-Modell-konformen Items (ohne Item 4, 5, 8 und 9) herangezogen. Die untersuchten Verfahren korrelierten, wie nachfolgend die Tabelle 12 zeigt, positiv. Der Korrelationskoeffizient zwischen FA und ES erreichte  $r_s = .350$  (Bestimmtheitsmaß  $R^2 = .123$ ). Demnach werden ca. 12% der Varianz der Leistung in Extended Syllogismen durch Figurale Analogien erklärt.

FRRT und ES korrelierten zu  $.375$  ( $R^2 = .141$ ). Die Leistung in Family-Relation-Reasoning erklärt demnach etwa 14% der Varianz jener in Extended Syllogismen.

Zwischen FA und FRRT lag der Rangkorrelationskoeffizient nach Spearman  $r_s$  bei  $.298$  ( $R^2 = .089$ ). Demnach werden etwa 9% der Varianz der Leistung in Figurale Analogien durch jene in Family-Relation-Reasoning erklärt. Das Ausmaß der Korrelationen weist auf sehr schwache Zusammenhänge hin. Unter Berücksichtigung der erklärten Varianzanteile, die zwischen 8,8% und 14,1% liegen, kann von einer konvergenten Validität nicht ausgegangen werden.

Für die Zusammenhänge zwischen den Verfahren war gemäß der Theorie der fluiden und kristallisierten Intelligenz von Cattell (1963) eine bestimmte Reihenfolge der Korrelationskoeffizienten zu den unterschiedlichen Konstrukten anzunehmen. Die zu erwartende Reihung der Korrelationen zwischen den Verfahren war  $r_s(\text{FA, FRRT}) > r_s(\text{FRRT, ES}) > r_s(\text{FA, ES})$  und konnte anhand der beobachteten Werte  $r_s(\text{FRRT, ES}) = .375$ ;  $r_s(\text{FA, ES}) = .350$ ;  $r_s(\text{FA, FRRT}) = .298$  in dieser Form nicht bestätigt werden. Gemäß Literatur (Campbell & Fiske, 1959) wären zudem auch höhere Korrelationen zu erwarten gewesen.

Im Rahmen der Analyse der korrelativen Zusammenhänge zwischen den Reasoning-Tests FA und FRRT war dies auch zu erwarten, weil sie denselben Intelligenzbereich erfassen. Für die gegenständliche Studie wurde entsprechend, in Anlehnung an die vorangehende Arbeit von Ünal (2014), ein Grenzwert von  $.40$  festgelegt. Im Vergleich dazu sollten jene Verfahren, die unterschiedliche Konstrukte (fluid vs. kristallisiert) erfassen, eine Korrelation maximal so hoch wie dieser Grenzwert aufweisen. Diese theoretischen Vorannahmen lehnen sich an Studien von Cattell (1987) an. Es konnte bei Untersuchungen zu unterschiedlichen

Intelligenzkonzepten und -verfahren belegt werden, dass die Faktoren *fluide* und *kristallisierten Intelligenz* zwischen .40 bis .50 positiv korrelieren (Cattell, 1987, S. 116). Cattells Annahme hierzu war, dass die Entwicklungen der kristallisierten Intelligenz teilweise von der fluiden abhängig sind und sich demnach die Korrelation erklären lässt.

Zusammenfassend ist festzuhalten, dass unter Berücksichtigung der sich überschneidenden Konfidenzintervalle der Korrelationskoeffizienten keine bestimmte Reihenfolge, wie vorab postuliert, beobachtet werden konnte. Zudem wurde der festgelegte Grenzwert der Korrelationsstärken von .40 nur ansatzweise erreicht.

Die Tabelle 12 zeigt die in der vorliegenden Studie beobachteten Interkorrelationen der untersuchten Verfahren.

Tabelle 12 *Interkorrelationen für die Skalenzusammenhänge gemäß Koeffizient nach Spearman  $r_s$  mit 95% - Konfidenzintervall und erklärtem Varianzanteil  $R^2$  ( $n = 297$ )*

Verfahren	$r_s$		95% KI [UG; OG]	$R^2$
FA x Extended Syllogismen	.350	sign.	.233; .442	12,3%
FRRT x Extended Syllogismen	.375	sign.	.273; .471	14,1%
FA x FRRT	.298	sign.	.186; .396	8,8%

UG = untere Grenze, OG = obere Grenze,  $R^2$  = Bestimmtheitsmaß

## 5 Diskussion und Ausblick

Das Ziel der vorliegenden Arbeit war unter anderem eine erneute Durchführung von Rasch-Modell-Analysen für das Verfahren Extended Syllogismen (Treiber & Kubinger, in Vorb.). Ebenfalls wurde die diskriminante Validität zu zwei weiteren Verfahren, dem Family-Relation-Reasoning-Test (Kubinger, Poinstingl, Weidinger & Skoda, in Vorb.) unter Berücksichtigung der Figuralen-Analogien (Ünal & Kubinger, in Vorb.) erfasst.

Nach Ausschluss von vier Items, aus insgesamt 20 Items im Extended-Syllogismen-Test, konnte a posteriori Modellgültigkeit angenommen werden. Für die Items 4 und 8 konnte ein höherer Schwierigkeitsparameter für Personen mit einem niedrigeren Syllogismen-Score angenommen werden, während für Item 9 ein höherer Schwierigkeitsparameter für Personen mit einem höheren Syllogismen-Score zeigte. Item 5 wiederum fiel Personen mit einer höheren Ausbildung leichter. Eine inhaltliche Begründung für die fehlende Rasch-Modell-Konformität der auszuschließenden Items könnte für das Item 5 aufgrund seiner Position in den beiden Testheftversionen argumentiert werden: Dieses Item war im Vergleich zu den anderen auszuschließenden Items jeweils früher zu bearbeiten und ein Reihenfolgeeffekt ist nicht auszuschließen. Warum die Items 4, 8 und 9 keine Rasch-Modell-Konformität erreichten, kann möglicherweise durch Konzentrationseffekte bei sorgfältigem Durcharbeiten, oder wie bereits in Bezug auf Item 5 angeführt, an der (inhaltlichen) Itemposition erklärt werden. Diese Annahme kann mit dem *Teilungskriterium Testform* alternativ noch überprüft werden.

Des Weiteren wurden zwei weitere Tests vorgegeben, der Family-Relation-Reasoning-Test und Figurale-Analogien. Für den Family-Relation-Reasoning-Test konnte von Beginn an für alle vorgegebenen Items, insgesamt 20 Items, Modellgültigkeit festgestellt werden. Aus der aktuellen Untersuchung kann eine weitere Absicherung der Rasch-Modell-Gültigkeit in Bezug auf die vorgegebenen Items, in dem noch unveröffentlichten Test, ermittelt werden.

Die diskriminante Validität wurde durch Ermittlung der Rangkorrelationskoeffizienten in den Leistungsbereichen der drei Verfahren zum *Reasoning* berechnet.

Die beiden Verfahren zur kristallisierten Intelligenz, Family-Relation-Reasoning-Test und Figurale-Analogien, korrelieren zu  $r_s = .298$ . Die gleiche Fähigkeit wird somit erfasst, obwohl sich das Aufgabenmaterial in lexikalisch und figural unterscheidet. Das Bestimmtheitsmaß von  $R^2 = 8,8\%$  weist darauf hin, dass die beiden Verfahren das Konstrukt der kristallisierten Intelligenz verschieden erfassen. Der Zusammenhang zwischen den beiden Tests Family-

Relation-Reasoning-Test und Extended Syllogismen beträgt  $r_s = .375$  mit einem Bestimmtheitsmaß von  $R^2 = 14,1\%$  erklärter Varianz.

Der Family-Relation-Reasoning-Test soll kristallisierte Intelligenz und der Extended-Syllogismen-Test fluide Intelligenz erfassen. Der niedrige Korrelationskoeffizient weist demnach darauf hin, dass tatsächlich unterschiedliche Bereiche der Intelligenz gemessen werden. Der Zusammenhang zwischen Figurale-Analogien und Extended-Syllogismen beträgt  $r_s = .350$  mit einem Bestimmtheitsmaß von  $R^2 = 12,3\%$ . Die zu erwartende Reihung der Korrelationen zwischen den drei Verfahren gemäß Cattell (1963) konnte anhand der aktuell beobachteten Werte nicht bestätigt werden.

Da im Rahmen der vorliegenden Arbeit eine anfallende Stichprobe, bestehend aus auszubildenden Personen der Gesundheits- und Krankenpflege, ohne Zufallsziehung aus der Grundgesamtheit akquiriert wurde, ist die eingeschränkte Repräsentativität als Limitation zu betrachten. Die untersuchte Stichprobe entspricht mit Anteilswerten zum Geschlecht (rund 70% Frauenanteil) nicht der Verteilung in der Population. Die Quoten der höchst abgeschlossenen Ausbildung der Stichprobe entsprechen ebenso nicht jenen der Gesamtbevölkerung in Österreich, die im Speziellen einen höheren Akademikeranteil aufweist. Somit sind nicht nur zufällige Fehler nicht auszuschließen, sondern es kann auch ein systematischer Bias vorliegen. Die Erhebungen fanden im Rahmen eines Gruppensettings in Gesundheits- und Krankenpflegeschulen statt, wobei auch die Freiwilligkeit der Teilnahme anzuführen ist. Anzumerken ist ebenfalls, dass die Testungen im Rahmen des Unterrichts zumeist am Ende des Schultages stattfanden und demnach Ermüdungs- und Demotivationseffekte nicht auszuschließen sind. Demgegenüber konnte während der Bearbeitung der Testhefte ein im Allgemeinen leistungsbemühtes Verhalten der Teilnehmenden beobachtet werden.

Als Ausblick für weiterführende Arbeiten und nachfolgende Studien ist die Berücksichtigung der Weiterentwicklungen des einparametrig-logistischen Testmodells nach Rasch zu empfehlen. Bei weiteren Untersuchungen zur Geltung des Rasch-Modells besteht die Möglichkeit des Adaptiven Testens, in Form des *tailored testing* (Kubinger, 1996), welche als besonders ökonomische Form des Testens gesehen werden kann.

## 6 Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurde die Eindimensionalität des Verfahrens Extended-Syllogismen mit dem Rasch-Modell überprüft. Es konnte a posteriori Modellgültigkeit nach dem Ausschluss von vier Items erreicht werden. Jene Items mussten aufgrund ihrer Auffälligkeit in den Modellkontrollen ausgeschlossen werden. Die Itemselektion kann möglicherweise aufgrund von Konzentrationseffekten oder der (inhaltlichen) Itemposition erklärt werden und kann alternativ mit dem *Teilungskriterium Testform* noch überprüft werden.

Die Erhebungen wurden im Rahmen von Gruppentestungen in den Unterrichtsräumlichkeiten von Gesundheits- und Krankenpflegeschulen durchgeführt. Neben einem Datenblatt zur Erfassung der demographischen Daten bestand die Testbatterie aus den drei Reasoning-Tests sowie der Kurzversion des Anstrengungsvermeidungstests für ältere Schüler, dessen Ergebnisse als externes Teilungskriterium innerhalb der Rasch-Modell-Analysen herangezogen wurden. Es wurden neben dem Test Extended-Syllogismen, welcher schlussfolgerndes Denken im Bereich der fluiden Intelligenz erfassen soll, die Verfahren Figurale-Analogien und der Family-Relation-Reasoning-Test vorgegeben.

Der Zusammenhang zwischen FRRT und ES mit  $r_s = .375$  zeigte sich als etwas höher als die restlichen Korrelationen des FA und ES mit einem Korrelationskoeffizienten von  $r_s = .350$  und einem  $r_s = .298$  zwischen FRRT und FA. Diese Reihung der Zusammenhänge entspricht nicht der theoriegeleiteten Annahme. Zu erwarten wäre gewesen, dass Figurale-Analogien und der Family-Relation-Reasoning-Test am stärksten korrelieren, da beide Verfahren kristallisierte Intelligenz erfassen und sich lediglich das Aufgabenmaterial in figural und lexikalisch unterscheidet.

Als mögliche Erklärung wäre denkbar, dass zur Lösung der Items im Family-Relation-Reasoning Test auch Anteile des Teilbereiches der fluiden Intelligenz benötigt werden. Es kann hier nicht nur auf vergangene, umweltspezifische Lernerfahrungen zurückgegriffen, sondern es wird abstraktes Schlussfolgern gefordert. In der vorliegenden Arbeit wurde ein Grenzwert von .40 vorab festgelegt, womit die diskriminante Validität der drei Verfahren insgesamt als gegeben angesehen werden kann. Die drei Verfahren unterscheiden sich demnach in den zu erfassenden Konstrukten der fluiden und kristallisierten Intelligenz und erfassen diese auf unterschiedliche Art und Weise.

## 7 Literaturverzeichnis

- Anastasi, A. (1958). *Differential Psychology*. New York: MacMillan.
- Andersen, E.B. (1973). Conditional Inference for Multiple-Choice Questionnaires. *Br. J. math. statist. Psychol.*, 26, 31-44.
- Andrich, D. (1978). Application of a psychometric rating model to ordered categories with are scored with successive integers. *Applied Psychological Measurement*, 38, 665-680.
- Anzenbacher, A. (1985). *Einführung in die Philosophie*. Wien: Herder.
- Beckmann, J. F., & Guthke, J. (1999). *Psychodiagnostik des schlußfolgernden Denkens: Handbuch zur Adaptiven Computergestützten Intelligenz-Lerntestbatterie für Schlußfolgerndes Denken (ACIL)*. Göttingen: Hogrefe.
- Birnbaum, A. (1968). Some latent trait models and their use in interferred an examinee's ability. In F.M. Lord & M.R. Novick (eds.), *Statistical theories of mental test scores* (pp. 395-479). Reading/Mass.: Addison-Wesley.
- Brentano, F. (1956). *Die Lehre vom richtigen Urteil: nach den Vorlesungen über Logik mit Benützung anderer Manuskripte aus dem Gebiet der Erkenntnistheorie*. Bern: Francke.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the Multitrait Multimethod Matrix. *Psychological Bulletin Volume 56*(2), S. 81105.
- Cattell, R. B. (1957). Fluid and Crystallized Intelligence. In J. J. Jenkins & D. G. Paterson (Eds.), *Studies in individual differences. The search for intelligence* (pp. 738-746). London: Methuen & Co.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54 (1), 1-22.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: Elsevier
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Hans Huber.
- Formann, A. K. & Piswanger, K. (1979). *Wiener Matrizen-Test. WMT*. Weinheim: Beltz-Test.
- Gashi, A. (2011). *Untersuchung über Beziehungen zwischen Anstrengungsvermeidung, Selbststeuerung und Lernemotionen bei Jugendlichen aus psychologischer Sicht*. Diplomarbeit, Universität Wien.
- Hofstätter, P. R. (1977). *Persönlichkeitsforschung*. Stuttgart: Alfred Kröner.



- 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.
- Jäger, A. O. (1973). *Dimensionen der Intelligenz* (3. Auflage). Göttingen: Hogrefe.
- Kubinger, K.D. (1996). Methoden der Psychologischen Diagnostik. In E. Erdfelder, R. Mausfeld, T. Meiser & G. Rudinger (Hrsg.), *Handbuch Quantitative Methoden* (S.567-576). Weinheim: Beltz.
- 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. überarbeitet und erweiterte Aufl. Göttingen: Hogrefe.
- Kubinger, K. D., Pointstingl, H., Weidinger, G., & Skoda, S. (in Vorb.). FamilyRelation-Reasoning-Test (FRRT).
- Kubinger, K. D., Rasch, D. & Yanagida, T. (2011a). *Statistik in der Psychologie: Vom Einführungskurs bis zur Dissertation*. Göttingen: Hogrefe.
- Kubinger, K. D., & Ünal, S. (in Vorb.). Sprachliche Analogien.
- Mair, P., Hatzinger, R. & Maier M. J. (2015). *eRm: Extended Rasch Modeling*. 0.15-6. <http://erm.r-forge.r-project.org/>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, 38(1), 115-142.
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2. Auflage). Berlin: Springer-Verlag.
- Pointstingl, H., Mair P., & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modeling) Anwendung des Rasch-Modells (1-PL Modell) - deutsche Version*. Wien.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Kopenhagen: Paedagogiske Institut.
- Rasch, G. (1961). On general laws and the meaning of measurement in Psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 4: *Contributions to Biology and Problems of Medicine*. Berkeley, Calif: University of California Press, 321-333.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests (Exp. ed.)*. University of Chicago, Chicago.

- Rollett, B. & Bartram, M. (1998). *Anstrengungsvermeidungstest (AVT)*. (3.Auflage). Göttingen: Hogrefe Verlag.
- Rollett, B. & Rollett, W. (2010). Anstrengungsvermeidung. In Rost, D. H. (Hrsg.), *Handwörterbuch der Pädagogischen Psychologie*. (S. 17-22). Weinheim: Beltz PVU.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (Zweite, vollständig überarbeitete und erweiterte Auflage). Bern: Verlag Hans Huber.
- Skoda, S. (2005). *Der Verwandtschaften-Reasoning-Test*. Konzept und erste empirische Erprobung. Unveröffentlichte Diplomarbeit, Universität Wien
- Spearman, C. (1904). "General Intelligence" objectively determined and measured. *The American Journal of Psychology*, 15 (2), 201-292.
- Spearman, C. (1914). The theory of two factors. *Psychological Review*, 21(2), 101-115.
- Spearman, C. (1927). *The Abilities of Man*. London: MacMillan & Co.
- Srp, G. (1994). *Syllogismen*. Frankfurt am Main: Swets Test Service.
- Stemmler, G., Hagemann, D., Amelang, M., & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung*. 7. überarb. Aufl. Stuttgart: Kohlhammer.
- Stern, W. (1912). *Die psychologischen Methoden der Intelligenzprüfung und deren Anwendung an Schulkindern*. 5. Kongress der Experimentellen Psychologie, Berlin.
- Schüling, H. (2002). *Die Genesis des schlussfolgernden Denkens*. Hildesheim: Georg Olms.
- Süß, H.-M. (2003). 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*. Chicago: University of Chicago.
- Treiber, S. (2010). *Konstruktion eines neuen Syllogismen-Tests und dessen erste testtheoretische Prüfung anhand des Rasch-Modells*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Ünal, S. (2014). *Figurale Analogien: Ein Test zur Messung des kristallisierten schlussfolgernden Denkens mit Hilfe figuralen Materials*. Unveröffentlichte Diplomarbeit, Universität Wien, Wien.
- Ünal, S., & Kubinger, K. D. (in Vorbereitung). *Figurale Analogien*.
- Waldmann, M., & Weinert, F. E. (1990). *Intelligenz und Denken: Perspektiven der Hochbegabungsforschung*. Göttingen: Hogrefe.
- Wechsler, D. (1956). *Die Messung der Intelligenz Erwachsener*. Bern: Hans Huber.

## 8 Abbildungsverzeichnis

<i>Abbildung 1.</i> Wege der Erkenntnisgewinnung (modifizierte Darstellung).....	7
<i>Abbildung 2.</i> Beispielitem des Family-Relation-Reasoning-Test (FRRT) .....	16
<i>Abbildung 3.</i> Beispielsitem für Figurale Analogien .....	17
<i>Abbildung 4.</i> Beispielitem für Extended Syllogismen.....	18
<i>Abbildung 5.</i> Beispielitems für Anstrengungsvermeidung (AV) und Pflichteifer (PE).....	19
<i>Abbildung 6.</i> Verteilung des Lebensalters in Jahren bezüglich des Geschlechts der Teilnehmenden.....	22
<i>Abbildung 7.</i> Verteilung der AVT-Rohwerte .....	23
<i>Abbildung 8.</i> Verteilung der FRRT-Rohwerte.....	23
<i>Abbildung 9.</i> Verteilung der Extended Syllogisms-Rohwerte .....	24
<i>Abbildung 10.</i> Graphische Modellkontrolle, internes Teilungskriterium FRRT-Score .....	27
<i>Abbildung 11.</i> Graphische Modellkontrolle, externes Teilungskriterium Bildung.....	28
<i>Abbildung 12.</i> Graphische Modellkontrolle, externes Teilungskriterium AVT-Median .....	28
<i>Abbildung 13.</i> Graphische Modellkontrolle, externes Teilungskriterium Erstsprache .....	29
<i>Abbildung 14.</i> Graphische Modellkontrolle, externes Teilungskriterium Geschlecht .....	29
<i>Abbildung 15.</i> Graphische Modellkontrolle, externes Teilungskriterium Altersmedian .....	30
<i>Abbildung 16.</i> Graphische Modellkontrolle, internes Teilungskriterium ES-Score .....	32
<i>Abbildung 17.</i> Graphische Modellkontrolle, internes Teilungskriterium Altersmedian .....	33
<i>Abbildung 18.</i> Graphische Modellkontrolle, externes Teilungskriterium AVT-Median .....	33
<i>Abbildung 19.</i> Graphische Modellkontrolle, externes Teilungskriterium Geschlecht .....	34
<i>Abbildung 20.</i> Graphische Modellkontrolle, externes Teilungskriterium Erstsprache .....	34
<i>Abbildung 21.</i> Graphische Modellkontrolle, externes Teilungskriterium Ausbildung .....	35
<i>Abbildung 22.</i> Graphische Modellkontrolle, externes Teilungskriterium Ausbildung .....	36
<i>Abbildung 23.</i> Graphische Modellkontrolle, internes Teilungskriterium ES-Median .....	36
<i>Abbildung 24.</i> Graphische Modellkontrolle, internes Teilungskriterium ES-Median .....	37
<i>Abbildung 25.</i> Graphische Modellkontrolle, externes Teilungskriterium Ausbildung .....	38
<i>Abbildung 26.</i> Graphische Modellkontrolle, externes Teilungskriterium Ausbildung (Item 5 Ausbildung, $p = .005$ ).....	39
<i>Abbildung 27.</i> Graphische Modellkontrolle, internes Teilungskriterium ES-Median .....	39
<i>Abbildung 28.</i> Graphische Modellkontrolle, externes Teilungskriterium Ausbildung .....	40

## 9 Tabellenverzeichnis

Tabelle 1 Überblick zu den Konstrukten und Aufgabenmaterialien der angewendeten Verfahren.....	21
Tabelle 2 Vorgabereihenfolge der Verfahren gemäß Testheftversion .....	21
Tabelle 3 Matrix der rphi-Koeffizienten mit Signifikanzbeurteilungen ( $p$ ) zu den Kreuztabellen der dichotomen Teilungskriterien (FRRT) .....	26
Tabelle 4 Internes und externe Teilungskriterien für den FRRT .....	26
Tabelle 5 rphi-Koeffizienten mit Signifikanzbeurteilungen ( $p$ ) zu den Kreuztabellen der dichotomen Teilungskriterien (ES) .....	31
Tabelle 6 Internes und externe Teilungskriterien für ES.....	31
Tabelle 7 Internes und externe Teilungskriterien für ES nach Ausschluss von Item 4.....	35
Tabelle 8 Internes und externe Teilungskriterien für ES nach Ausschluss der Items 4 und 8.....	37
Tabelle 9 Internes und externe Teilungskriterien für ES nach Ausschluss der Items 4, 8 und 9.....	38
Tabelle 10 Internes und externe Teilungskriterien für ES nach Ausschluss der Items 4, 5, 8 und 9....	40
Tabelle 11 Im Rahmen der Analyse nach dem Rasch-Modell als nicht-konform, resultierende und daher ausgeschlossene Items des Extended-Syllogismen-Tests .....	41
Tabelle 12 Interkorrelationen für die Skalenzusammenhänge gemäß Koeffizient nach Spearman $r_s$ mit 95% - Konfidenzintervall und erklärtem Varianzanteil $R^2$ ( $n = 297$ ) .....	43

## 10 Abkürzungsverzeichnis

Abkürzung (Akronym)	Beschreibung
AVT	Anstrengungs-Vermeidungs-Test
AV	Anstrengungsvermeidung
ES	Extended Syllogismen
FA	Figurale Analogien
FRRT	Family-Relation-Reasoning-Test
IC	Item-Charakteristik
IRT	Item-Response-Theory
IQR	Innerquartilabstand
KI	Konfidenz-Intervall
LRT	Likelihood-Ratio-Test
LQR	Likelihood-Quotienten-Test
M	Mittelwert ( <i>Mean</i> )
PE	Pflichteifer
PCM	Partial Credit Model
$R^2$	Bestimmtheitsmaß (erklärter Varianzanteil)
RSM	Rating Scale Model
SD	Standardabweichung ( <i>Standard Deviation</i> )

## 11 Anhang



### **Genehmigung für Interviews & Umfragen in Einrichtungen des Wiener Krankenanstaltenverbandes**

Wiener Krankenanstaltenverband  
Generaldirektion  
Geschäftsbereich Pflegemanagement  
Bürocarier TownTown  
Thomas-Masch-Platz 31b  
1030 Wien  
Tel.: +43 1 43422 70111  
Fax: +43 1 43422 55 10111  
E-Mail: [post@ed\\_gpm@wienkav.at](mailto:post@ed_gpm@wienkav.at)  
[www.wienkav.at](http://www.wienkav.at)

1. Die Teilnahme der Befragten muss stets auf freiwilliger Basis erfolgen.
2. Die Bestimmungen des Datenschutzes sind einzuhalten. Die im Rahmen Ihres Studiums unterzeichnete Datenschutzverpflichtungserklärung besitzt auch für die Genehmigung Ihres Ansuchens Gültigkeit.
3. Für persönliche Befragungen von PatientenInnen bzw. MitarbeiterInnen/SchülerInnen sind schriftliche Zustimmungen (Einverständniserklärungen) einzuholen. In der Einverständniserklärung muss der/die Betroffene explizit auf das Recht auf Widerruf hingewiesen werden.
4. Der Dienstbetrieb darf durch die Befragung nicht beeinträchtigt werden.
5. Die Anonymität der Befragten muss gewährleistet sein bzw. kann die Verwendung der Namen nur mit schriftlicher Zustimmung der Betroffenen erfolgen.
6. Die fertige wissenschaftliche Arbeit (Bachelor-, Master- oder Diplomarbeit) ist dem Wiener Krankenanstaltenverband, Geschäftsbereich Pflegemanagement vor einer Veröffentlichung zur Einsicht vorzulegen.

Oben stehende allgemein gültige Punkte sind vor Beginn des/der geplanten Projektes/Umfrage/Interviews im Wiener Krankenanstaltenverband von dem/der AntragstellerIn verpflichtend zur Kenntnis zu nehmen und deren Einhaltung durch Unterzeichnung zu bestätigen:

NAME des/der AntragstellerIn:.....

DATUM:.....

TITEL des Ansuchens:.....

THEMA:.....

.....  
Unterschrift



universität  
wien

Kerstin Grafi  
Redtenbachergasse 22-32/1/10  
1160 Wien

Tel.: 0699/19820822  
Mail: [kerstin.grafi@gmail.com](mailto:kerstin.grafi@gmail.com)

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

#### **Datenerhebung im Rahmen der Diplomarbeit:**

**Titel: „Tripelweiser Dimensionalitätsvergleich von Reasoningtests“**

Im Rahmen der Diplomarbeit am Arbeitsbereich Psychologische Diagnostik an der Fakultät für Psychologie der Universität Wien beschäftigen wir uns mit der Analyse und dem Vergleich psychologischer Verfahren die schlussfolgerndes Denken erfassen.

Bei dem vorzuziehenden Testmaterial handelt es neben im deutschen Sprechraum etablierten Tests um einige neu entwickelte psychologische Verfahren zur Erfassung der Fähigkeit zum logischen Schlussfolgern. Es soll untersucht werden, inwieweit diese Verfahren auch tatsächlich die zu erfassende Fähigkeit messen und inwiefern sich die verschiedenen Verfahren ähneln bzw. unterscheiden. Damit soll auch ein Beitrag zur Überprüfung der Validität geleistet werden.

Wir untersuchen daher um Unterstützung bzw. Mitwirkung an unserem Projekt. Die Studie koordinieren wir von Wien aus (Betreuung: Univ.-Prof. Mag. Dr. Klaus Kubinger).

Die Testungen werden als „Gruppentestung“ durchgeführt. Die Bearbeitungszeit dauert ca. 60 Minuten (bis maximal 80 Minuten) pro Gruppe / Klasse. Wir sind dabei bemüht, den Aufwand für die jeweiligen Personen und Institutionen so gering wie möglich zu halten.

Die teilnehmenden Personen tragen ganz wesentlich zur Sicherung der Qualität psychologischer Leistungsdiagnostik bei, welche eine optimale Beratung hinsichtlich leistungsbezogener Aspekte und Fähigkeiten gewährleisten soll.

Es besteht die Möglichkeit, sollte dies gewünscht sein, den teilnehmenden Personen als Dankeschön für die Teilnahme einen kurzen schriftlichen Ergebnisbericht über ihre kognitiven Stärken und Schwächen auszubändigen.

### **Testablauf und Untersuchungsziel:**

Da die Testbatterie jedem/r zu testenden Person schriftlich vorgegeben wird, ist es möglich, die Aufgaben gleichzeitig einer Gruppe von Personen vorzulegen. Es handelt sich um eine Gruppentestung. Um adäquate Aussagen hinsichtlich der Güte der zu analysierenden Tests treffen zu können, ist es nötig, dass jede Person einen kurzen Fragebogen und die drei unten angeführten Verfahren bearbeitet.

Es werden nur Personen getestet, die sich freiwillig dazu bereiterklären. Die Testung ist anonym. Insgesamt sind 300 Personen (Schüler und Schülerinnen) im Alter zwischen 16 und 26 Jahren für die Erhebung der Daten vorgesehen.

Bei den Testungen werden wir von anderen Diplomand/Innen des Arbeitsbereichs Psychologische Diagnostik unterstützt. Die gewonnenen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke verwendet und anonymisiert.

### **Informationen zur Testbatterie:**

**Figurale Analogien** (Ünal&Kubinger, in Vorb.): Mit diesem Verfahren wird Reasoning sowie Allgemeinwissen mithilfe von inhaltlich sinnvollen Bildern erhoben. Zwischen jeweils zwei Bildern soll eine Beziehung erkannt werden, die sodann auf zwei weitere Bilder zu übertragen und das fehlende Bild auszuwählen ist. (Testdauer ca. 10 Minuten)

**Syllogismen** (Treiber, in Vorb.): Dieser Test stellt ein Verfahren zur Erfassung schlussfolgernden Denkens im verbalen Bereich dar, wobei aus vorgegeben Aussagen logische Schlüsse abgeleitet werden sollen. (Testdauer ca. 10 Minuten)

**Family-Relation-Reasoning Test** (Kubinger, Poinstingl, Weidinger&Skoda): Dieser Test dient zur Messung des kristallisierten, schlussfolgernden Denkens mit Hilfe lexikalischen Materials. (Testdauer ca. 10 Minuten)

**Fragebogen zum Anstrengungsvermeidungsverhalten** (Rollet&Bertram): Dabei handelt es sich um ein Raschmodell - konformes Fragebogenverfahren zur Erfassung von negativ gerichteten motivationalen Tendenzen von Personen. Die Fragen umfassen die gängigsten „Ausreden“ von Personen, um sich einer Lern – bzw. Leistungssituation zu entziehen und können den Beiden Skalen Anstrengungsvermeidung und Pflichteifer zugeordnet werden. (Testdauer ca. 15 Minuten)

Diese Untertests werden zusammen mit dem Anstrengungsvermeidungsfragebogen in einem gesamten Testheft bei der Testung pro Person in der Gruppe ausgehändigt. Anwesend bei der Testung ist eine Testleiterin (Kerstin Grafi), die die Informationen zur Bearbeitung gibt und die Instruktionen erklärt. Weiters ist eine Hilfsperson (Diplomandin) dabei, die beim Austeilen und Einsammeln der Testhefte und Stoppen der Testzeit Unterstützung gibt.

Die Testdauer gesamt ergibt ca. 1 Stunde pro Gruppe/Klasse. Die organisatorischen Rahmenbedingungen, sowie die Absprachen und Koordination mit den einzelnen Direktoren/Innen werden von der Diplomandin und der Hilfsperson übernommen.

Die Auswertung findet anhand einer Raschmodellanalyse mittels dem Statistikprogramm „R“ statt. Die Daten werden auf Raschmodell Konformität untersucht, ob die Aufgaben jeweils das messen, was sie vorgeben zu messen.

Wie oben bereits erwähnt, gibt es die Möglichkeit für die teilnehmenden Personen ihre Leistungsergebnisse, anonym mittels „Testpersonencode“ anzufordern und eine Ergebnismeldung zu erhalten.



Probandencode:

Für Ergebnismeldung bitte Mail an folgende Adresse:

[kerstin.grafl@gmail.com](mailto:kerstin.grafl@gmail.com)

Im E-Mailbetreff bitte Personencode eintragen. Die Rückmeldung kann einige Monate dauern.  
Bitte um Ihr Verständnis.

DANKE für die Teilnahme.

1. **Einleitungstext/Testleiterinstruktion (siehe S.3)**

2. Es ist jedem/jeder Schüler(in) eine **Testbatterie** auszuteilen. Beim Austeilen ist darauf zu achten, dass die Testhefte genauso liegen gelassen werden, wie sie ausgeteilt wurden und nicht umhergeblättert wird.

3. Die Testbatterie besteht aus 3 Untertest, dem AVT (Fragebogen) und einem Blatt für die demographischen Daten in folgender **Reihenfolge**:

- Demographisches Datenblatt
- Extended Syllogisms
- Figurale Analogien
- Family Relation Reasoning Test
- AVT

Die Testhefte Extended Syllogism und Figurale Analogien liegen in Form A + B vor. Der Untertest Family Relation Reasoning Test in Form A, B + C. Der AVT in einer Form.

SchülerInnen, die nebeneinander sitzen, sollten jeweils eine unterschiedliche Testform erhalten. Das heißt die Testhefte werden abwechselnd an die SchülerInnen ausgeteilt.

Die Tests sind vor der Testung in der richtigen Reihenfolge sowie auch abwechselnd nach Testform (siehe Kästchen) zu sortieren, um ein einfaches Austeilen zu gewährleisten.

SchülerIn	Reihenfolge der Testhefte
1.	0 – 1A – 2A – 3A – AVT
2.	0 – 1B – 2B – 3B – AVT
3.	0 – 1A – 2A – 3C – AVT
4.	0 – 1B – 2B – 3A – AVT
5.	0 – 1A – 2A – 3B – AVT
6.	0 – 1B – 2B – 3C – AVT
7.	Wieder wie 1, usw. ...

4. Zeitpunkt des **Beginns der Testung** (ab Einleitungstext) und des **Endes der Testung** (sobald die Bearbeitung des letzten Tests vorbei ist) sollten bitte am **Testleiterprotokoll** notiert werden, wie auch **Datum, Schule und Klasse** sowie Namen der **Testleiter**.

5. Um den SchülerInnen **Rückmeldung über ihre Testergebnisse** geben zu können, sollen sie sich den **Zettel vorne** mit meiner E-Mail-Adresse und dem Probandencode mitnehmen.

6. Die SchülerInnen sollen einen **Stift** bereitlegen. Sollten sie eine bereits markierte Antwort ändern wollen, so sollen sie bitte die **falsche Lösung durchstreichen** und die neu gewählte Antwort wieder ankreuzen und **deutlich markieren** (z.B.: durch Einkreisen)

7. Wenn eine Instruktion nicht gleich verstanden wird (z.B.: Beispielaufgabe) oder Fragen dazu sind, so wird diese bitte nochmals langsam wiederholt oder mit anderen Worten erklärt.
8. Für **jeden Untertest** gibt es eine **feste Bearbeitungszeit**, die in der Testleiteranweisung angeführt ist. Nach Ablauf der Bearbeitungszeit sind die Testpersonen jeweils zu bitten, einen **Strich unter die zuletzt bearbeitete Aufgabe zu ziehen**.
9. Sollten **vor Ablauf der Bearbeitungszeit** alle Testpersonen bereits fertig sein, so kann schon früher zum nächsten Untertest übergegangen werden. Die Testpersonen sollten **aber bitte trotzdem einen Strich unter die zuletzt bearbeitete (in diesem Fall letzte) Aufgabe** als Kontrolle ziehen. Die tatsächlich benötigte Zeit wird dann auf dem Testleiterprotokoll notiert.
10. Sollten SchülerInnen nachfragen, ob sie bei Unsicherheit raten sollen oder die Aufgabe auslassen sollen, ist ihnen zu antworten, dass sie lieber raten sollen. Dies aber bitte nur auf Nachfrage sagen!
11. Bitte die SchülerInnen **animieren, die Aufgaben der Reihe nach zu bearbeiten**, das ist für die anschließende Auswertung sehr wichtig. Wenn eine Aufgabe aber viel zu schwierig ist, dann kann man schon sagen, dass sie dann die nächste Aufgabe versuchen sollen, weil auch wieder einfachere kommen. Aber es soll nicht animieren, zwischen den Testaufgaben „herumzuspringen“.
12. Bitte darauf achten, dass die SchülerInnen auch umblättern, wenn „**BITTE UMBLÄTERN**“ am Ende der Seite steht.
13. Am Ende eines Tests müssen die SchülerInnen warten, bis sie zum Umblättern aufgefordert werden. Bitte darauf achten, dass sie **nicht** zum nächsten Untertest oder während der Testinstruktion zu den Testaufgaben **weiterblättern, ohne dass dazu aufgefordert wurden**.
14. Es ist **darauf zu achten**, dass die SchülerInnen **nicht vom Nachbarn abschreiben** oder miteinander reden. Liegt der Verdacht nahe, dass ein/e SchülerIn abgeschrieben hat, so ist das am besten beim Absammeln am Testheft zu notieren.
15. Bitte eventuelle **Auffälligkeiten und Probleme** pro Untertest am **Testleiterprotokoll notieren**.
16. Den Testpersonen darf bei der Bearbeitung der Aufgaben nicht geholfen werden, auch keine Hinweise zur Lösung gegeben werden. Bitte auch keine inhaltlichen Fragen beantworten (z.B.: bei figuralen Analogien: was ein bestimmtes Bild darstellen sollte). Bei derartigen

Fragen bitte sagen, dass man diese Fragen leider nicht beantworten darf.

17. Es ist darauf zu achten, dass alle SchülerInnen die Testhefte wieder **abgeben** und nicht mit nach Hause nehmen. Bitte ebenfalls darauf achten, dass **keine Fotos** von den Tests gemacht werden (Datenschutz)

# Testleiter – Instruktion

## Einleitungstext und das Austeilen der Testhefte

*Ich wünsche euch einen guten Tag. Mein Name ist \_\_\_\_\_ Und das ist /sind meine Kollgen/Innen \_\_\_\_\_*

*Wir möchten euch herzlich zu unserer heutigen Testung **begrüßen** und uns vorweg dafür bedanken, dass ihr euch für die Testung im Rahmen meiner Diplomarbeit bereit erklärt habt teilzunehmen. Wie ihr von den Vorinformationen wisst, werdet ihr heute ein paar **Aufgaben bearbeiten**.*

*Ich möchte euch **vorweg kurz erklären**, wie die gesamte Testung ablaufen wird. Ihr werdet als erstes die Testhefte ausgeteilt bekommen. Ein **Testheft** besteht jeweils aus 3 Untertests, einem Fragebogen und einem demographischen Datenblatt, auf dem ihr bitte eure **persönlichen Daten** ausfüllt*

*Wenn ihr das **Testheft** bekommt, lasst es bitte noch vor euch liegen.*

*(Testleiter/In beginnt jetzt mit dem Austeilen)*

*Ich werde euch zu Beginn vor jedem einzelnen Untertest erklären, was bei den jeweiligen Aufgaben zu tun ist und anhand der Beispielaufgaben gemeinsam üben. Ihr habt dann eine jeweils **festgelegte Bearbeitungszeit für jeden Untertest** zu Verfügung, die ich euch immer vorher nennen werde. Mit der Bearbeitung der Testaufgaben bitte **immer erst beginnen, wenn ich es euch sage**. Bitte bearbeitet die **Aufgaben der Reihe nach**, das ist für die anschließende Auswertung der Daten sehr wichtig. Solltet ihr eine Antwort ändern wollen, dann streicht bitte die erste Antwort durch und markiert die neue Antwort deutlich durch z.B.: Einkreisen.*

***Nach Ablauf der Zeit** werde ich euch Bescheid geben, woraufhin ihr bitte aufhört zu arbeiten. Bitte immer nach Bearbeitung der **letzten Aufgabe**, einen **Strich** unter diese ziehen*

*Bitte achtet auch darauf, dass während der Bearbeitung des jeweiligen Untertest auf das **jeweilige „bitte umblättern“** geachtet wird und ihr selbstständig umblättert, um die darauffolgende Aufgabe zu lösen.*

***Arbeitet immer so zügig wie möglich, gewissenhaft und genau.***

***Habt ihr bisher alles verstanden? Gibt es noch Fragen?***

## Erklärung „Rückmeldungs-CODE“ - Blatt

*Ganz vorne findet ihr auf eurem Testheft einen **kleinen Zettel**, auf dem ihr **meine E-Mail-** Adresse und einen Code seht. Das ist der **gleiche Code** wie auf eurem Testheft. Ihr könnt euch den **Zettel mitnehmen** und wenn ihr eine Rückmeldung über eure Testergebnisse möchtet schreibt mir bitte eine E-Mail mit dem Code und ich schicke euch eure **Ergebnisse** sobald ich mit der Auswertung fertig bin. Dies kann jedoch ein einige Wochen dauern, bitte um euer Verständnis.*

*Die **Testung** ist **anonym**, das heißt in der Auswertung wird niemand persönlich nachvollziehbar sein und dient ausschließlich wissenschaftlichen Zwecken. Deshalb erhält ihr auch für die Rückmeldung einen „Code“.*

## Demographisches Datenblatt

*Bitte füllt jetzt zu Beginn die Daten zu eurer Person auf der ersten Seite aus. Wenn es von eurer Seite keine Fragen mehr gibt, können wir mit der Testung beginnen!*

## Untertest Extended Syllogisms

Wenn die SchülerInnen mit dem Ausfüllen der persönlichen Daten fertig sind, wird wie folgend instruiert:

*Wir beginnen jetzt mit der Bearbeitung der einzelnen Tests. Bitte blättert jetzt zum ersten Test „Extended Syllogisms“ um.*

Der/die TestleiterIn liest Folgendes vor, die SchülerInnen können im Testheft mitlesen:

*In dem folgenden Test werden Ihnen bei jeder Aufgabe Aussagen und mögliche Schlussfolgerungen angeboten. Sie müssen entscheiden, welche Schlussfolgerung aufgrund der beiden vorgegebenen Aussagen zwingend logisch richtig ist.*

*Beispielsweise führen die Aussagen „Alle A sind B. Alle B sind C“ zu der Lösung:*

*„Alle A sind C“. Da alle „A“ „B“ sind und alle „B“ „C“ sind, müssen alle „A“ auch „C“ sein.*

*Schaut euch bitte zuerst folgende Beispielaufgaben an und versuchen Sie diese zu lösen.*

**Jemanden aus der Klasse die Antwort nennen lassen!**

***B1: Jeder Baum ist eine Pflanze. Jede Birke ist ein Baum.***

- a) Einige Birken sind keine Pflanzen.*
- b) Jede Birke ist eine Pflanze.*
- c) Keine Birke ist eine Pflanze.*
- d) Einige Birken sind Pflanzen.*
- e) Alle Birken sind keine Pflanzen.*

***Die Lösung lautet: b) Jede Birke ist eine Pflanze***

***B2: Jedes Auto ist ein Fahrrad. Einige Kraftfahrzeuge sind Autos.***

- a) Alle Kraftfahrzeuge sind Fahrräder.*
- b) Einige Kraftfahrzeuge sind keine Fahrräder.*
- c) Einige Kraftfahrzeuge sind Fahrräder.*
- d) Kein Kraftfahrzeug ist ein Fahrrad.*
- e) Alle Kraftfahrzeuge sind keine Fahrräder.*

***Die Lösung lautet: c) Einige Kraftfahrzeuge sind Fahrräder.***

***Wie bei diesen Beispielen gibt es auch bei allen weiteren Aufgaben immer nur eine Lösung. Es ist möglich, dass Sie zu Lösungen kommen, die Ihnen absurd oder inhaltlich falsch vorkommen. Dies beruht darauf, dass man auch aus absurden Aussagen logisch richtige Schlüsse ableiten kann. Ob ein logischer Schluss richtig ist, hängt nicht von der „Wahrheit“ der Aussage in der realen Welt ab. Für die Bearbeitung stehen Ihnen 13 Minuten zur Verfügung. Falls Sie vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig sind, warten Sie bitte auf das Zeichen des Testleiters. Gibt es von Ihrer Seite noch Fragen?***

***LOS GEHT'S! BITTE UMBLÄETTERN!***

Nach 6 Minuten und 30 Sekunden ist den SchülernInnen mitzuteilen, dass die Hälfte der Bearbeitungszeit verbraucht ist.

Nach insgesamt 13 Minuten ist mit dem Wort „STOPP“ abubrechen.

Die SchülerInnen werden gebeten unter der zuletzt bearbeiteten Aufgabe einen Strich zu ziehen.

***Bitte macht einen Strich unter die Aufgabe, die Ihr zuletzt bearbeitet habt. Blättert jetzt bitte auf die nächste Seite zum Test Figurale Analogien!***

## **Untertest Figurale Analogien**

Der/die TestleiterIn liest vor:

*Auf den nächsten Seiten finden Sie 27 Aufgaben. Bei jeder Aufgabe werden Ihnen drei Bilder vorgegeben. Zwischen dem ersten und zweiten Bild besteht eine Beziehung. Eine ähnliche Beziehung besteht zwischen dem dritten und einem der fünf Wahlbilder. Dieses Bild sollen Sie herausfinden.*

*Sehen Sie sich bitte zuerst das folgende Beispiel an!*

**Beispielaufgabe B:**

*Die ersten zwei Bilder zeigen eine männliche und eine weibliche Person. Dieselbe Beziehung besteht auch zwischen Hahn und einer Henne. Daher ist „D“ die richtige Lösung.*

*Wie bei diesem Beispiel, markieren Sie bitte Ihre Antwort durch Ankreuzen des Buchstabens unter dem Bild. Beachten Sie bitte, dass es immer nur eine Lösung gibt.*

*Für die Bearbeitung stehen Ihnen 8 Minuten zur Verfügung. Falls Sie vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig sind, warten Sie bitte auf ein Zeichen des Testleiters.*

*Gibt es von eurer Seite noch Fragen?*

**LOS GEHT'S! BITTE UMBLAETTERN!**

Für die Bearbeitung stehen 8 Minuten zur Verfügung. Nach 4 Minuten ist den Schülern/Innen mitzuteilen, dass nur mehr die Hälfte der Zeit übrig ist.

Nach insgesamt 8 Minuten ist mit dem Wort „STOPP“ abzubrechen.

Die Schüler/Innen werden anschließend darum gebeten, im Testheft unter der zuletzt bearbeiteten Aufgabe einen Strich zu ziehen.

**Bitte macht nun einen Strich unter der von euch zuletzt bearbeiteten Aufgabe!**

Wenn alle unter der letzten Aufgabe einen Strich gezogen haben, wird wie folgt instruiert:

**Blättert jetzt bitte um auf die nächste Seite zum Family Relation Reasoning Test**



## Untertest Family Relation Reasoning Test

Der/die TestleiterIn liest vor:

*Auf den nächsten Seiten werden Ihnen Sätze über Familienbeziehungen präsentiert. Sie sollen mit Hilfe der gegebenen Informationen die fehlenden Beziehungen finden. Sehen Sie sich bitte zuerst folgende Beispielaufgaben an:*

Jemandem aus der Klasse die Antwort nennen lassen!

- 1. Karin hat die Tochter Tina. Peter ist der Bruder von Tina. Karin ist Peters MUTTER*
- 2. Harry ist der Sohn von Edith. Edith hat die Tochter Lisa. Lisa ist Harrys SCHWESTER*

*Dabei kommen ausschließlich echte Geschwister und echte Kinder eines Paares vor. Es gibt also keine Halb- oder Stiefgeschwister, keine Adoptivkinder und verheiratete Elternteile haben ausschließlich gemeinsame Kinder. Selbiges gilt auch für die Großeltern und Onkel bzw. Tanten. Wenn also Anna die Mutter von Simon ist und Peter der Ehemann von Anna, dann ist Simon automatisch der Sohn von Peter. Somit sind die folgenden Aufgaben rein durch logisches Kombinieren lösbar. Alle zu Lösung benötigten Informationen sind im Text der jeweiligen Aufgabe enthalten.*

*Jede Aufgabe hat nur eine Lösung. Wenn Sie bei einer Aufgabe nicht weiterkommen, gehen Sie einfach zur nächsten über. Bei der Bearbeitung der Aufgaben dürfen Sie sich keine Notizen machen.*

*Für die Testbearbeitung stehen Ihnen 15 Minuten zur Verfügung. Falls Sie vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig sind, warten Sie bitte auf ein Zeichen des Testleiters.*

*Gibt es von Ihrer Seite aus noch Fragen? LOS GEHT'S! BITTE UMBLÄTTERN!*

Für die Bearbeitung stehen den SchülernInnen 15 Minuten zur Verfügung. Nach 7 Minuten und 30 Sekunden ist den SchülernInnen mitzuteilen, dass die Hälfte der Bearbeitungszeit vorbei ist.

Nach den gesamten 15 Minuten ist mit dem Wort „STOPP“ abubrechen. Die SchülerInnen werden anschließend gebeten unter der zuletzt bearbeiteten Aufgabe einen Strich zu ziehen.

*Bitte macht wieder einen Strich unter der zuletzt bearbeiteten Aufgabe.*

Wenn alle einen Strich unter die zuletzt bearbeitete Aufgabe gemacht haben, wird wie folgt instruiert:

*Blättert jetzt bitte auf die nächste Seite zum Fragebogen.*

## **Fragebogen AVT**

*Auf den folgenden Seiten finden Sie eine Reihe von Aussagen, die ihr Fühlen und Handeln beim Lernen bzw. Arbeiten betreffen. Bitte nehmen Sie zu diesen Aussagen Stellung.*

*Kreuzen Sie bitte die Antwort an, die in der Regel, also nicht nur gerade heute, zutrifft. Hinter jedem Satz sind 4 Kästchen, die für die Antworten „stimmt“, „stimmt eher nicht“ und „stimmt nicht“ stehen. Entscheiden Sie sich bei jeder Aussage, in wie weit sie zutrifft oder nicht und kreuzen Sie das entsprechende Kästchen an.*

*Gibt es von Ihrer Seite noch Fragen?*

**LOS GEHT'S! BITTE UMBLÄTTERN!**

*Vielen Dank für eure Teilnahme.*

Bitte vorm Absammeln noch mal kurz kontrollieren, ob jeder/jede das Datenblatt anfangs ausgefüllt hat. Ebenfalls bitte darauf achten, dass niemand das Testheft mit nach Hause nimmt.

Anschließend noch bei den Lehrpersonen bedanken, dass diese für die Testung eine Unterrichtsstunde zur Verfügung stellten.