



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Tripelweiser Dimensionalitätsvergleich der Reasoning-Tests
Numerische Topologien, Family-Relation-Reasoning-Test und
Wiener Matrizen-Test 2“

verfasst von / submitted by

Zoé Josephine Babett Kresnik, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2015 / Vienna, 2015

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

A 298

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

Psychologie

Betreut von / Supervisor:

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

*Ich widme diese Diplomarbeit meiner Mutter Nicola Penning,
die mich nicht nur während des Verfassens dieser Arbeit
tatkraftig unterstützt hat, sondern mich stets ermutigt meinen
Weg zu gehen und mir immer helfend zur Seite steht.*

Danksagung

Viele Menschen haben mich beim Verfassen dieser Diplomarbeit unterstützt. Nicht alle kann ich hier namentlich nennen, aber ich möchte mich auf diesem Weg bei Euch für eure Hilfe und Zeit bedanken.

Danken möchte ich in erster Linie meinem Betreuer, Herr Univ.-Prof. Klaus Kubinger und Frau Mag. Serra Ünal, für ihre fachliche und organisatorische Unterstützung und ihre richtungsweisenden Ratschläge bei der Erstellung dieser Diplomarbeit.

Daneben gilt mein besonderer Dank Tobias Grüner und meinem Freund Sven Schwake für ihre Freundschaft, Geduld und große Hilfe bei dieser Arbeit.

Ein Dankeschön möchte ich auch an meine Kolleginnen Bernadette Schnait und Larissa Bartok aussprechen, die mir trotz eigenem Diplomarbeitsstress viele Fragen beantwortet und ihre kostbare Zeit gewidmet haben.

Auch Frau Dr. Karin Waldherr möchte ich für ihre Unterstützung danken.

Ich bedanke mich bei dem Lehrpersonal der SLAK Traiskirchen und den Krankenpflegeschulen Baumgartner Höhe und Ybbs, die mir die Durchführung meiner Testungen ermöglichten.

Nicht zuletzt möchte ich mich bei meinen Eltern Nicola Penning und Johann Kresnik bedanken, die mir mein Studium erst ermöglicht haben. Auch meinem Onkel Andreas Penning bin ich für seine Unterstützung sehr dankbar.

ABSTRACT (Deutsch)

Im Rahmen dieser Diplomarbeit wurden die beiden neu entwickelten Tests *N-TOP* (Kubinger & Heuberger, in Vorb.) und *FRRT* (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.) auf Konformität mit dem dichotomen logistischen Modell von Rasch und somit ihre Eindimensionalität überprüft. Weiterhin wurden diese Verfahren zusammen mit dem bereits etablierten *WMT 2* (Formann, Waldherr, & Piswanger, 2011) auf korrelative Zusammenhänge bzw. ihre diskriminante Validität untersucht. Die Daten wurden im Rahmen von Gruppentestungen an österreichischen Krankenpflegeschulen und einer Sicherheitsakademie erhoben. Die Stichprobe bestand aus 292 Schülern und Schülerinnen zwischen 18 bis 54 Jahren. Die Auswertung erfolgte mittels Likelihood-Ratio-Test, graphische Modellkontrolle sowie z -Test. Die Rasch-Modell-Analysen ergaben Modellgeltung der Items des *FRRT* sowie, nach Ausschluss eines Items, des *N-TOP*. Die Koeffizienten der Produkt-Moment-Korrelationen zwischen den Personenparametern der Verfahren lagen in einem Bereich von $r = 0,278$ bis $r = 0,379$. Aufgrund theoretischer Überlegungen wurde a-priori ein Grenzwert festgelegt, den Verfahren, die demselben Intelligenzbereich sensu Cattell (z.B. 1957; 1963) (fluid vs. kristallisiert) zugeordnet wurden, nicht unterschreiten sollten. Entgegen dieser Annahme überschritten die Korrelationskoeffizienten den Grenzwert nicht. Für die Tests, die unterschiedlichen Intelligenzkonzepten zugeordnet wurden, ließ sich die gegenteilige Hypothese durch Unterschreitung des Grenzwertes bestätigen. Limitierungen des verwendeten Testmaterials lassen einen endgültigen Schluss bezüglich der Gründe für die geringe Korrelation zwischen *WMT-2* und *N-TOP* nicht zu. Nähere Untersuchung der Diskrepanz zwischen erwarteter und gemessener Korrelation könnte Gegenstand zukünftiger Forschungsarbeiten sein.

Schlüsselwörter: schlussfolgerndes Denken, Reasoning, fluid, kristallisiert, Rasch-Modell, diskriminante

Validität, *N-TOP*, *FRRT*, *WMT 2*

ABSTRACT (English)

In the scope of this diploma thesis the newly developed psychological reasoning-tests *N-TOP* (Kubinger & Heuberger, in Vorb.) and *FRRT* (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.) are evaluated in regards to the Rasch model, assessing their dimensionality. Furthermore these tests and the already established *WMT 2* (Formann, Waldherr, & Piswanger, 2011) are examined for correlations and their discriminant validity. Data was gathered by testing groups at Austrian nursing schools and a police academy. The sample consisted of 292 students aged between 18 to 54 years. Collected data was then analysed employing the Likelihood-Ratio-Test, graphical modelling test and the z -Test. The results of the Rasch model analysis showed that this model holds a-posteriori for all the examined items of *FRRT*, as well as – after exclusion of one item – for those of *N-TOP*. The coefficients of the pearsons-correlations between personparameters of the tests were in the range of $r = .278$ to $r = .379$. Based on theoretical considerations, a threshold value was set a-priori. Against the assumption the correlation coefficients of those tests which are designed on the same intelligence concept sensu Cattell (e.g., 1957; 1967) (fluid vs. crystallized) didn't exceed this threshold. For tests that can be associated with different intelligence constructs, the null hypothesis was verifiable by the threshold value not being reached. Limitations of the testing-materials used do not allow for a final conclusion with regards to the reason of discrepancy between expected and measured correlations. This could be the subject of further inquiry.

Keywords: reasoning, fluid, crystallized, Rasch model, discriminant validity, *N-TOP*, *FRRT*, *WMT 2*

INHALTSVERZEICHNIS

I.	EINLEITUNG	1
II.	THEORETISCHER TEIL	3
1.	Der Intelligenzfaktor Reasoning – Schlussfolgerndes Denken	3
	1.1 Induktives Denken	6
	1.2 Deduktives Denken	7
2.	Intelligenztheoretische Grundlagen in Bezug auf den Faktor Reasoning	8
	2.1 Die Zwei-Faktoren-Theorie von Spearman	9
	2.2 Das Primärfaktorenmodell von Thurstone	10
	2.3 Die Theorie der fluiden und kristallisierten Intelligenz von Cattell	11
	2.4 Das Strukturmodell der Intelligenz von Guilford	14
	2.5 Das Berliner Intelligenzmodell von Jäger	16
	2.6 Die „Three Stratum“-Theorie von Carroll	18
3.	Testtheoretische Grundlagen	19
	3.1 Die Item-Response-Theorie	19
	3.2 Das dichotome logistische Testmodell von Rasch	20

III.	EMPIRISCHER TEIL	23
4.	Ziele der Untersuchung und Hypothesen	23
5.	Beschreibung der Testbatterie	25
	5.1 <i>Numerische Topologien (N-TOP)</i>	26
	5.2 <i>Family-Relation-Reasoning-Test (FRRT)</i>	27
	5.3 <i>Wiener Matrizen-Test 2 (WMT 2)</i>	28
	5.4 <i>Anstrengungsvermeidungstest für ältere Schüler (AVT-Sch)</i>	28
6.	Planung und Durchführung der Untersuchung	29
7.	Stichprobe	32
8.	Darstellung der Ergebnisse	35
	8.1 Analysen zur Überprüfung der Geltung des Rasch-Modells	35
	8.1.1 Ergebnisse der Rasch-Modell-Analysen beim <i>N-TOP</i>	37
	8.1.2 Ergebnisse der Rasch-Modell-Analysen beim <i>FRRT</i>	50
	8.2 Korrelative Zusammenhänge zwischen den Reasoning-Tests	55
IV.	DISKUSSION UND AUSBLICK	58
V.	ZUSAMMENFASSUNG	62
VI.	LITERATURVERZEICHNIS	64
VII.	ANHANG	69

ABBILDUNGSVERZEICHNIS

Abbildung 1: Itembeispiel aus <i>N-TOP</i> (Kubinger & Heuberger, in Vorb.).....	27
Abbildung 2: Itembeispiel aus <i>FRRT</i> (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.).....	27
Abbildung 3: Verteilung der Stichprobe in Bezug auf das Alter in Jahren ($n = 291$).....	33
Abbildung 4: Verteilung der Stichprobe in Bezug auf die Anstrengungsvermeidungstendenz ($n = 291$).....	34
Abbildung 5: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (<i>N-TOP</i> , alle Items).....	38
Abbildung 6: Grafische Modellkontrolle für das Teilungskriterium „Alter“ (<i>N-TOP</i> , alle Items).....	39
Abbildung 7: Grafische Modellkontrolle für das Teilungskriterium „Anstrengungsvermeidungstendenz“ (<i>N-TOP</i> , alle Items).....	40
Abbildung 8: Grafische Modellkontrolle mit Konfidenzellipsen für das Teilungskriterium „Bildungsniveau“ mit Beschriftung der auffälligen Items (<i>N-TOP</i> , alle Items).....	41
Abbildung 9: Grafische Modellkontrolle mit Konfidenzellipsen für das Teilungskriterium „Alter“ mit Beschriftung der auffälligen Items (<i>N-TOP</i> , alle Items).....	42
Abbildung 10: Grafische Modellkontrolle mit Konfidenzellipsen für das Teilungskriterium „Anstrengungsvermeidungstendenz“ mit Beschriftung der auffälligen Items (<i>N-TOP</i> , alle Items).....	43
Abbildung 11: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (<i>N-TOP</i> , ohne Item 7b).....	47
Abbildung 12: Grafische Modellkontrolle für das Teilungskriterium „Alter“ (<i>N-TOP</i> , ohne Item 7b).....	48

Abbildung 13: Grafische Modellkontrolle für das Teilungskriterium „Anstrengungs- vermeidungstendenz“ (<i>N-TOP</i> , ohne Item 7b).....	49
Abbildung 14: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (<i>FRRT</i> , alle Items).....	51
Abbildung 15: Grafische Modellkontrolle für das Teilungskriterium „Alter“ (<i>FRRT</i> , alle Items).....	52
Abbildung 16: Grafische Modellkontrolle für das Teilungskriterium „Anstrengungs- vermeidungstendenz“ (<i>FRRT</i> , alle Items).....	53

TABELLENVERZEICHNIS

Tabelle 1: Überblick über die untersuchten Reasoning-Tests.....	25
Tabelle 2: Testheftvarianten.....	26
Tabelle 3: Bearbeitungszeit und Itemanzahl pro Verfahren.....	32
Tabelle 4: Verteilung der Stichprobe in Bezug auf die höchste abgeschlossene Ausbildung ($n = 289$).....	34
Tabelle 5: Ergebnisse der Likelihood-Ratio-Tests ($N-TOP$, alle Items).....	37
Tabelle 6: Ergebnisse des z -Tests nach Andersen für das Teilungskriterium „Bildungsniveau“ ($N-TOP$, alle Items).....	44
Tabelle 7: Ergebnisse des z -Tests nach Andersen für das Teilungskriterium „Alter“ ($N-TOP$, alle Items).....	45
Tabelle 8: Ergebnisse der Likelihood-Ratio-Tests ($N-TOP$, ohne Item 7b).....	46
Tabelle 9: Ergebnisse der Likelihood-Ratio-Tests ($FRRT$, alle Items).....	50
Tabelle 10: Ergebnisse des z -Tests nach Andersen für das Teilungskriterium „Anstrengungs- vermeidungstendenz“ ($FRRT$, alle Items).....	54
Tabelle 11: Interkorrelationsmatrix der Reasoning-Tests.....	56
Tabelle 12: Matrix der Bestimmtheitsmaße der Reasoning-Tests.....	56

I. EINLEITUNG

Die Fähigkeit zum schlussfolgernden Denken ist ein wesentlicher Bestandteil der menschlichen Intelligenz und wird teilweise als eine fundamentale Denkfunktion betrachtet (siehe bspw. Jäger, 1973). Dementsprechend wird diese Fähigkeit in zahlreichen Intelligenztestbatterien, aber auch in spezifischen Tests erhoben.

Die vorliegende Diplomarbeit beschäftigt sich mit der Analyse und dem Vergleich psychologisch-diagnostischer Tests, die schlussfolgerndes Denken – in der Fachliteratur häufig mit dem englischen Äquivalent *Reasoning* bezeichnet – erfassen. Häufig finden derartige Verfahren im Rahmen von Berufs- sowie Bildungsberatungen, Auswahlverfahren oder in der Diagnostik von Lern- und Leistungsproblemen Anwendung. Die vielfältigen Einsatzbereiche und die Relevanz von Reasoning-Tests für die heutige Praxis spiegeln sich in der steigenden Zahl von neu entwickelten Verfahren zur Erhebung des schlussfolgernden Denkens wieder. Auch die Tatsache, dass Reasoning-Tests, die auf dem Konzept der fluiden Intelligenz von Cattell (z.B. 1957; 1963) basieren, sprachfrei und weitestgehend unabhängig von kulturspezifischem Wissen konzipiert werden können, spricht angesichts einer wachsenden Globalisierung für sie und ihre Notwendigkeit.

Bei dem für diese Diplomarbeit verwendeten Testmaterial handelt es sich neben einem im deutschen Sprachraum bereits etablierten Test, um zwei neu entwickelte Verfahren. Es wurde untersucht, inwieweit diese tatsächlich die zu erfassende Fähigkeit Reasoning erheben und ob sich die zugrunde liegenden Konstrukte der Tests ähneln bzw. unterscheiden. Grundlegend basieren die drei Verfahren auf dem Konzept der fluiden und der kristallisierten Intelligenz sensu Cattell (1957; 1963). Im Speziellen wurde mittels einer Überprüfung auf Rasch-Modell-Konformität untersucht, ob die beiden neuen Tests die Fähigkeit Reasoning eindimensional messen. Damit soll auch ein Beitrag zur Überprüfung der Validität geleistet werden.

Zu Beginn dieser Diplomarbeit widmet sich der theoretische Teil den intelligenztheoretischen Grundlagen. Um die Bedeutsamkeit der Fähigkeit Reasoning aufzuzeigen, wird diese in Bezug zu grundlegenden Definitionen von Intelligenz gesetzt. Anschließend führt ein kurzer Diskurs über schlussfolgerndes Denken, über die den angewendeten Verfahren zugrunde liegende Definition von Reasoning und über die Teilaspekte dieser Fähigkeit in die Thematik ein. Danach wird der Faktor Reasoning innerhalb verschiedener Intelligenzmodelle positioniert. Das Kapitel testtheoretische Grundlagen bietet einen Überblick über die Item-Response-Theorie und das dichotome logistische Modell von Rasch, nach dem die untersuchten Verfahren konstruiert wurden. Im empirischen Teil der Arbeit werden die Ziele, die Testbatterie sowie die Untersuchung und deren Ergebnisse vorgestellt.

II. THEORETISCHER TEIL

1. Der Intelligenzfaktor Reasoning – Schlussfolgerndes Denken

In diesem Kapitel wird der Intelligenzfaktor Reasoning sowie die Formen dieser Fähigkeit näher betrachtet. Der Faktor Reasoning, in diesem Zusammenhang meist mit schlussfolgerndem Denken übersetzt¹, ist ein Teilaspekt des theoretischen Konstrukts der Intelligenz. Die Bedeutsamkeit dieser Komponente innerhalb der Intelligenzforschung beschreibt Lohmann (2005) folgendermaßen: „Even though there is more to intelligence than reasoning, reasoning is a crucial aspect of any understanding of human intelligence“ (S. 227). Jene verschiedenen Sichtweisen auf Intelligenz seien im Folgenden anhand der wichtigsten Intelligenzdefinitionen beschrieben.

Ein Grund für die Vielzahl an Definitionen mag sein, dass Intelligenz kein direkt beobachtbares Merkmal ist und nur indirekt über das Verhalten in Leistungssituationen erhoben werden kann. Dennoch spielt dieses Merkmal in zahlreichen Forschungsbereichen eine gewichtige Rolle und hat seit Beginn seiner Erforschung vor mehr als 100 Jahren nicht im Geringsten an Interesse verloren (Süß, 2003).

Die Entwickler des – nach heutigem Verständnis – ersten Intelligenztests, Alfred Binet und Théodore Simon, teilten die Auffassung, dass Intelligenz „die Art der Bewältigung einer aktuellen Situation (. . .) gut urteilen, gut verstehen und gut denken“ sei (Binet & Simon, 1905, zitiert nach Stemmler, Hagemann, Amelang, & Bartussek, 2011, S. 138).

¹ In dieser Arbeit werden die Begriffe Reasoning und schlussfolgerndes bzw. logisch-schlussfolgerndes Denken synonym verwendet.

William Stern (1912), Begründer der Differentiellen Psychologie, beschreibt Intelligenz als „die allgemeine Fähigkeit des Individuums, sein Denken bewußt auf neue Forderungen einzustellen; sie ist [die] allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens“ (S. 3).

David Wechsler (1956) legt Intelligenz als „die zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen“ (S. 13) fest.

Peter R. Hofstätter (1977) definiert Intelligenz als die Fähigkeit zur Auffindung von Ordnungen in der Welt.

Trotz dieser teilweise „leeren Tautologien“, wie Hofstätter (1977, S. 190) sie nennt, umschreiben all diese Auffassungen von Intelligenz die Fähigkeiten, Probleme zu lösen und sich an neue Bedingungen anzupassen, auch wenn „[im allgemeinen] unausgesprochen bleibt (. . .), was zur Lösung neuer Probleme bzw. zur geistigen Anpassung an neue Aufgaben erforderlich sei“ (Hofstätter, 1977, S. 191). Nach Stemmler et al. (2011) haben diese „verbalen Definitionen“ (S. 139) für das Verständnis des Konstrukts Intelligenz keinen großen Beitrag geleistet. Demgegenüber stehen die sogenannten „operationalen Definitionen“ (S. 139). Diese resultieren aus der „Spezifizierung der zum Zwecke seiner [des Merkmals] Messung ausgeführten empirischen-experimentellen Handlungen“ (S. 139). Somit entstehen Definitionen häufig über die Operationalisierung eines Merkmals (Hofstätter, 1977; Stemmler et al., 2011). In diesem Sinne mag auch die saloppe Definition Edwin Borings von Intelligenz als das, was Intelligenztests messen, verstanden werden (Boring, 1923, zitiert nach Süß, 2003, S. 217).

Die in Kapitel 2 vorgestellten Intelligenztheorien werden die in den genannten Definitionen beschriebenen Fähigkeiten oder äußerst ähnlich definierte Fähigkeiten als zentrale Komponenten der Intelligenz positionieren. Wie im Weiteren gezeigt wird, decken diese sich zu großen Teilen mit dem Konstrukt Reasoning.

Als Ursache für schlussfolgerndes Denken können nach Schüling (2002) Wahrnehmungsdefizite ausgemacht werden. Diese Defizite sind dergestalt, dass zukünftiges, vergangenes, weit entferntes, sehr kleines oder lückenhaftes sich, angesichts seiner Natur, der Wahrnehmung entzieht. Zweck des Schlussfolgerns ist dementsprechend das Begehren, trotz dieser Defizite zu einer Erkenntnis zu gelangen – einem Informationsgewinn. Erste wissenschaftliche Arbeiten zum schlussfolgernden Denken lassen sich bei den Griechen und später den Formallogikern finden (Schüling, 2002). Die sogenannte *formale Logik* als Teilgebiet der Philosophie beschäftigt sich dabei mit Schlussfolgerungsprozessen und umfasst die Lehre vom Begriff, vom Urteil und vom Schluss (Beckmann & Guthke, 1999; Kutschera & Breitkopf, 2000).

In der Logik (. . .) versteht man unter Schließen die Anwendung eines geregelten Verfahrens (sog. Kalküle), um von bestimmten Aussagen bzw. Urteilen (Prämissen eines Schlusses) zu einer weiteren Aussage bzw. Urteil (Konklusion des Schlusses) zu gelangen.

(Beckmann & Guthke, 1999, S. 1)

Die formale Logik widmet sich jedoch allein den Elementen und Vorgängen des korrekten, d.h. streng logischen Regeln folgenden Schlussfolgerns und lässt menschliche Komponenten dabei außen vor. Erst die Wissenschaft der Psychologie bezog bei ihren Untersuchungen –

neben den Erkenntnissen der formalen Logik – auch die menschliche Informationsverarbeitung mit ein (Beckmann & Guthke, 1999).

So leitet Kubinger (2009) aus dem zur Verfügung stehenden Testinventar die Definition von Reasoning als die Fähigkeit „Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge erkennen und zweckentsprechend verwerten zu können“ (S. 206) ab. Schlussfolgerndes Denken lässt sich dabei in zwei grundlegende Formen unterteilen – induktives und deduktives Denken.

1.1 Induktives Denken

Induktives Denken bezeichnet das Schlussfolgern vom Besonderen auf das Allgemeine. Dabei werden aus gegebenen Prämissen neue Informationen generiert (Beckmann & Guthke, 1999). Der induktive Schluss führt somit von der Kenntnis einzelner Fakten zu Verallgemeinerungen. Da eine solche Verallgemeinerung jedoch nicht zwangsläufig für alle möglichen Bedingungen gelten mag, ist die induktive anfälliger gegenüber Irrtümern als die deduktive Schlussfolgerung (Beckmann & Guthke, 1999). In diesem Sinne bezeichnet Kant den Induktionsschluss auch als Wahrscheinlichkeitsschluss (Beckmann & Guthke, 1999). Waldmann und Weinert (1990) definieren induktives Denken als

Prozesse, die ein Individuum dazu befähigen, aus spezifischen Erfahrungen allgemeine Regeln, Muster, Konzepte oder Gesetzmäßigkeiten abzuleiten und auf neue Ereignisse anzuwenden. Induktives Denken bezeichnet damit eine Kernleistung menschlicher Intelligenz, nämlich die Fähigkeit, aus Erfahrungen zu lernen.

(S. 25-26)

Beispiele für psychologisch-diagnostische Verfahren, die induktives Inferieren erheben, sind Matrizentests, Analogie-, Reihenfortsetzungs- oder Klassifikationsaufgaben (Waldmann & Weinert, 1990).

1.2 Deduktives Denken

Eine Deduktion beschreibt eine Schlussfolgerung, die von einem allgemeinen Sachverhalt auf einen Einzelfall übergeht. Im Gegensatz zur Induktion erfolgt dabei zum einen keine Erweiterung der Informationsmenge und zum anderen kann bei Korrektheit der Regeln und Wahrheit der Prämissen auch von der Wahrheit der Schlussfolgerung ausgegangen werden (Beckmann & Guthke, 1999). In diesem Sinne bezeichnet Brentano (1956) deduktive Schlüsse auch als „Gewißheitsschlüsse“ (S. 202). Waldmann und Weinert (1990) definieren Deduktion als „einen systematischen Denkprozeß, der von einer bestimmten Menge von Prämissen unter Anwendung logischer Regeln zu einer Menge von Schlußfolgerungen führt“ (S. 47).

Klassische Aufgabentypen, die deduktives Denken erfassen, stellen Syllogismen dar. Derartige Reasoning-Aufgaben bestehen aus Prämissen, von denen logische Konklusionen abgeleitet werden sollen.

Kritisch anzumerken ist, dass das Unterscheidungsmerkmal der Schlussrichtung zwischen Induktion und Deduktion als unzureichend betrachtet werden muss. Findet beispielsweise im Zuge einer vollständigen Induktion – diese enthält bereits alle relevanten Einzelfälle in Form von Prämissen für die Schlussfolgerung – keine Generierung neuer Informationen statt, so lässt sich diese Folgerung auch als Deduktion bezeichnen (Beckmann & Guthke, 1999). Daher sollte nach Beckmann und Guthke (1999) die Schlussrichtung lediglich als Oberflächenmerkmal betrachtet und vielmehr der Charakter der Konklusion –

Informationen der Prämissen werden erweitert/nicht erweitert – als Differenzierungsmerkmal genutzt werden. Überdies bemerken Beckmann und Guthke (1999), dass

jede Deduktion das Resultat einer vorausgegangenen Induktion [ist]; das Resultat induktiven Denkens manifestiert sich letztendlich über einen deduktiven Denkakt. So erfolgt die Prüfung der Fähigkeit zum induktiven Denken im Rahmen der Psychodiagnostik nicht selten explizit über Deduktion.

(S. 6)

In diesem Sinne erscheint eine Differenzierung von Aufgabentypen, die schlussfolgerndes Denken erfassen, nicht immer sinnvoll.

2. Intelligenztheoretische Grundlagen in Bezug auf den Faktor Reasoning

Psychologische Verfahren zur Messung der Intelligenz – aber auch im Allgemeinen - müssen immer in Hinblick auf ihre zugrunde liegenden theoretischen Konzepte und Annahmen hin interpretiert werden. Die theoretischen Überlegungen spielen bei der Erhebung der Intelligenz insofern eine wichtige Rolle, als Tests meist nur Teilaspekte bzw. einzelne Faktoren der Intelligenz erfassen und diese – je nach Intelligenzmodell – unterschiedliche Definitionen besitzen können (Stemmler et al., 2011).

In den folgenden Kapiteln (2.1 bis 2.5) werden daher verschiedene Intelligenztheorien vorgestellt, welche im Rahmen dieser Arbeit für das Verständnis und die Definition des Faktors Reasoning von Bedeutung sind.

2.1 Die Zwei-Faktoren-Theorie von Spearman

1904 entwarf der britische Psychologe Charles Spearman als Erster ein explizites Modell der Intelligenz (Funke & Vaterrodt-Plünnecke, 1998). Aufgrund der Ergebnisse von Korrelationsberechnungen und späteren Faktorenanalysen postulierte Spearman (1904; 1914; 1927) in seinem Modell – der sogenannten *Zwei-Faktoren-Theorie* – dass jede Testleistung im Intelligenzbereich auf einen gemeinsamen Intelligenzfaktor g und einen zweiten spezifischen Faktor s zurückzuführen sei.

Der erste Faktor – von Spearman als g -Faktor (für engl. *general factor*) bezeichnet – meint die allgemeine Intelligenz, genauer den Varianzanteil, den alle Fähigkeiten im Intelligenzbereich gemeinsam haben. Diese Annahme begründete Spearman (1904; 1914) damit, dass kognitive Leistungen von Testpersonen in unterschiedlichen Tests gewöhnlich positiv miteinander korrelierten und demzufolge ein gemeinsamer Faktor zugrunde liege. Der zweite Faktor – von Spearman s -Faktor (für engl. *specific*) genannt – beschreibt dabei den Varianzanteil der Leistung, die nicht durch g erklärbar und spezifisch für die jeweilige Aufgabenart ist. Dementsprechend gibt es nach Spearman theoretisch so viele s -Faktoren wie es Verfahren zur Messung kognitiver Fähigkeiten gibt. Die Ausprägungen der Faktoren g und s können dabei stets unterschiedlich stark ausfallen (Funke & Vaterrodt-Plünnecke, 1998).

Nach Spearman (1946, zitiert nach Beckmann & Guthke, 1999, S. 6) stellen in Tests, die hoch auf dem g -Faktor laden, das Erkennen von Beziehungen und das Finden von Regeln wichtige Kernkompetenzen dar. Ein Beispiel für derartige Verfahren wären figurale Matrizentests.

Kritisch anzumerken ist, dass sich Spearmans Zwei-Faktoren-Theorie für die Erklärung der Varianzen kognitiver Fähigkeiten in der Empirie nicht durchsetzen konnte. Es zeigte sich, dass einige Aufgabentypen stärker untereinander korrelierten als ihre Ladungen auf dem g -Faktor vermuten ließen (Holling, Preckel, & Vock, 2004).

2.2 Das Primärfaktorenmodell von Thurstone

Aufbauend auf den Arbeiten von Spearman beschäftigte sich auch der Psychologe Louis Thurstone mit den Faktoren der Intelligenz (Jäger, 1973). Im Gegensatz zu Spearman vertrat Thurstone in seinem *Primärfaktorenmodell* (auch *Modell mehrerer gemeinsamer Faktoren* genannt) die Annahme, dass es anstelle eines einzigen, übergeordneten g-Faktors mehrere, voneinander unabhängige, generelle Faktoren gebe, 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). Demzufolge seien bei Denkleistungen non-hierarchische Gruppenfaktoren in unterschiedlichen Gewichtungsverhältnissen beteiligt. Auf Basis von Ergebnissen multipler Faktorenanalysen extrapolierte Thurstone folgende sieben Primärfaktoren (engl. *primary mental abilities*), die sich wiederholt belegen ließen (Guilford & Hoepfner, 1976; Stemmler et al., 2011; Süß, 2003):

- **V** (*verbal comprehension*): Wortschatz, Sprachverständnis, verbal-logisches Denken
- **W** (*word fluency*): Wortflüssigkeit
- **N** (*number*): Fähigkeit zum Umgang mit Zahlen, Rechenfähigkeit
- **S** (*space*): räumliches Vorstellungsvermögen
- **M** (*memory*): Merkfähigkeit, Kurzzeitgedächtnis
- **P** (*perceptual speed*): Wahrnehmungsgeschwindigkeit
- **R** (*reasoning*): schlussfolgerndes Denken, Fähigkeit zum induktiven und deduktiven

Denken

Thurstone unterteilte den Faktor Reasoning zunächst in die Faktoren **I** (*induction*), **D** (*deduction*) und **R**, wobei **R** namentlich nicht näher ausgeführt wird. *Induction* definierte er als die Fähigkeit zum Erkennen einer dem Aufgabenmaterial zugrunde liegenden

Gesetzmäßigkeit. *Deduction* entspricht dabei der Anwendung dieser Gesetzmäßigkeit bzw. Regel und die Anwendung des Faktors *R* findet dort statt, „wo irgendeine Art Restriktion bei der Lösung einer Aufgabe gefordert [wird]“ (Jäger, 1973, S. 80). Später fasste Thurstone diese drei Faktoren als *Reasoning* zusammen, welcher unter den sieben Primärfaktoren oft als *der* Kernfaktor der Intelligenz angesehen wird (Beckmann & Guthke, 1999; Jäger, 1973).

Im Vergleich mit Spearmans Theorie bietet das Modell von Thurstone den Vorteil einer differenzierteren Beschreibung der Leistungsfähigkeit einzelner Personen, da sich Intelligenz hier aus mehreren, unabhängigen Faktoren zusammensetzt (Funke & Vaterrodt-Plünnecke, 1998). Zudem betrachtete Thurstone sein Modell als ein offenes System, welches jederzeit ergänzt und modifiziert werden könne (Jäger, 1973). Gemein ist beiden Modellen, dass sowohl Spearman als auch Thurstone schlussfolgerndes Denken als eine der zentralen Hauptkomponenten der Intelligenz betrachteten.

2.3 Die Theorie der fluiden und kristallisierten Intelligenz von Cattell

1957 veröffentlichte Raymond B. Cattell seine bis heute wegweisende Theorie der fluiden und kristallisierten Intelligenz. Sein hierarchisches Modell der Intelligenz vereint dabei u. a. Aspekte der Zwei-Faktoren-Theorie nach Spearman sowie des Primärfaktorenmodells von Thurstone (Cattell, 1957, 1963).

Cattell ging davon aus, dass sich der *g*-Faktor nach Spearman aus zwei voneinander abhängigen Faktoren – der „fluiden allgemeinen Intelligenz“ (*gf*-Faktor) und der „kristallisierten allgemeinen Intelligenz“ (*gc*-Faktor) – zusammensetzt. Diese beiden generellen Faktoren ordnete er auf einer (zweiten) Ebene über sogenannte, miteinander korrelierte Primärfaktoren an, welche auf Testleistungen in verschiedenen Aufgabenarten

basieren (Cattell, 1963; Horn & Cattell, 1966; siehe dazu Funke & Vaterrodt-Plünnecke, 1998; Stemmler et al., 2011).

Darüber hinaus vertraten Cattell (1957; 1963) sowie sein Arbeitskollege Horn (1968), mit dem er seine Theorie weiterentwickelte, die Annahme, dass die fluide Intelligenz ihren Höhepunkt vergleichsweise rasch erreicht und bereits im frühen Erwachsenenalter ein Abbau dieser Fähigkeiten einsetzt. Dem gegenüber entwickelt sich die kristallisierte Intelligenz langsamer und über die gesamte Lebensspanne hinweg und baut nur geringfügig oder gar nicht ab. Cattells und Horns Annahmen über die Entwicklungsverläufe der Faktoren g_f und g_c konnten in einer neueren Längsschnittstudie untermauert werden (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002).

Die fluide Intelligenz (engl. *fluid intelligence*, g_f) basiert größtenteils auf genetisch determinierten, kognitiven Prozessen (Cattell, 1957; siehe dazu Süß, 2003). Sie zeichnet sich durch Fähigkeiten aus, die unabhängig von bisherigen Lernerfahrungen, zur Bewältigung neuer Situationen und Probleme benötigt werden. Maßgeblich wird die fluide Intelligenz dabei von der Entwicklung des zentralen Nervensystems und der Sinnesorgane bestimmt; sie bildet die Grundlage für abstraktes und logisch-schlussfolgerndes Denken (Cattell, 1963; 1987; Horn & Cattell, 1966; siehe dazu Stemmler et al., 2011; Süß, 2003). Gustafsson (1988, zitiert nach Lohmann, 2005, S. 226-227) geht dabei so weit, den g_f -Faktor mit induktivem Schlussfolgern und der allgemeinen Intelligenz g gleich zu setzen. Nach Horn (1968) eignen sich zur Messung der fluiden Intelligenz vor allem solche Aufgaben, die induktives Schlussfolgern, figurale Beziehungen sowie intellektuelle Geschwindigkeit erfassen. Verfahren, die fluide Intelligenz erheben, sind aufgrund ihrer Unabhängigkeit von bisherigen

Lernerfahrungen und kulturspezifischem Wissen in Bezug auf die Aufgabenstellung häufig *culture-fair*².

In so far as the theory can be expressed in a few familiar words, it is that *fluid ability* („gf“) is a *general relation-perceiving capacity* which operates in all fields. It is measured, most saturatedly, by relation and correlate education of a complexity appropriate to the subjects, in any field where acquired skills are not differentiating; for it is independent of sensory content.

(Cattell, 1957, S. 743)

Die kristallisierte Intelligenz (engl. *crystallized intelligence*, *g_c*) basiert auf Fertigkeiten, bei denen Lernerfahrungen bzw. erworbenes Wissen – sogenanntes kristallisiertes Wissen – zur Lösung von Problemen angewendet werden (Cattell, 1963; 1987; Horn & Cattell, 1966). Somit spielen Lern- und Sozialisationsprozesse während der Entwicklung der kristallisierten Intelligenz eine entscheidende Rolle. Nach Cattell (1963) zählen beispielsweise Thurstones Primärfaktoren *verbal comprehension* (Wortschatz, Sprachverständnis, verbal-logisches Denken) und *number* (Rechenfähigkeit) sowie Kenntnisse über Geographie oder Geschichte zu diesem Bereich.

Crystallized ability („gc“), on the other hand, must be measured on a basis of careful cultural sampling, for it is not independent of area and is simply a *sum of particular relation-perceiving skills acquired in specific fields*.

(Cattell, 1957, S. 743)

² Verfahren, welche als „culture-fair“ (zu Deutsch: „kulturfrei“, „kulturunabhängig“) bezeichnet werden, verwenden Instruktionen und Testmaterialien, die sprachfrei und (größtenteils) unabhängig von kulturspezifischem Wissen bearbeitet werden können (Kubinger, 2009, S. 206–207).

Mit der Beziehung zwischen fluider und kristallisierter Intelligenz beschäftigte sich Cattell (1987) ausführlicher im Rahmen seiner sogenannten Investment-Theorie. Basierend auf Ergebnissen von Korrelationsuntersuchungen beider Faktoren, stellte Cattell (1987, S. 116), unabhängig vom Alter, einen positiven Zusammenhang von $r = 0,4 - 0,5$ fest. Diesem Resultat entnahm Cattell, dass die Entwicklung der kristallisierten Intelligenz maßgeblich von der fluiden Intelligenz abhängig sei. Einhergehend mit der Entwicklung des Gehirns entwickle sich die Fähigkeit zum Erkennen von Beziehungen – der *gf*-Faktor – als erster Baustein der Intelligenz. Diese Fähigkeit beeinflusse den Erwerb weiterer, komplexerer Fähigkeiten wie beispielsweise die Lerngeschwindigkeit, welche wiederum Auswirkungen auf die Entwicklung des *gc*-Faktors habe. Demnach ist Wissen bzw. kristallisierte Intelligenz, investierte fluide Intelligenz (Cattell, 1987). Es sollte jedoch nicht außer Acht gelassen werden, dass Faktoren wie Motivation, Interesse und Lernumgebung ebenfalls Einfluss auf den Erwerb des *gc*-Faktors haben (Cattell, 1987).

Ebenso wie die zuvor behandelten Intelligenztheorien wird das Erkennen von Beziehungen in Cattells Modell als eine zentrale Fähigkeit der Intelligenz angesehen. Reasoning wird dabei vorrangig dem Bereich der fluiden Intelligenz zugeordnet. Spätere Untersuchungen zeigen jedoch, dass der Faktor Reasoning ebenfalls im *gc*-Faktor hoch lädt und daher auch mit wissensgebundenen Denkaufgaben erhoben werden kann (Beckmann & Guthke, 1999).

2.4 Das Strukturmodell der Intelligenz von Guilford

Einen gänzlich anderen Ansatz zur Struktur der Intelligenz verfolgte der Persönlichkeits- und Intelligenzforscher Joy Paul Guilford. Auf der Basis von Kontrolluntersuchungen diverser Intelligenzmodelle die ständig zu neuen Faktoren führten, vertrat Guilford die Annahme, dass

sich Intelligenz nicht wie bei Spearman aus einem übergeordneten Generalfaktor oder wie bei Thurstone aus mehreren Primärfaktoren zusammensetze, sondern aus vielen unabhängigen Einzelfaktoren, die untereinander kaum oder gar nicht miteinander korrelieren, bestehe (Funke & Vaterrodt-Plünnecke, 1998). Darauf aufbauend publizierte Guilford 1959 sein *Strukturmodell der Intelligenz* (engl. *Structure of Intellect*), auch bekannt als Würfelmodell der Intelligenz (Funke & Vaterrodt-Plünnecke, 1998; Jäger, 1973). Graphisch in Form eines dreidimensionalen Quaders dargestellt, besteht das Modell aus drei Dimensionen (*Operationen, Inhalte, Produkte*) (siehe hierzu Guilford, 1971, S. 63). Intelligentes Verhalten setzt sich demnach aus intellektuellen *Operationen*, deren *Inhalten* und den jeweiligen *Produkten* zusammen (Funke & Vaterrodt-Plünnecke, 1998; Guilford & Hoepfner, 1976; Hofstätter, 1977). Die Dimensionen sind wie folgt unterteilt (Guilford & Hoepfner, 1976):

1. *Operationen*: Kognition, Gedächtnis, divergente Produktion, konvergente Produktion, Evaluation
2. *Inhalte*: figural, symbolisch, semantisch, Verhalten
3. *Produkte*: Einheiten, Klassen, Beziehungen, Systeme, Transformationen, Implikationen

Jede Kombination der Unterkategorien der Dimensionen entspricht einer kognitiven Fähigkeit. Daraus resultieren 120 ($= 5 \times 4 \times 6$) mögliche theoretische, unabhängige Intelligenzfaktoren (Guilford, 1971), von denen ein großer Teil bereits empirisch bestätigt werden konnte (Funke & Vaterrodt-Plünnecke, 1998; Jäger, 1973).

Der Faktor Reasoning wird im Strukturmodell der Intelligenz in Induktion und Deduktion unterteilt. Induktion wird dabei der Operation *Kognition* und Deduktion der Operation *konvergente Produktion* zugeordnet (Guilford, 1956, zitiert nach Putz-Osterloh, 1981, S. 37).

2.5 Das Berliner Intelligenzstrukturmodell von Jäger

Einen integrativen und bisherige Intelligenzmodelle-umfassenderen Ansatz zur Struktur der Intelligenz, versuchte Adolf Otto Jäger mit seinem *Berliner Intelligenzstrukturmodell* zu verwirklichen.³ Ausgangspunkt bot die Problematik, dass bei der Untersuchung möglicher Intelligenzfaktoren nur diejenigen aufgespürt und erfasst werden konnten, welche dem Testmaterial entsprechend erhoben wurden (Jäger, 1967; 1982, zitiert nach Holling et al., 2004, S. 23; siehe dazu Süß, 2003). Aus diesem Grund katalogisierte Jäger zunächst „alle“ in der Literatur beschriebenen Aufgabenarten zur Intelligenz- sowie Kreativitätsmessung (was in etwa 2000 Itemtypen entsprach). Dieser Itempool von Leistungsvariablen wurde in weiterer Folge auf 98 Aufgabenblöcke mit insgesamt 191 Items reduziert. Bei der Auswahl wurde im Besonderen darauf geachtet der Vielfalt des Aufgabenmaterials sowie den konkurrierenden Strukturmodellen gerecht zu werden (Stemmler et al., 2011; Süß, 2003).

Cluster- und exploratorische Faktorenanalysen führten schließlich zu vier Operations- und drei Inhaltsklassen, aus deren Kombination – ähnlich den Facetten des Strukturintelligenzmodell nach Guilford – sich insgesamt zwölf spezifische Fähigkeitsbereiche ergaben (Stemmler et al., 2011). Die *Operationen* und *Inhalte* sind dabei einem übergeordneten g-Faktor sensu Spearman untergeordnet und in unterschiedlicher Gewichtung an allen intellektuellen Leistungen beteiligt.

³ Konkret vereint das Berliner-Intelligenzstrukturmodell von Jäger Elemente aus den Modellen von Spearman, Thurstone und Guilford (siehe dazu z.B. Jäger, 1973).

Sie sind wie folgt unterteilt (Stemmler et al., 2011; Süß, 2003):

1. Ebene: g-Faktor – allgemeine Intelligenz
2. Ebene: *Operationen:* Bearbeitungsgeschwindigkeit, Merkfähigkeit, Einfallsreichtum, Verarbeitungskapazität
Inhalte: figural-bildhaft, verbal, numerisch
3. Ebene: Spezifische Fähigkeitsbereiche; z. B. Einfallsreichtum im verbalen Bereich

Die Gültigkeit des hierarchischen und bimodalen⁴ Strukturmodells von Jäger konnte mehrfach in kulturübergreifenden sowie unterschiedliches Aufgabenmaterial verwendenden Untersuchungen gezeigt werden (Süß, 2003).

Der Faktor Reasoning findet sich in diesem Modell unter der operativen Fähigkeit *Verarbeitungsgeschwindigkeit* wieder (Jäger, 1973). Kernelemente der Verarbeitungsgeschwindigkeit nach Jäger (1973) sind „Problemlösen und Produktives Denken auf höherem Niveau“ (S. 129). Darunter zu verstehen sind die Verarbeitung und Beurteilung komplexer Informationen, formallogisches Denken sowie das Erkennen von Beziehungen (Holling et al., 2004). Schlussfolgerndes Denken im Sinne Jägers als ein primärer Fähigkeitsbereich der Intelligenz lässt sich dabei auf allen drei Inhaltsebenen erfassen – mit anschauungs-, sprach- sowie zahlengebundenem Aufgabenmaterial.

⁴ Die Zuschreibung „bimodal“ begründet sich darauf, dass die spezifischen Leistungsfaktoren auf zwei Faktoren, einem operativen und einem inhaltsgebundenen, laden; man spricht daher ebenfalls von Bimodalität der Operations- und Inhaltsfaktoren (vgl. bspw. Jäger, 1973).

2.6 Die „Three Stratum“-Theorie von Carroll

Mit seiner „Three Stratum“-Theorie versuchte John B. Carroll (1993) – wie bereits Jäger zuvor – der anhaltenden Unstimmigkeit über Anzahl und Konstellation verschiedener Intelligenzfaktoren ein Ende zu bereiten und ein integratives Intelligenzstrukturmodell zu entwickeln (siehe auch Stemmler et al., 2011). Anhand einer umfangreichen Metaanalyse von 460 Studien aus der Intelligenzforschung aus einen Zeitraum von 1930 bis 1985, leitete Carroll (1993) ein hierarchisches Intelligenzmodell, bestehend aus drei Schichten, ab.

Auf der ersten und untersten Schicht, die Carroll (1993) als „Stratum I“ bezeichnete, befinden sich 68 Einzelfaktoren der Intelligenz, die eine ähnliche Spezifität wie die Faktoren aus Guilfords Strukturmodell der Intelligenz besitzen. Aufgrund ihrer starken Korrelation extrahierte Carroll eine zweite, mittlere Hierarchieebene, das „Stratum II“, mit insgesamt acht Faktoren, welche inhaltlich ungefähr den Primärfaktoren nach Thurstone entsprechen (Stemmler et al., 2011). Die oberste Schicht, das sogenannte Stratum III, bildet die allgemeine Intelligenz, die durch komplexe kognitive Prozesse gekennzeichnet ist und allen intellektuellen Aktivitäten, ähnlich dem g-Faktor nach Spearman, zugrunde liegt. Die Stratum-II-Faktoren unterscheiden sich dabei in ihren Ladungen auf die allgemeine Intelligenz. So weisen beispielsweise die Faktoren fluide und kristallisierte Intelligenz die höchsten Ladungen auf (Stemmler et al., 2011).

Die Schichten sind wie folgt unterteilt (Stemmler et al., 2011):

Stratum III: allgemeine Intelligenz

Stratum II: fluide Intelligenz, kristallisierte Intelligenz, Gedächtnis & Lernen, visuelle Wahrnehmung, auditive Wahrnehmung, Retrieval, kognitive Schnelligkeit, Verarbeitungsgeschwindigkeit

Stratum I: spezifische Einzelfaktoren, z.B. Leseverständnis, Konzeptflüssigkeit etc.

Die Fähigkeit zum schlussfolgernden Denken wird in diesem Modell dem Faktor der fluiden Intelligenz zugeordnet.

3. Testtheoretische Grundlagen

Im Folgenden werden die testtheoretischen Grundlagen, auf denen diese Arbeit basiert, vorgestellt. Es wird ein Überblick über die Item-Response-Theorie sowie im Speziellen über das dichotome logistische Modell von Georg Rasch gegeben. Das Modell wird im Rahmen dieser Untersuchung zur Überprüfung auf Eindimensionalität der Verfahren *N-TOP* und *FRRT* herangezogen (siehe Kapitel 8.1).

3.1 Die Item-Response-Theorie

Die Testtheorie bildet die Grundlage für die Konstruktion psychologischer Verfahren – sogenannter Messinstrumente. Gegenstand der Testtheorie im Allgemeinen ist es, zu erfüllende Anforderungen an diese Messinstrumente zu formulieren und Methoden zu deren Überprüfung zu entwickeln – entsprechend einer Qualitätssicherung. Dabei lassen sich beispielsweise laut Kubinger (2003b) historisch wie auch wissenschaftstheoretisch zwei unterschiedliche Ansätze festhalten – die *Klassische Testtheorie* und die *Probabilistische Testtheorie*. Im Gegensatz zur probabilistischen Testtheorie, die in der internationalen Literatur vorzugsweise auch als *Item-Response-Theorie* oder *Latent-Trait-Theorie* bekannt ist, basiert die klassische Testtheorie primär auf korrelationsstatistischen Betrachtungen und stellt ein nicht prüfbares Modell dar. Die Item-Response-Theorie hingegen beruht auf einer probabilistischen Funktion und beinhaltet auch überprüfbare Modelle (Kubinger, 2003b). Konkret geht die Item-Response-Theorie davon aus, dass manifeste, d.h. beobachtbare Reaktionen einer Person, auf latente Eigenschaften zurückgeführt werden können. Die

Beziehung zwischen manifester Reaktion und latenter Eigenschaft kann mit Hilfe einer Wahrscheinlichkeitsfunktion dargestellt werden. Grundlegend basiert die Item-Response-Theorie auf der logistischen Funktion (Fischer, 1974; Kubinger, 2003b; 2009; Kubinger, Rasch, & Yanagida, 2011; Molenaar, 1995).

3.2 Das dichotome logistische Testmodell von Rasch

Das dichotome logistische Testmodell des Statistikers Georg Rasch (abgekürzt Rasch-Modell) bildet die Grundlage der Item-Response-Theorie. Ausgehend von diesem Modell wurden zahlreiche weitere probabilistische Modelle entwickelt, die als Spezialfälle des Rasch-Modells oder dessen Verallgemeinerung zu betrachten sind (Kubinger, 2003b; 2005).

Wie in Kapitel 3.1 kurz angedeutet, beschreibt das dichotome logistische Modell von Rasch den Zusammenhang einer beobachtbaren Reaktion und deren zugrunde liegende latente Eigenschaft mit Hilfe einer Wahrscheinlichkeitsfunktion. Der Begriff „dichotom“ bezieht sich dabei auf die Konstruktion des Modells auf Basis von zweikategoriellen (sog. dichotomen) Daten, d.h. Aufgaben werden nur als „gelöst“ („+“) oder „nicht gelöst“ („-“) kodiert; die Bezeichnung „logistisch“ beruht auf der Verwendung einer logistischen Funktion, wie sie im Folgenden dargestellt ist (Kubinger et al., 2011):

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

Konkret beschreibt das Rasch-Modell die Wahrscheinlichkeit, dass eine bestimmte Testperson v das Item i löst („+“). Grundlage dieser Beziehung bilden zwei Parameter. Zum einen ein sogenannter Personenparameter ξ_v , welcher eine reelle Zahl darstellt und das Fähigkeitsniveau der Testperson v in einer latenten Dimension x beschreibt. Zum anderen ein Itemparameter σ_i ,

welcher das Schwierigkeitsniveau bzw. den Grad der Schwierigkeit des Items i in selbiger Dimension x angibt. Bei konstant bleibender Schwierigkeit des Items i und zunehmender Fähigkeit der Person würde dementsprechend auch die Wahrscheinlichkeit steigen dieses Item i zu lösen. Die Fähigkeit einer Testperson sowie die Schwierigkeit eines Items werden dabei jeweils durch nur einen Parameter geschätzt. Dies bedeutet, dass beide Parameter Ausprägungen einer zugrunde liegenden, latenten Dimension repräsentieren und *eindimensional* erhoben werden (Kubinger, 2003b; 2005; Kubinger et al., 2011; Waldherr, 2001).

Daraus ergeben sich einige Vorteile des Rasch-Modells, wie beispielsweise die *lokal stochastische Unabhängigkeit*. Diese besagt, dass die Wahrscheinlichkeit dafür ob eine Person ein Item löst unabhängig davon ist ob, welche und wie viele Items sie zuvor bearbeitet bzw. gelöst hat. Die Lösungswahrscheinlichkeit hängt bei jeder Aufgabe allein von deren Schwierigkeit, der Fähigkeit der Person und vom Zufall ab. Dadurch sind auch mögliche Lerneffekte während der Bearbeitung eines Tests ausgeschlossen (Kubinger, 2003b; 2005; Kubinger et al., 2011).

Ein weiteres Merkmal des Rasch-Modells ist die *spezifische Objektivität*. Sie besagt einerseits, dass Testpersonen miteinander vergleichbar sind, unabhängig davon wie viele und welche Items sie zuvor bearbeitet haben; andererseits können auch Items miteinander verglichen werden, unabhängig davon welche Personen sie bearbeitet haben. In diesem Sinn ist das Rasch-Modell als stichprobenunabhängig zu bezeichnen. Die Wahl der Stichprobe hat also, testtheoretisch gesehen, keinen Einfluss auf den Vergleich der Items (Dimitz & Formann, 1979; Kubinger, 2003b; 2005; Waldherr, 2001).

Der Rohwert (Anzahl der gelösten Aufgaben) eines nach dem Rasch-Modell-kalibrierten Verfahrens wird als Testwert herangezogen und enthält im Sinne einer *erschöpfenden Statistik* die gesamte Information über die interessierende Fähigkeit (Dimitz &

Formann, 1979; Fischer, 1974; Molenaar, 1995). Die Anzahl gelöster Aufgaben entspricht einer adäquaten Abbildung von empirischen Verhaltensrelationen (Kubinger, 2003b; Kubinger et al., 2011). Entspricht ein Test den genannten Annahmen und Anforderungen, so erfüllt dieser das Gütekriterium *Skalierung*⁵ (Kubinger, 2009).

Für die Überprüfung von Tests auf Rasch-Modell-Konformität wird primär das Merkmal der Stichprobenunabhängigkeit genutzt. Bei einer willkürlichen Teilung der Gesamtstichprobe in zwei Untergruppen, z.B. männlich und weiblich, müssen die Itemparameterschätzungen in beiden Teilstichproben gleich ausfallen. Anderenfalls würden die jeweiligen Items, bei denen es in den Untergruppen zu unterschiedlichen Schätzergebnissen kommt, einer der beiden Gruppen leichter fallen und diese Personen „bevorzugen“. Die entsprechenden Items wären in diesem Fall nicht Modell-konform und aus dem Test auszuschließen (Dimitz & Formann, 1979; Kubinger et al., 2011). Inferenzstatistisch erfolgt die Überprüfung eines Verfahrens auf Rasch-Modell-Gültigkeit mit Hilfe des Likelihood-Ratio-Tests nach Andersen (1973). Für die Unterteilung der Gesamtstichprobe kann sowohl ein *internes* Teilungskriterium, welches die Gruppe anhand der Anzahl gelöster Aufgaben unterteilt, als auch ein *externes* Teilungskriterium, wie z.B. das Geschlecht, das Alter oder die Muttersprache, gewählt werden (Kubinger, 2003b; 2005; Kubinger et al., 2011). Zusätzlich kann eine sogenannte graphische Modellkontrolle durchgeführt werden. Dabei werden die Parameterschätzungen pro Item der jeweiligen Teilstichproben in einem rechtwinkligen Koordinatensystem gegenübergestellt (Kubinger, 2009; Kubinger et al., 2011). Eine weitere Möglichkeit zur Überprüfung des Rasch-Modells bietet der z-Test nach Fischer und Scheiblechner (1970, zitiert nach Kubinger et al., 2011, S. 556-557).

Die praktische Anwendung dieser Prüfverfahren wird in Kapitel 8.1 ausführlich beschrieben.

⁵ Für eine ausführliche Beschreibung des Gütekriteriums *Skalierung* siehe beispielsweise Kubinger (2009; 2003a).

III. EMPIRISCHER TEIL

4. Ziele der Untersuchung und Hypothesen

Ziel der vorliegenden Diplomarbeit war ein Dimensionalitätsvergleich dreier Reasoning-Tests, wobei untersucht werden sollte, ob die Tests jeweils unterschiedliche und unabhängige Konstrukte von Reasoning erfassen. Folgende Verfahren wurden für die Untersuchung verwendet:

- *Numerische Topologien (N-TOP)* (Kubinger & Heuberger, in Vorb.)
- *Family-Relation-Reasoning-Test (FRRT)* (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.)
- *Wiener Matrizen-Test 2 (WMT 2)* (Formann, Waldherr, & Piswanger, 2011)⁶

Anhand des dichotomen logistischen Modells sollte zudem die Eindimensionalität respektive Rasch-Modell-Konformität der Items des *N-TOP* sowie des *FRRT*, welche sich im Gegensatz zum hinreichend untersuchten *WMT 2* beide noch in Entwicklung befinden, überprüft werden (siehe Abschnitt 8.1).

Anhand der Personenparameter erfolgte anschließend eine Untersuchung korrelativer Zusammenhänge zwischen den drei Reasoning-Tests (siehe Abschnitt 8.2). Auf diese Weise sollte die diskriminante Validität⁷ der Verfahren bestimmt werden. Damit sich die Untersuchung nicht auf eine rein deskriptive Auswertung beschränkt, wurde ein Grenzwert

⁶ Der *Wiener Matrizen-Test 2* wurde freundlicherweise vom Hogrefe Verlag für diese Untersuchung zur Verfügung gestellt. Für nähere Informationen, wie Testinstruktion, Beispielitems etc. wenden Sie sich bitte direkt an Hogrefe bzw. sei auf das Manual verwiesen.

⁷ Die diskriminante Validität ist eine Facette der Konstruktvalidität, welche zum übergeordneten Gütekriterium Validität zählt. Sie dient zur Validierung eines Messinstrumentes. Konkret bezieht sie sich auf Korrelationen zwischen Verfahren, die vorgeben unterschiedliche Konstrukte zu erheben. Die Korrelationen sollten folglich im Idealfall gegen Null gehen. Den Gegenpol zur diskriminanten Validität bildet die konvergente Validität (siehe z.B. Kubinger, 2009, S. 55-67; 2003b, S. 199-200).

anhand theoretischer Überlegungen festgelegt (Moosbrugger & Kelava, 2012). Dieser Grenzwert bestimmt die maximal bzw. minimal zu tolerierende Korrelation zwischen Verfahren, die vorgeben unterschiedliche oder gleiche Leistungsbereiche zu erfassen.

Aus den genannten Zielen lassen sich folgende Hypothesen ableiten:

$H_0(1)$

Die Items des Verfahrens *N-TOP* sind konform mit dem dichotomen logistischen Modell von Rasch.

$H_0(2)$

Die Items des Verfahrens *FRRT* sind konform mit dem dichotomen logistischen Modell von Rasch.

$H_0(3)$

Der Zusammenhang zwischen den Reasoning-Tests, die denselben Intelligenzbereich (fluid vs. kristallisiert) erheben ist größer als der festgelegte Grenzwert, während die Reasoning-Tests, die unterschiedliche Konstrukte erfassen einen maximal dem Grenzwert entsprechenden Zusammenhang aufweisen.

5. Beschreibung der Testbatterie

Im Rahmen dieser Untersuchung wurde eine Testbatterie, bestehend aus drei Reasoning-Tests, einem Fragebogen und einem Datenblatt zur Erhebung der demographischen Daten erstellt. In den folgenden Unterkapiteln (5.1 – 5.4) werden die einzelnen Verfahren, die Fähigkeiten, die sie erfassen, sowie das Aufgabenmaterial und die Aufgabenstellungen näher erläutert.

Um einen möglichst großen Pool an Items testen zu können und Abschreiben während der Gruppentestungen zu verhindern, wurden vom *N-TOP* und dem *FRRT* verschiedene Testversionen mit teilweise unterschiedlichen Items verwendet. In Abhängigkeit aller zur Verfügung stehenden Items wurden vom *N-TOP* zwei und vom *FRRT* drei Versionen erstellt. Somit ergaben sich sechs mögliche Kombinationen von Testheften. Tabelle 1 gibt einen Überblick über die untersuchten Reasoning-Tests und ihre Zuordnung zum fluiden bzw. kristallisierten Intelligenzbereich sensu Cattell sowie die Art des jeweilig verwendeten Materials. Tabelle 2 illustriert die sechs Testheftvarianten.

Tabelle 1

Überblick über die untersuchten Reasoning-Tests

Verfahren	Zuordnung sensu Cattell	Material
<i>N-TOP</i>	fluid	numerisch
<i>FRRT</i>	kristallisiert	lexikalisch
<i>WMT 2</i>	fluid	figural

Tabelle 2*Testheftvarianten*

Testheftversion	Deckblatt	<i>FRRT</i>	<i>N-TOP</i>	<i>WMT 2</i>	<i>AVT</i>
1	A	A	A	A	A
2	A	B	B	A	A
3	A	C	A	A	A
4	A	A	B	A	A
5	A	B	A	A	A
6	A	C	B	A	A

5.1 *Numerische Topologien (N-TOP)*

Das Verfahren *Numerische Topologien* (Kubinger & Heuberger, in Vorb.) ist ein Test zur Erfassung des fluiden schlussfolgernden Denkens mithilfe numerischen Materials.⁸ Dabei soll die Testperson Gesetzmäßigkeiten erkennen und diese entsprechend anwenden. Wie in Abbildung 1 demonstriert, werden drei große Kästchen präsentiert, in denen sich Zahlen befinden. Anhand der gegebenen Zahlen soll eine oder mehrere Gesetzmäßigkeiten gefunden werden. Im rechten großen Kästchen soll folgerichtig in das leerstehende Viereck die fehlende Zahl eingetragen werden. Aus den zur Verfügung stehenden Items wurden zwei Testversionen (A und B) mit jeweils 16 Aufgaben erstellt. Die Bearbeitungsdauer betrug 10 Minuten.

⁸ Für vorhandene Untersuchungen zur Rasch-Modell-Gültigkeit des *N-TOP* sei auf die Diplomarbeiten von Ünal (2014) und Gamsjäger (2012) verwiesen.

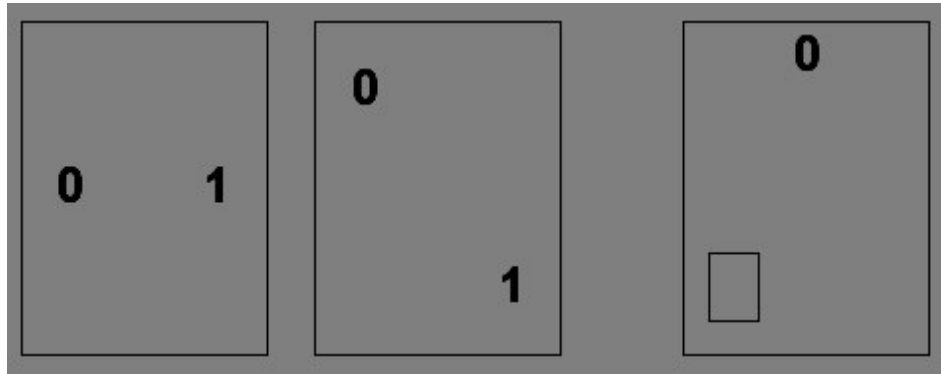


Abbildung 1. Itembeispiel aus *N-TOP* (Kubinger & Heuberger, in Vorb.).

5.2 *Family-Relation-Reasoning-Test (FRRT)*

Der *FRRT* (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.) ist ein Verfahren zur Messung des kristallisierten schlussfolgernden Denkens.⁹ Er bedient sich dabei lexikalischem Material. Die Testperson erhält die Aufgabe aus Sätzen über Familienbeziehungen durch logisches Kombinieren eine fehlende Verwandtschaftsbeziehung zu finden. Zur Veranschaulichung ist in Abbildung 2 ein Beispielitem angeführt. Aus einem Itempool von 29 Items wurden insgesamt drei Testhefte erstellt (Version A, B, C), die jeweils aus 20 der insgesamt 29 Items bestehen. Für deren Bearbeitungen hatten die Testpersonen 15 Minuten Zeit.

Karin hat die Tochter Tina. Peter ist der Bruder von Tina. Karin ist Peters _____.
--

Abbildung 2. Itembeispiel aus *FRRT* (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.).

⁹ Zur Vertiefung sei auf die Diplomarbeit von Skoda (2005) verwiesen.

5.3 *Wiener Matrizen-Test 2 (WMT 2)*

Der *WMT 2* (Formann, Waldherr, & Piswanger, 2011) ist ein Matrizentest, der die Fähigkeit zum logisch-schlussfolgernden Denken im fluiden Intelligenzbereich sensu Cattell erhebt. Er bedient sich dabei figuralem Material und erhebt das Konstrukt Reasoning in sprachfreier Form. Die Aufgaben bestehen jeweils aus einer 3x3-Matrix mit Figuren, wobei das rechte unterste Feld immer leer steht. Aus acht Antwortalternativen soll die Testperson der gegebenen Ordnung folgend die fehlende Figur auswählen. Dabei gilt es sowohl horizontale als auch vertikale Gesetzmäßigkeiten zwischen den einzelnen Figuren zu erkennen.

Der *WMT 2* besteht insgesamt aus 18 Items für deren Bearbeitung die Testpersonen eine Zeitvorgabe von 15 Minuten hatten. Für weitere Informationen sowie Beispielitems wird auf das Manual des *WMT 2* verwiesen.

5.4 *Anstrengungsvermeidungstest für ältere Schüler (AVT-Sch)*

Der *Anstrengungsvermeidungstest für ältere Schüler*, kurz *AVT-Sch* (Rollett & Rollett, 2000, zitiert nach Gashi, 2011, S. 41), ist ein Fragebogen zur Erhebung des Ausmaßes von negativ gerichteten motivationalen Tendenzen. Der *AVT-Sch* ist eine Adaption des *Anstrengungsvermeidungstests für Kinder und Jugendliche* von Rollett und Bartram (1998). „Anstrengungsvermeidung“ wird nach Rollett und Bartram (1998, S. 7-8) als die Tendenz bezeichnet, bei der durch den aktiven Einsatz geeigneter Vermeidungsstrategien versucht wird, sich den emotionell negativ erlebten Anstrengungen zu entziehen, die mit dem notwendigen Leistungseinsatz in bestimmten Tätigkeitsfeldern verbunden werden. Im Zuge dieser Untersuchung wurde der *AVT-Sch* als zusätzliches Teilungskriterium der Stichprobe für die Überprüfung auf Rasch-Modell-Konformität verwendet.

Da der Fragebogen ursprünglich für ältere Schülerinnen und Schüler konzipiert wurde, wurden für diese Untersuchung einzelne Items für den Erwachsenenbereich umformuliert und das Aufgabenmaterial gekürzt. Beispielsweise wurden statt Wörtern wie „Schule“ oder „Lehrer“, „Arbeit“ und „man“ verwendet. Überdies wurde das Antwortformat von sechs auf vier Antwortkategorien reduziert. Insgesamt bestand der Fragebogen aus 32 Items. Eine explizite Zeitvorgabe zum Ausfüllen gab es nicht. Die Testpersonen wurden jedoch darum gebeten den Fragebogen, als letzten Teil der Testbatterie, zügig zu bearbeiten. Für weitere Informationen zum *AVT-Sch* wird auf die Diplomarbeit von Gashi (2011) verwiesen.

6. Planung und Durchführung der Untersuchung

Für die in Kapitel 4 ausführlich beschriebenen Ziele der Untersuchung wurde im Vorfeld der Stichprobenakquirierung festgelegt, dass die Stichprobe eine Größe von ca. 300 Testpersonen umfassen soll. Da Reasoning-Tests häufig in der Bildungsberatung sowie der Berufseignungsdiagnostik Anwendung finden, wurde versucht, die Erhebung in Bildungsinstitutionen durchzuführen, die derartige Verfahren - beispielsweise im Zuge von Auswahlverfahren - verwenden. Als Folge dieser Überlegungen wurden die Testungen in der Sicherheitsakademie Traiskirchen der österreichischen Bundespolizei und zwei Krankenpflegeschulen in Wien und Ybbs realisiert.

Um die Genehmigungen für die jeweiligen Institutionen zu erhalten, wurden Anfragen und entsprechende Anträge an das Bundesministerium für Inneres (siehe im Anhang unter B) sowie die Generaldirektion KAV des Geschäftsbereichs Pflegemanagement (siehe im Anhang unter C) gestellt. Nach der offiziellen Bewilligung der Untersuchung wurden in Absprache mit den jeweiligen Verantwortlichen der Bildungseinrichtungen Termine für die Testungen vereinbart. In der Sicherheitsakademie wurde ein Termin für insgesamt fünf

Gruppentestungen ausgemacht. In der Krankenpflegeschule Baumgartner Höhe in Wien wurden drei Termine für ebenfalls fünf Klassen und in der Krankenpflegeschule Ybbs zwei Termine für insgesamt drei Klassen vereinbart. Dabei wurde im Besonderen darauf geachtet, den regulären Schulbetrieb so wenig wie möglich zu stören. Aus diesem Grund fanden die Testungen über einen Zeitraum von insgesamt vier Monaten statt. Mit einem Informationsschreiben wurden die zuständigen Lehrkräfte im Vorfeld über die geplante Erhebung unterrichtet (siehe im Anhang unter D). Sie wurden gebeten ihre Klassen über Zeitpunkt und Ablauf der Untersuchung zu informieren.

Die Testungen fanden innerhalb der regulären Unterrichtszeiten in den Räumlichkeiten der jeweiligen Schulen statt. Die Gruppengröße schwankte dabei zwischen 17 und 40 Teilnehmerinnen und Teilnehmern. Die erste Datenerhebung fand am 24.02.2015 in Traiskirchen statt. Die Testungen in Wien wurden am 27.03., 09.04. und 27.04.2015; jene in Ybbs am 22.04. sowie am 06.05.2015 durchgeführt. Bei jeder Testung waren mindestens zwei Testleiter anwesend um das Austeilen und Einsammeln der Testhefte zu beschleunigen und die Kontrolle bezüglich der Einhaltung der Testleiteranweisungen zu gewährleisten. Neben der Diplomandin waren alle Helfer ebenfalls Psychologiestudenten der Universität Wien, die im Vorfeld über den Erhebungsablauf und ihre Aufgaben informiert wurden (siehe im Anhang unter E). Die gesamte Durchführung einer Testung nahm je nach Klassengröße ca. 50 bis 60 Minuten in Anspruch.

Zu Beginn der Testungen wurden die Teilnehmerinnen und Teilnehmer darüber informiert, dass sämtliche Daten anonymisiert und freiwillig erhoben und für rein wissenschaftliche Zwecke genutzt werden würden. Es wurde ebenfalls versichert, dass die individuellen Ergebnisse der Testung keinerlei Auswirkungen auf schulische Bewertungen haben würden. Währenddessen teilte der Testhelfer an jede Teilnehmerin und jeden

Teilnehmer ein Testheft aus, mit der Bitte erst nach Aufforderung das Deckblatt umzublättern. Die Testhefte lagen in sechs Varianten vor (siehe Tabelle 2). Testpersonen, die nebeneinander saßen, erhielten jeweils eine andere dieser Varianten, so dass abwechselnd eine Person Version 1, die nächste Version 2, die nächste Version 3 usw. vor sich liegen hatten. Dabei wurde darauf geachtet, dass die einzelnen Varianten ungefähr gleich häufig eingesetzt wurden. Die genaue Itemanordnung der einzelnen Untertests der Testbatterie ist Anhang A zu entnehmen.

Um die Anonymität zu gewährleisten und trotzdem jeder Teilnehmerin und jedem Teilnehmer die Möglichkeit zu bieten, die eigenen Testergebnisse zu erfahren, wurde jeder Testperson ein sechstelliger Personencode mit den Kontaktdaten (E-Mail-Adresse) der Diplomandin zugeteilt. Der Personencode setzte sich aus drei Buchstaben und drei Ziffern zusammen, wobei der erste Buchstabe den Nachnamen der Diplomandin angab, die folgenden Buchstaben den Ort der Erhebung und die Ziffern eine fortlaufende Nummerierung der Testhefte wiedergaben. Im Anhang G ist ein Beispiel eines schriftlichen Ergebnisberichts für die Teilnehmerinnen und Teilnehmer zu finden.¹⁰

Nach den allgemeinen Erläuterungen und Anweisungen zur Erhebung wurden die Testpersonen zunächst gebeten, das Datenblatt auszufüllen. Hierbei wurden das Alter, das Geschlecht, die Muttersprache und die höchste abgeschlossene Ausbildung erfasst.

Danach wurde mit der Testbearbeitung begonnen. Bei jedem Test wurde die Instruktion¹¹ samt Aufgabenbeispielen vorgelesen und erklärt. Dabei wurden die Testpersonen aufgefordert, bei Verständnisschwierigkeiten nachzufragen, um spätere Störungen während der Testbearbeitung zu vermeiden. Den Teilnehmerinnen und Teilnehmern wurde vor Beginn der

¹⁰ Insgesamt baten 35 Personen aus der Stichprobe um eine Rückmeldung ihrer Ergebnisse.

¹¹ Die Instruktionen für die Bearbeitung der jeweiligen Tests sind im Anhang unter F zu finden.

einzelnen Tests die jeweilige Bearbeitungszeit sowie die Anzahl der vorgegeben Items pro Test mitgeteilt (siehe dazu Tabelle 3).

Tabelle 3

Bearbeitungszeit und Itemanzahl pro Verfahren

Verfahren	Itemanzahl	Bearbeitungszeit
<i>FRRT</i>	20	15 Minuten
<i>N-TOP</i>	16	10 Minuten
<i>WMT 2</i>	18	15 Minuten
<i>AVT-Sch</i>	32	keine Zeitbeschränkung

Nach Ablauf der jeweiligen Bearbeitungszeit wurden die Testpersonen gebeten einen Strich unter die zuletzt bearbeitete Aufgabe zu ziehen. Dies hatte den Grund bei der Dateneingabe unterscheiden zu können, ob eine Person ein Item zwar bearbeitet hatte, aber keine Antwort finden konnte und die Aufgabe deshalb nicht gelöst wurde oder ob eine Person aus Zeitmangel nicht mehr zur Bearbeitung des Items gekommen war.

7. Stichprobe

Die Stichprobe setzt sich aus insgesamt 292 Auszubildenden der Sicherheitsakademie Traiskirchen und der Krankenpflegeschulen Wien (Baumgartner Höhe) und Ybbs zusammen. Davon sind 50,2% weiblich und 49,8% männlich. Im Detail besteht die Stichprobe aus 113 Polizei-, 121 Krankenpflegeschülerinnen und -schülern aus Wien sowie 58 Krankenpflegeschülerinnen und -schülern aus Ybbs. Wie in Abbildung 3 veranschaulicht reicht die Spannweite des Alters der Testpersonen von 18 bis 54 Jahren ($Md_{Alter} = 24$ Jahre).

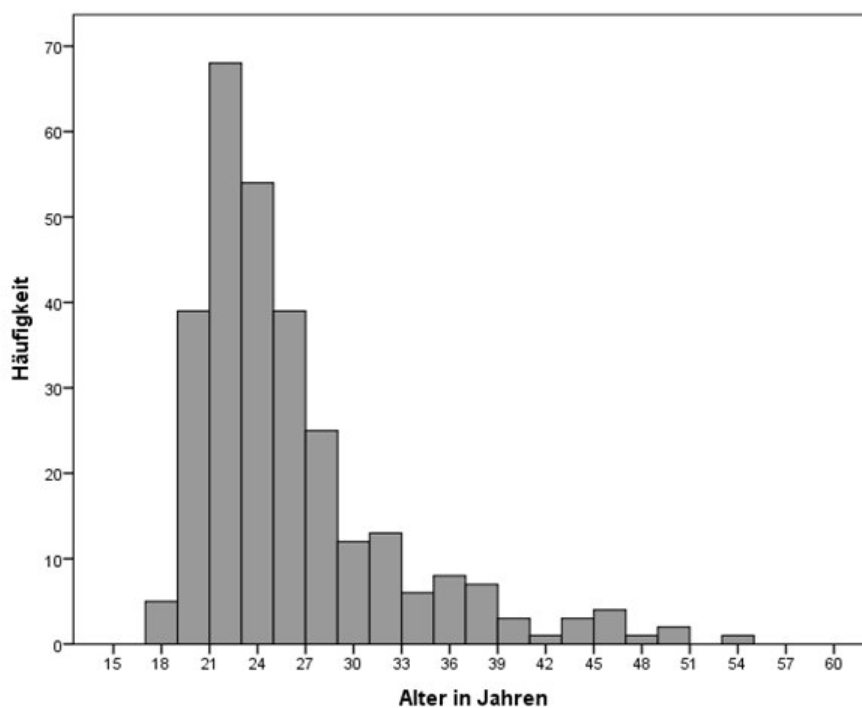


Abbildung 3. Verteilung der Stichprobe in Bezug auf das Alter in Jahren ($n = 291$)

93,5% geben Deutsch als ihre Muttersprache an. Die genaue Verteilung der höchsten abgeschlossenen Ausbildung geht aus Tabelle 4 hervor. Mit 31,5% haben die meisten Testpersonen einen Abschluss von einer allgemeinbildenden höheren Schule vorzuweisen, gefolgt von 29,5% mit einer abgeschlossenen Lehre.

Tabelle 4

Verteilung der Stichprobe in Bezug auf die höchste abgeschlossene Ausbildung (n = 289)

Höchste abgeschlossene Ausbildung	absolute Häufigkeit	Prozent
Pflichtschule	26	9,0
Lehre	86	29,8
AHS	92	31,8
BHS	64	22,1
Hochschule	7	2,4
Sonstiges	17	4,8

Anmerkung. Unter „Sonstiges“ fallen alle Angaben, die keiner der übrigen Kategorien zuzuordnen sind, bspw. Fachschulen.

In Bezug auf das Teilungskriterium Anstrengungsvermeidungsverhalten illustriert Abbildung 4 die Verteilung.

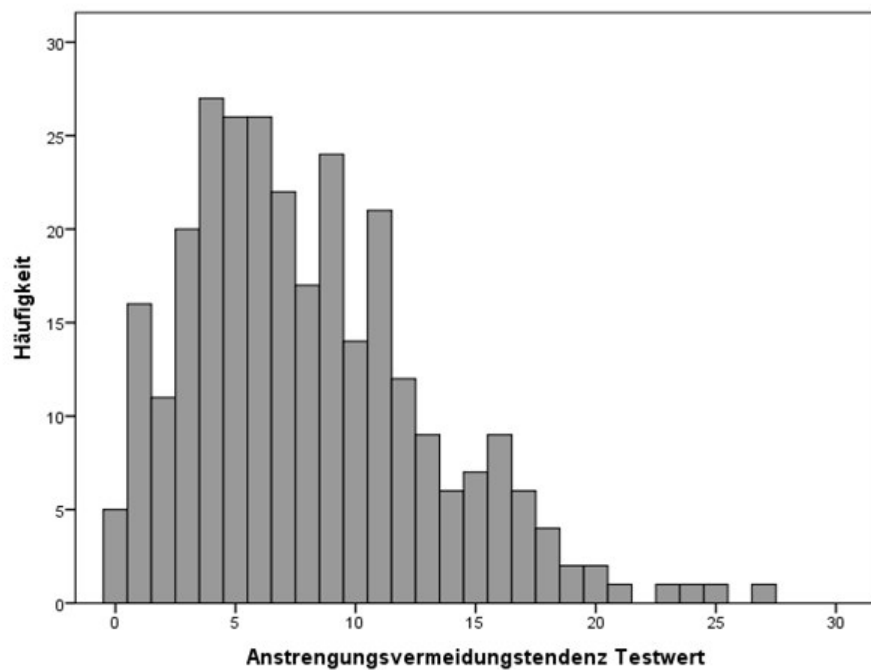


Abbildung 4. Verteilung der Stichprobe in Bezug auf die Anstrengungsvermeidungstendenz (n = 291)

8. Darstellung der Ergebnisse

Im Folgenden werden die Ergebnisse der Analysen zur Geltung des Rasch-Modells für den *N-TOP* und den *FRRT* im Kapitel 8.1 und den dazugehörigen Unterkapiteln detailliert vorgestellt. Das Kapitel 8.2 beschäftigt sich anschließend mit den Ergebnissen der Korrelationen zwischen den Reasoning-Tests *N-TOP*, *FRRT* und *WMT 2*.

Für die statistischen Analysen und die Auswertung der Daten wurden das Statistikprogramm IBM SPSS Statistics (Version 20.0.0) und die Open-Source-Statistik-Software R (Version 3.2.0; R Core Team, 2015) mit dem zusätzlichen Programmpaket *eRm* (extended Rasch modeling; Version 0.15-4; Mair, Hatzinger, & Maier, 2014) verwendet.¹²

8.1 Analysen zur Überprüfung der Geltung des Rasch-Modells

Für die Überprüfung der Geltung des Rasch-Modells der Items aus *N-TOP* und *FRRT* werden folgende Teilungskriterien verwendet:

- **Bildungsniveau:**
niedrig (Pflichtschule, Lehre, Sonstiges) vs. hoch (AHS, BHS, Hochschule)
- **Alter:**
18- bis 24-Jährige vs. > 24- bis 54-Jährige
- **Anstrengungsvermeidungstendenz:**
niedrig (Testwert $\leq Md_{AVT-Sch} = 7$) vs. hoch (Testwert $> Md_{AVT-Sch} = 7$)

¹² Zur Vertiefung mit dem Programm R und dem Zusatzpaket *eRm* sei auf das Manual von Poinstingl, Mair und Hatzinger (2007) sowie das Werk von Adler (2012) verwiesen.

Mithilfe des Likelihood-Ratio-Tests (LRT) nach Andersen (1973) wird überprüft, ob die Itemparameterschätzungen in den jeweiligen Teilstichproben statistisch gleich sind. Da hierbei mehrere Signifikanztests durchgeführt werden und dies das Risiko vergrößert, eine irrtümliche Abweichung vom Rasch-Modell festzustellen, wird das akzeptierte untersuchungsbezogene Risiko 1. Art auf $\alpha = 0,01$ festgelegt (Kubinger, 2005, S. 387-388; siehe auch Kubinger et al., 2011). Sollte es zu signifikanten Ergebnissen bei der Durchführung der Likelihood-Ratio-Tests kommen, so ist von einer Modellabweichung eines oder mehrerer Items auszugehen. Für diesen Fall werden die Daten einer grafischen Modellkontrolle unterzogen bei der jene Items, die weit entfernt einer Teilungsgeraden liegen, konkret deren 95%-Konfidenzellipsen die 45°-Gerade nicht schneiden, als nicht modellkonform bestimmt werden. Für jedes Item kann eine solche Konfidenzellipse erstellt werden. Diese umschließt den Bereich, in welchem mit einer Wahrscheinlichkeit von 95% die tatsächliche Itemschwierigkeit liegt. Beträgt die Differenz der Parameterschätzung eines Items dabei mehr als ein Zehntel der Spannweite der gesamten Itemschwierigkeitsparameterschätzung, kann die Modellabweichung des jeweiligen Items als relevant betrachtet werden (Goethals, 1994, zitiert nach Kubinger, 2005, S. 389-390).

Als zusätzliches Prüfkriterium wird der z-Test nach Fischer und Scheiblechner (1970, zitiert nach Kubinger et al., 2011, S. 556-557) gerechnet, der jedoch aufgrund der Überhöhung des Risikos 1. Art rein deskriptiv zu interpretieren ist.

Die als nicht modellkonform identifizierten Items sollen daraufhin iterativ aus dem Itempool entfernt werden, bis die Ergebnisse der Likelihood-Ratio-Tests für alle Teilungskriterien nicht mehr signifikant sind und somit a-posteriori Modellgeltung erreicht wird. Items, die von allen oder keiner Testperson einer Teilstichprobe gelöst wurden, können für das betreffende Teilungskriterium nicht geschätzt und daher in besagtem Fall auch nicht berücksichtigt werden. Die Anzahl der schätzbaren Items beträgt immer um eins mehr als die

Anzahl der Freiheitsgrade (df) und kann für alle Teilungskriterien den jeweiligen Tabellen über die Ergebnisse der Likelihood-Ratio-Tests entnommen werden .

8.1.1 Ergebnisse der Rasch-Modell-Analysen beim *N-TOP*

Der *N-TOP* (Kubinger & Heuberger, in Vorb.) besteht aus insgesamt 16 Items, die alle auf Modellgeltung überprüft wurden ($n = 292$).

Tabelle 5 zeigt, dass die Likelihood-Ratio-Tests nach Andersen in den Teilungskriterien „Alter“ und „Anstrengungsvermeidungstendenz“ nicht signifikant ausfallen. Im Teilungskriterium „Bildungsniveau“ kommt es jedoch zu einer signifikanten Modellabweichung.

Tabelle 5

Ergebnisse der Likelihood-Ratio-Tests (N-TOP, alle Items)

Teilungskriterium	χ^2 (LRT)	df	χ_{krit}^2	p
Bildungsniveau	31,722	15	30,578	0,007*
Alter	21,145	15	30,578	0,132
Anstrengungsvermeidungstendenz	23,528	15	30,578	0,074

Anmerkung. () = $p < 0,01$*

Die grafischen Modellkontrollen ohne Konfidenzellipsen sind in den Abbildungen 5, 6 und 7 für die drei Teilungskriterien zu finden.

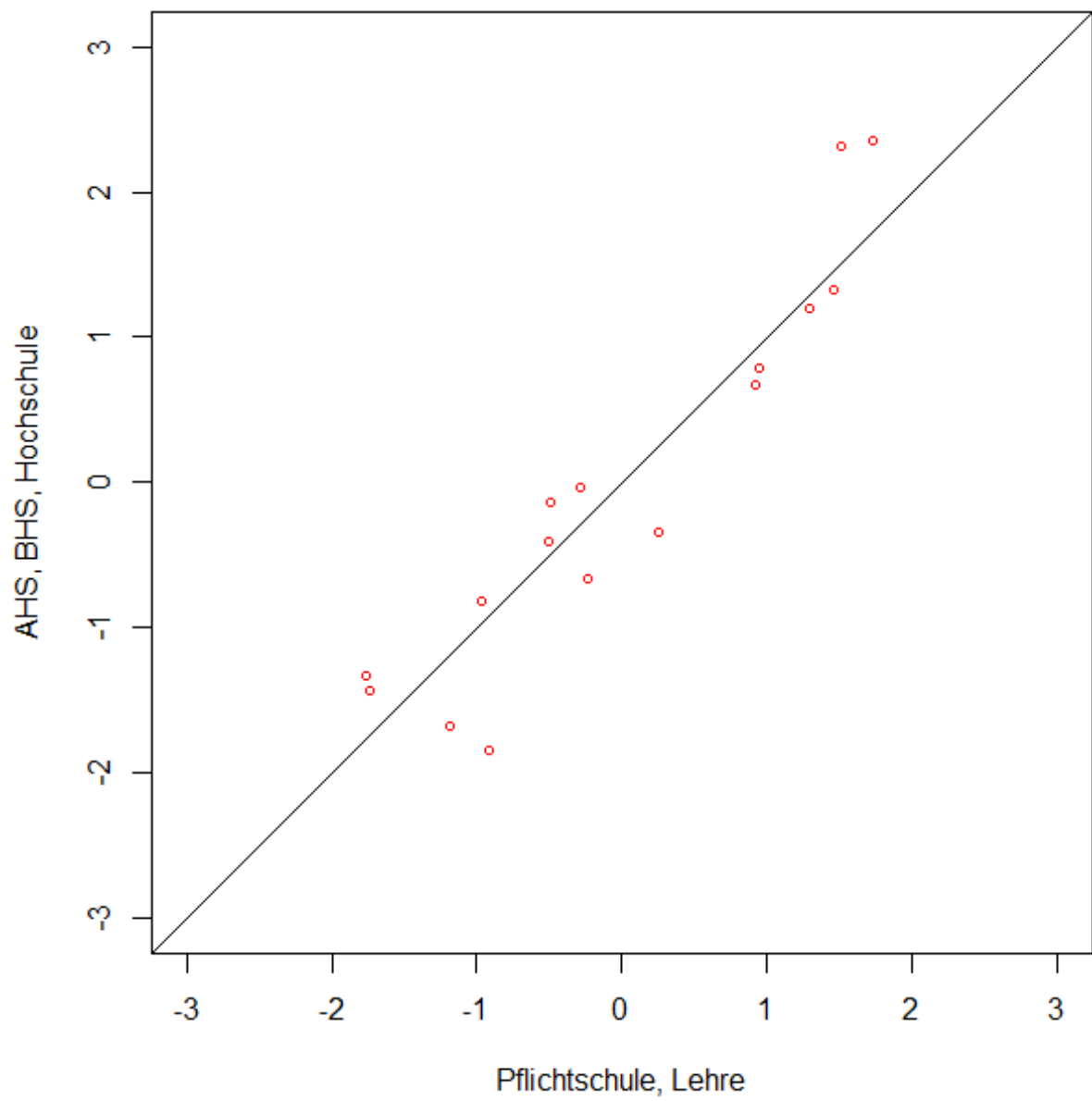


Abbildung 5. Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (*N-TOP*, alle Items)

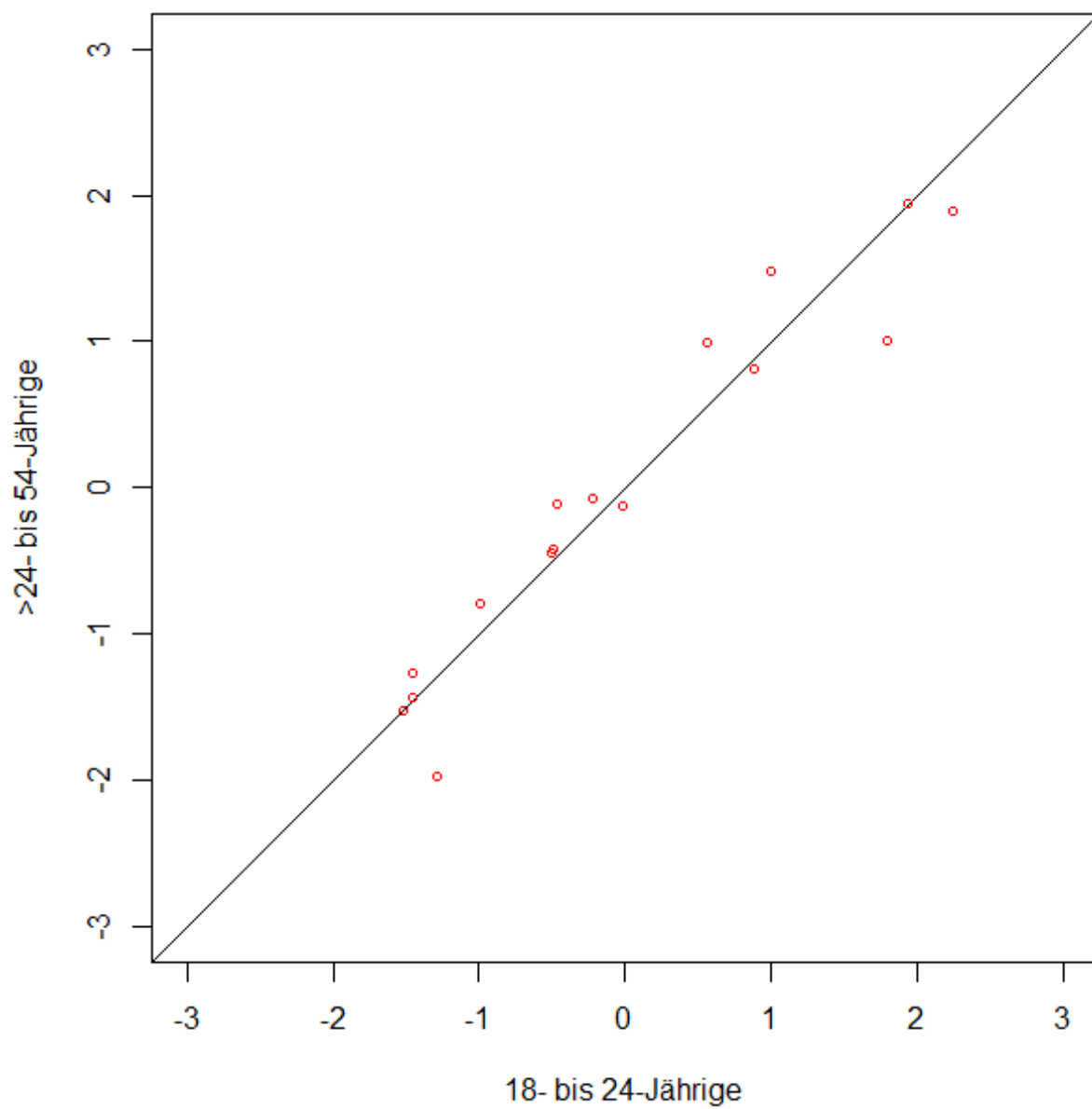


Abbildung 6. Grafische Modellkontrolle für das Teilungskriterium „Alter“ (*N-TOP*, alle Items)

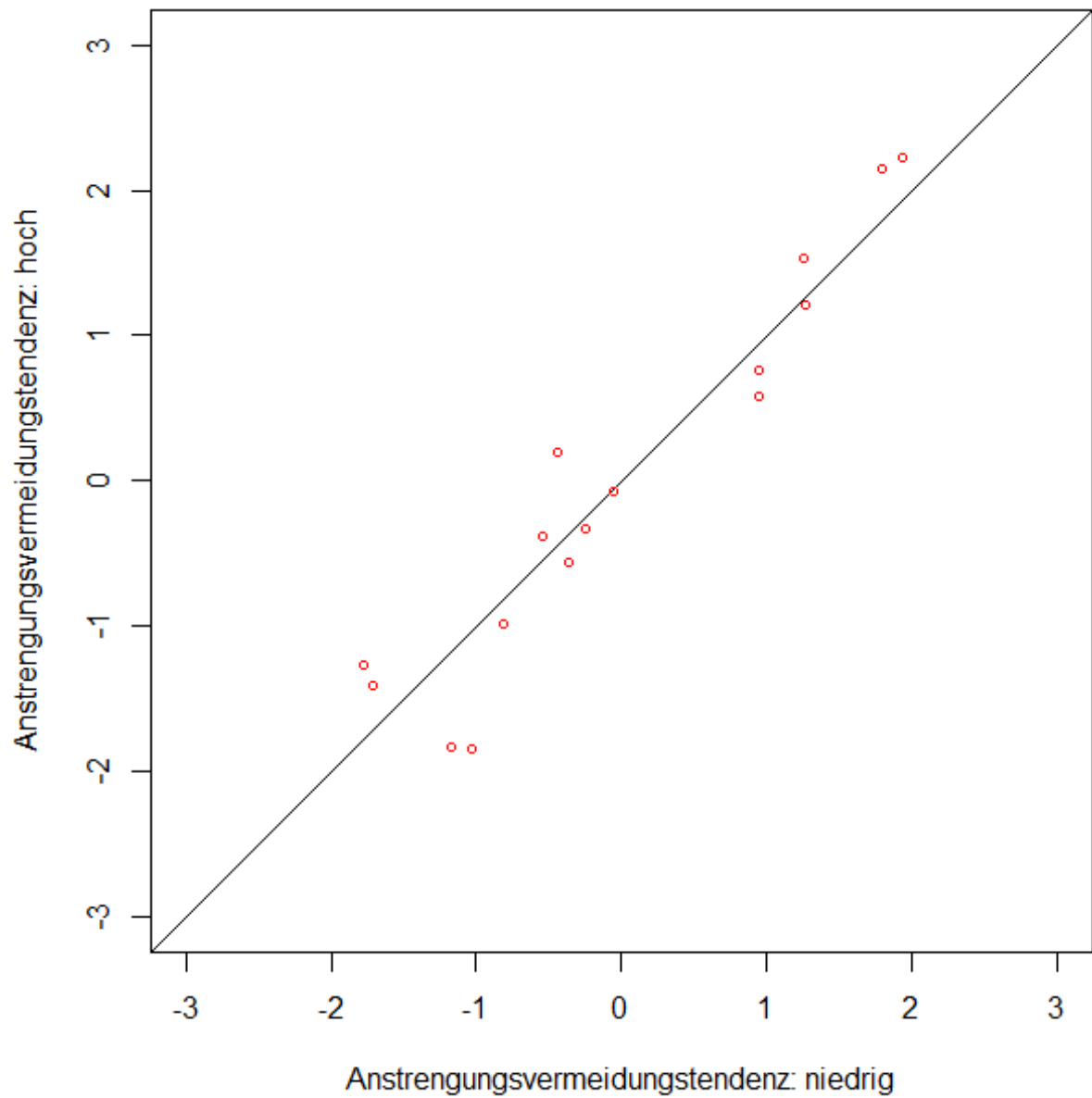


Abbildung 7. Grafische Modellkontrolle für das Teilungskriterium „Anstrengungsvermeidungstendenz“ (*N-TOP*, alle Items)

Im Folgenden zeigen die Abbildungen 8, 9 und 10 die grafischen Modellkontrollen der Teilungskriterien „Bildungsniveau“, „Alter“ und „Anstrengungsvermeidungstendenz“ mit beschrifteten Konfidenzellipsen der auffälligen Items.

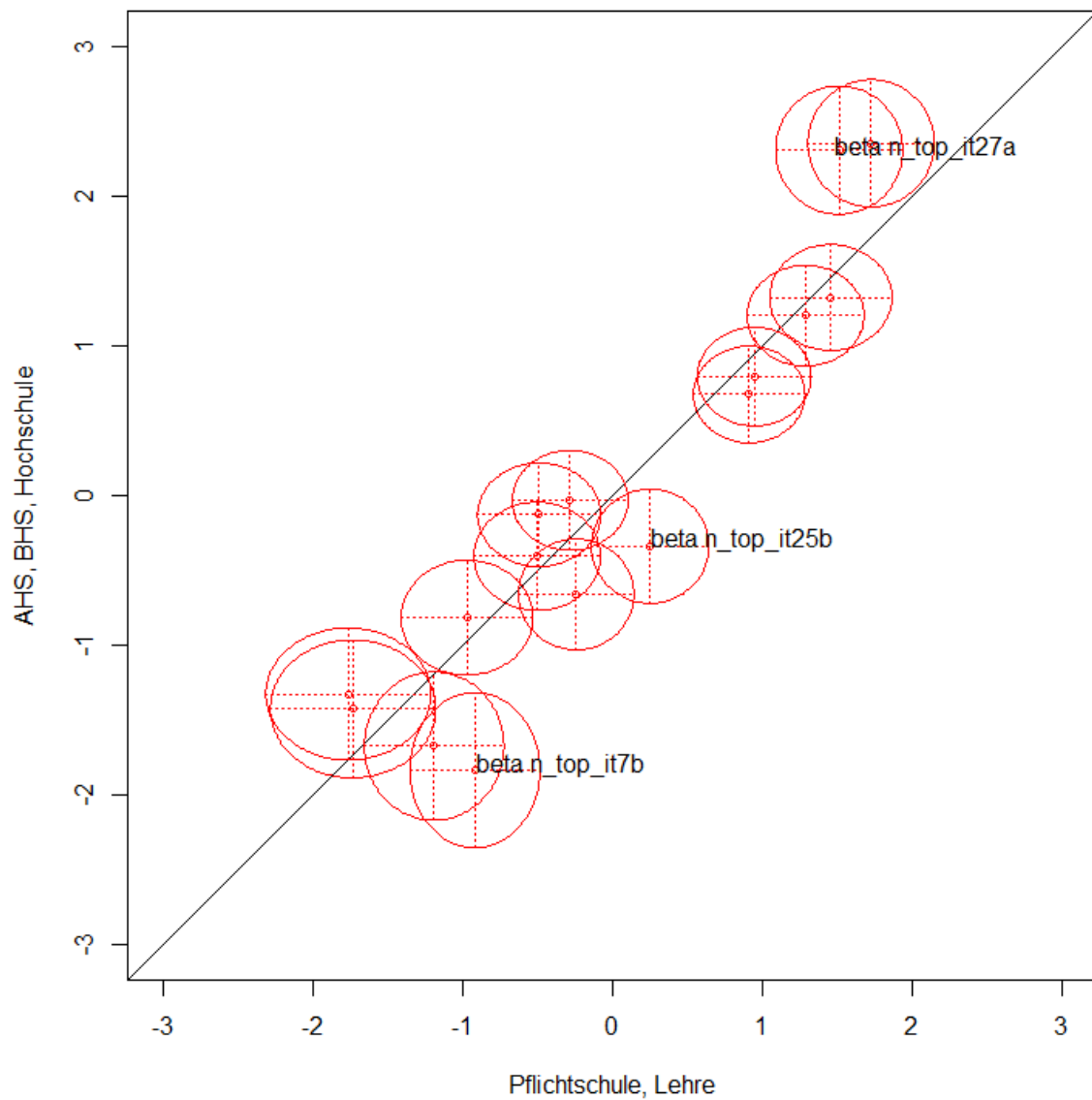


Abbildung 8. Grafische Modellkontrolle mit Konfidenzellipsen für das Teilungskriterium „Bildungsniveau“ mit Beschriftung der auffälligen Items (*N-TOP*, alle Items)

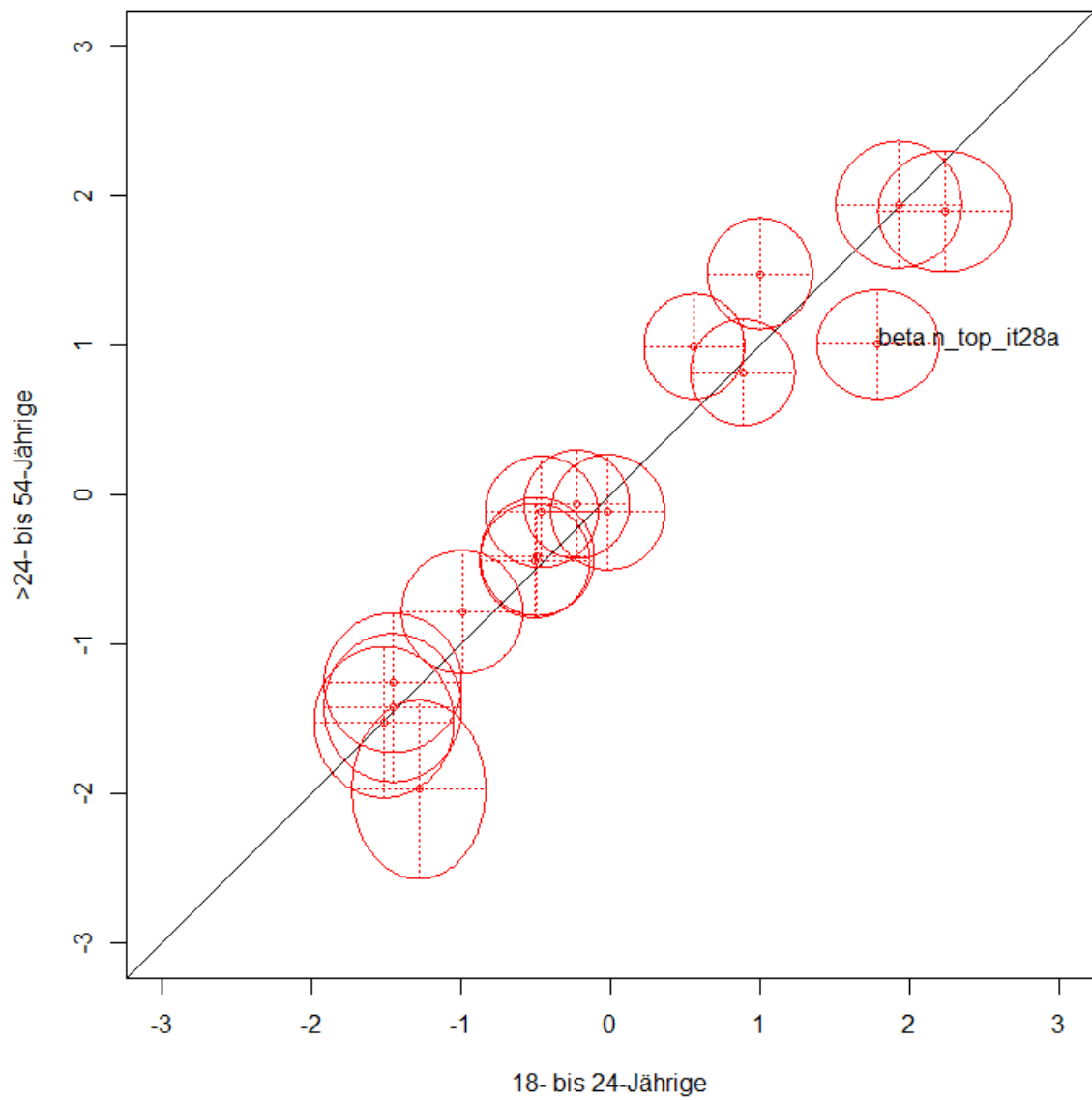


Abbildung 9. Grafische Modellkontrolle mit Konfidenzellipsen für das Teilungskriterium „Alter“ mit Beschriftung der auffälligen Items (*N-TOP*, alle Items)

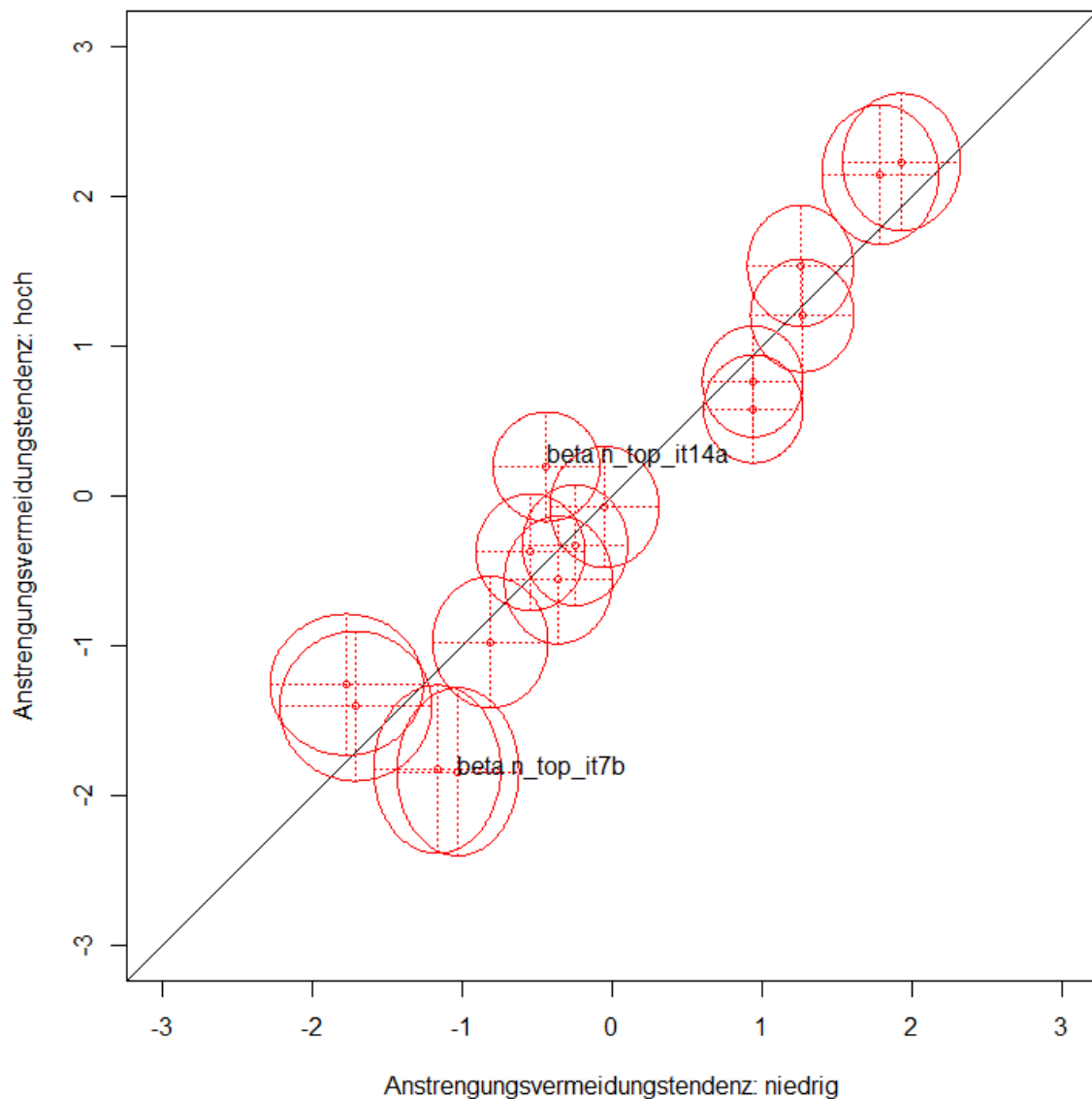


Abbildung 10. Grafische Modellkontrolle mit Konfidenzellipsen für das Teilungskriterium „Anstrengungsvermeidungstendenz“ mit Beschriftung der auffälligen Items (*N-TOP*, alle Items)

Die grafische Modellkontrolle des Teilungskriteriums „Bildungsniveau“ zeigt, dass die Konfidenzellipsen der Items 27a, 25b und 7b die 45°-Gerade nicht schneiden. Im Teilungskriterium „Alter“ ist das Item 28a als auffällig zu werten. In Bezug auf das Teilungskriterium „Anstrengungsvermeidungstendenz“ ist ebenfalls das Item 7b sowie das Item 14a auffällig.

Im Folgenden werden die Ergebnisse derjenigen z-Tests nach Fischer und Scheiblechner in den Tabellen 6 und 7 präsentiert, die signifikante Ergebnisse enthalten.

Tabelle 6

Ergebnisse des z-Tests nach Andersen für das Teilungskriterium „Bildungsniveau“ (N-TOP, alle Items)

Itemnummer	z-Wert	p
7b	-2,654	0,008*
18a	0,520	0,603
12b	-1,538	0,124
16a	-0,946	0,344
14a	0,957	0,339
17b	2,049	0,041
9a	1,212	0,225
11a	-0,333	0,739
21b	0,831	0,406
25b	-2,117	0,034
19b	-1,395	0,163
28a	-0,490	0,624
22a	-0,595	0,552
24b	1,317	0,188
26a	0,341	0,733
27a	2,580	0,010

Anmerkung. () = $p < 0,01$*

Tabelle 7*Ergebnisse des z-Tests nach Andersen für das Teilungskriterium „Alter“ (N-TOP, alle Items)*

Itemnummer	z-Wert	p
7b	0,582	0,561
18a	0,700	0,484
12b	0,237	0,813
16a	1,740	0,082
14a	0,617	0,537
17b	-1,111	0,266
9a	-0,025	0,980
11a	1,842	0,065
21b	-1,819	0,069
25b	-0,367	0,713
19b	0,076	0,940
28a	-2,809	0,005*
22a	-0,263	0,792
24b	1,273	0,203
26a	0,266	0,791
27a	0,045	0,964

Anmerkung. () = p-Wert < 0,01*

Die Ergebnisse der z-Tests bezüglich des Items 7b im Teilungskriterium „Bildungsniveau“ und des Items 28a im Teilungskriterium „Alter“ fallen signifikant aus. Unter Berücksichtigung der Ergebnisse der grafischen Modellkontrollen ist Item 7b mehrfach auffällig und dementsprechend als nicht Rasch-Modell-konform auszuschließen.

Wie aus Tabelle 8 ersichtlich ist, zeigen sich nach Ausschluss des Items 7b keine weiteren signifikanten Ergebnisse in den Likelihood-Ratio-Tests.¹³

Tabelle 8

Ergebnisse der Likelihood-Ratio-Tests (N-TOP, ohne Item 7b)

Teilungskriterium	χ^2 (LRT)	<i>df</i>	χ_{krit}^2	<i>p</i>
Bildungsniveau	24,96	14	29,141	0,035
Alter	20,606	14	29,141	0,112
Anstrengungsvermeidungstendenz	18,088	14	29,141	0,203

Ergänzend werden die grafischen Modellkontrollen für die drei Teilungskriterien in den Abbildungen 11, 12 und 13 nach Ausschluss von Item 7b dargestellt. Auch wenn sich hier noch Items als auffällig präsentieren, kann aufgrund der nicht signifikanten Ergebnisse der Likelihood-Ratio-Tests von Modellgültigkeit ausgegangen werden. Somit kann diesen Items a-posteriori Rasch-Modell-Konformität zugeschrieben werden. Das Verfahren *N-TOP* erhebt entsprechend eindimensional logisch-schlussfolgerndes Denken.

¹³ Bei Interesse können die auf Basis der Ergebnisse dieser Arbeit ausgeschiedenen Items beim Betreuer, Herr Univ.-Prof. Mag. Dr. Kubinger, eingesehen werden.

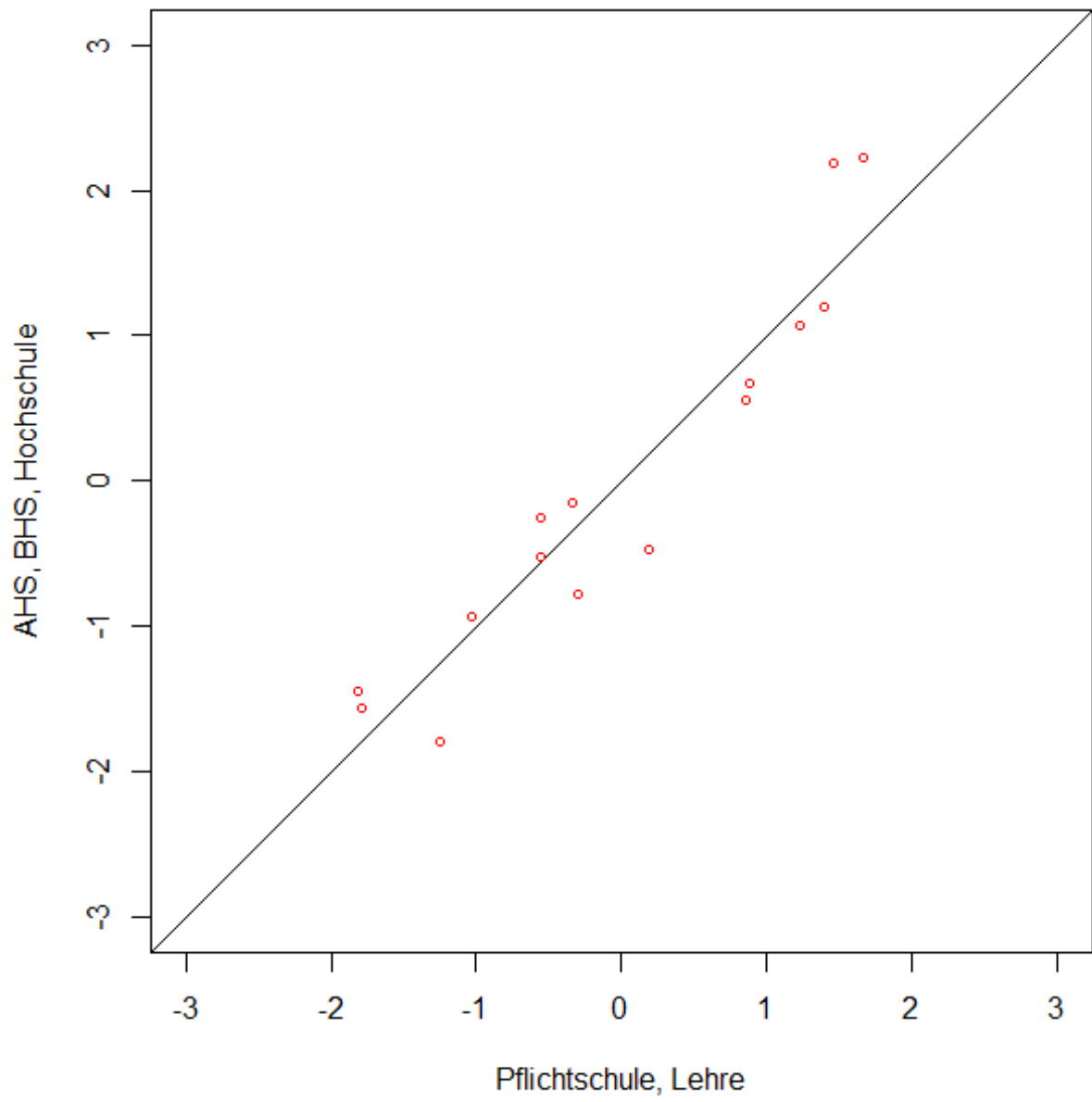


Abbildung 11. Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (*N-TOP*, ohne Item 7b)

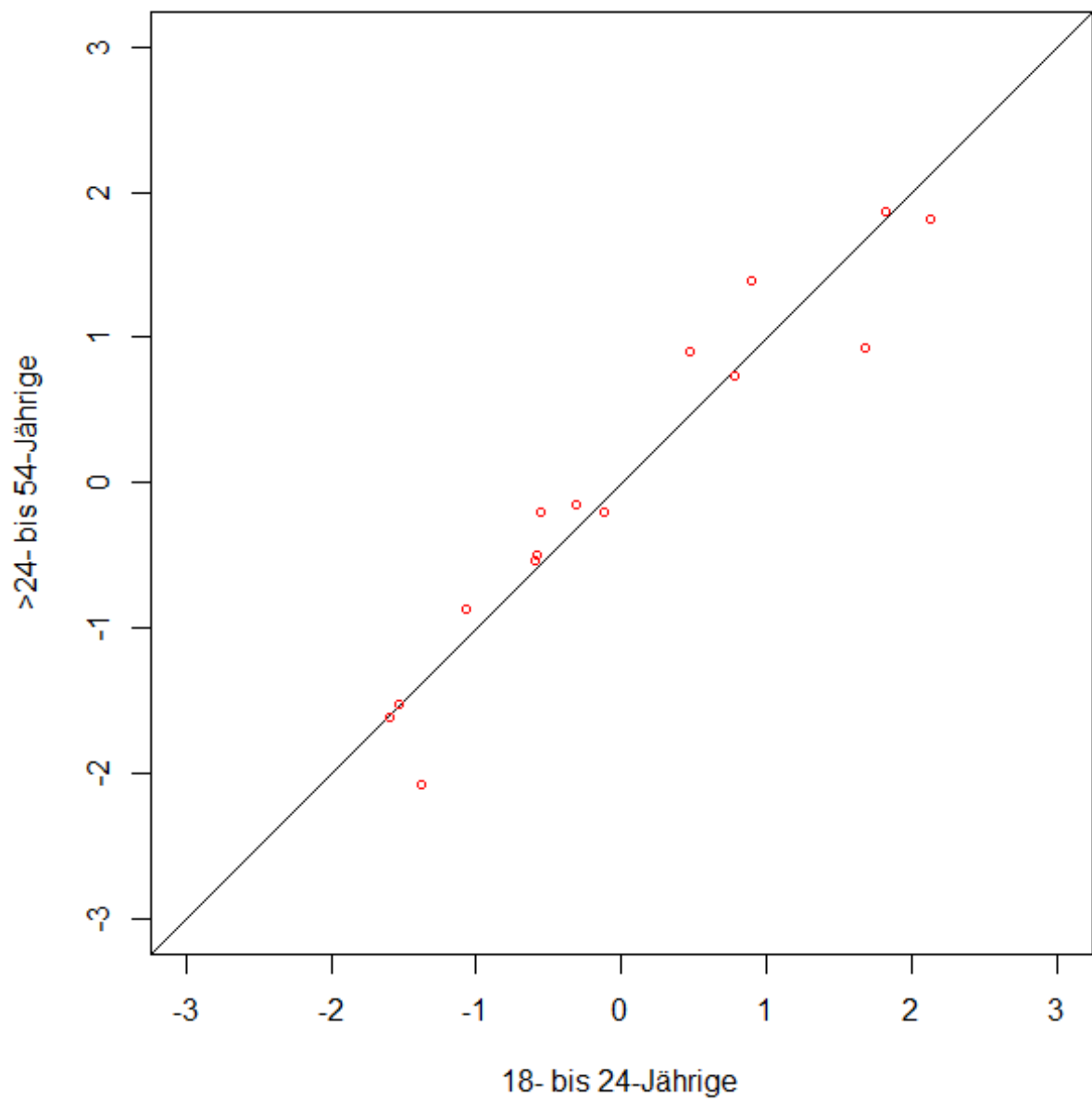


Abbildung 12. Grafische Modellkontrolle für das Teilungskriterium „Alter“ (*N-TOP*, ohne Item 7b)

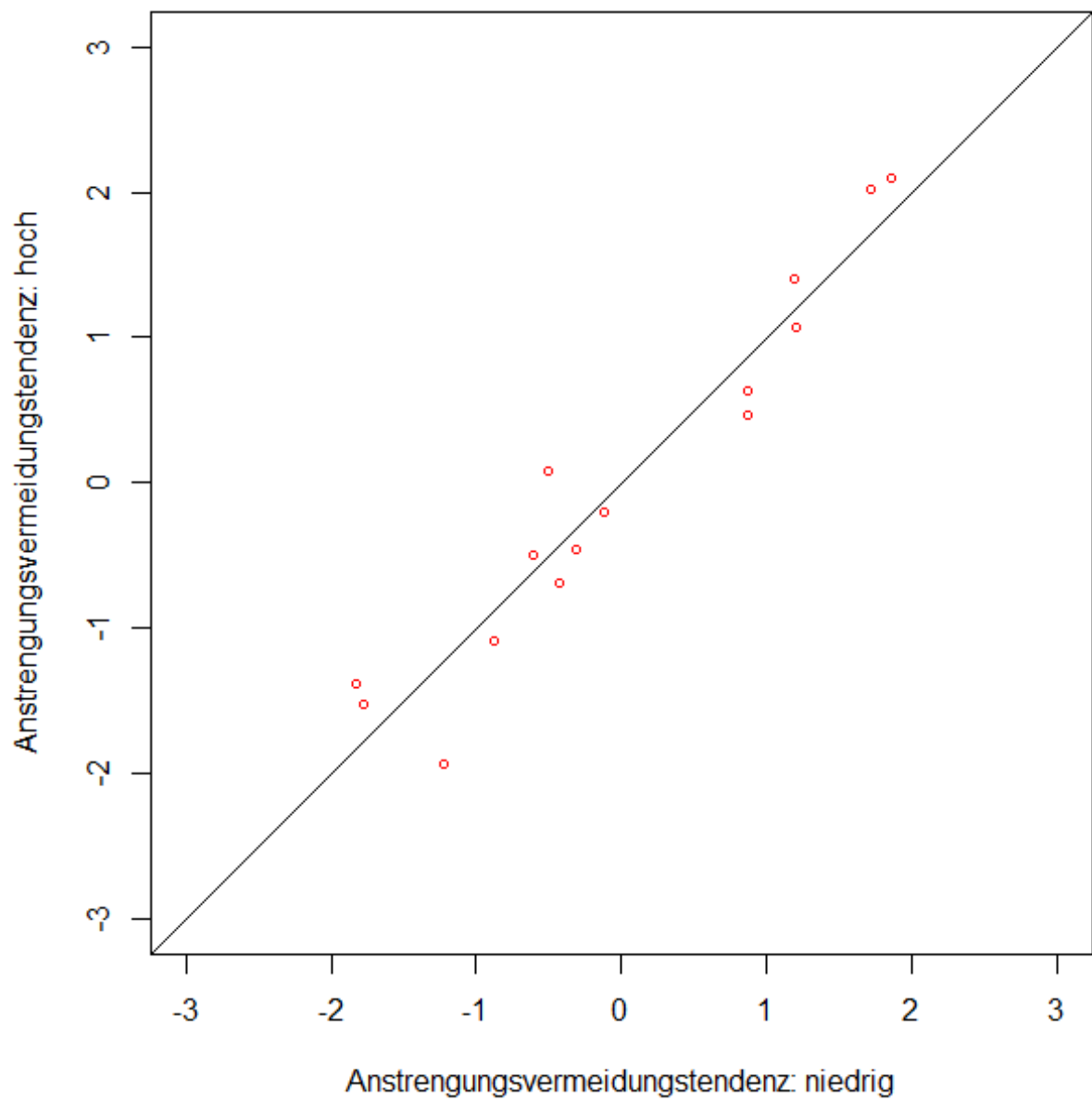


Abbildung 13. Grafische Modellkontrolle für das Teilungskriterium „Anstrengungsvermeidungstendenz“

(*N-TOP*, ohne Item 7b)

8.1.2 Ergebnisse der Rasch-Modell-Analysen beim *FRRT*

Der *FRRT* (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.) hat einen Itempool von insgesamt 29 Items, die alle auf Modellgültigkeit überprüft wurden ($n = 292$).

Das Item 64 musste von der Analyse der Teilungskriterien „Bildungsniveau“ und „Alter“ ausgeschlossen werden, weil es in den jeweiligen Teilstichproben von keiner Testperson gelöst werden konnte.

Tabelle 9 zeigt, dass alle Likelihood-Ratio-Tests nach Andersen in den drei Teilungskriterien nicht signifikant ausfallen. Dementsprechend werden keine Items ausgeschlossen. Die Abbildungen 14, 15 und 16 veranschaulichen die zugehörigen grafischen Modellkontrollen ohne Konfidenzellipsen, da in diesem Fall keine Items entfernt werden müssen. Diesen Ergebnissen entsprechend kann den 28 Items (ohne Item 64) a-posteriori Modellgültigkeit zugeschrieben werden. Folglich misst das Verfahren *FRRT* eindimensional logisch-schlussfolgerndes Denken.

Tabelle 9

Ergebnisse der Likelihood-Ratio-Tests (FRRT, alle Items)

Teilungskriterium	χ^2 (LRT)	<i>df</i>	χ_{krit}^2	<i>p</i>
Bildungsniveau	29,408	27	46,963	0,341
Alter	35,339	27	46,963	0,130
Anstrengungsvermeidungstendenz	36,618	28	48,278	0,128

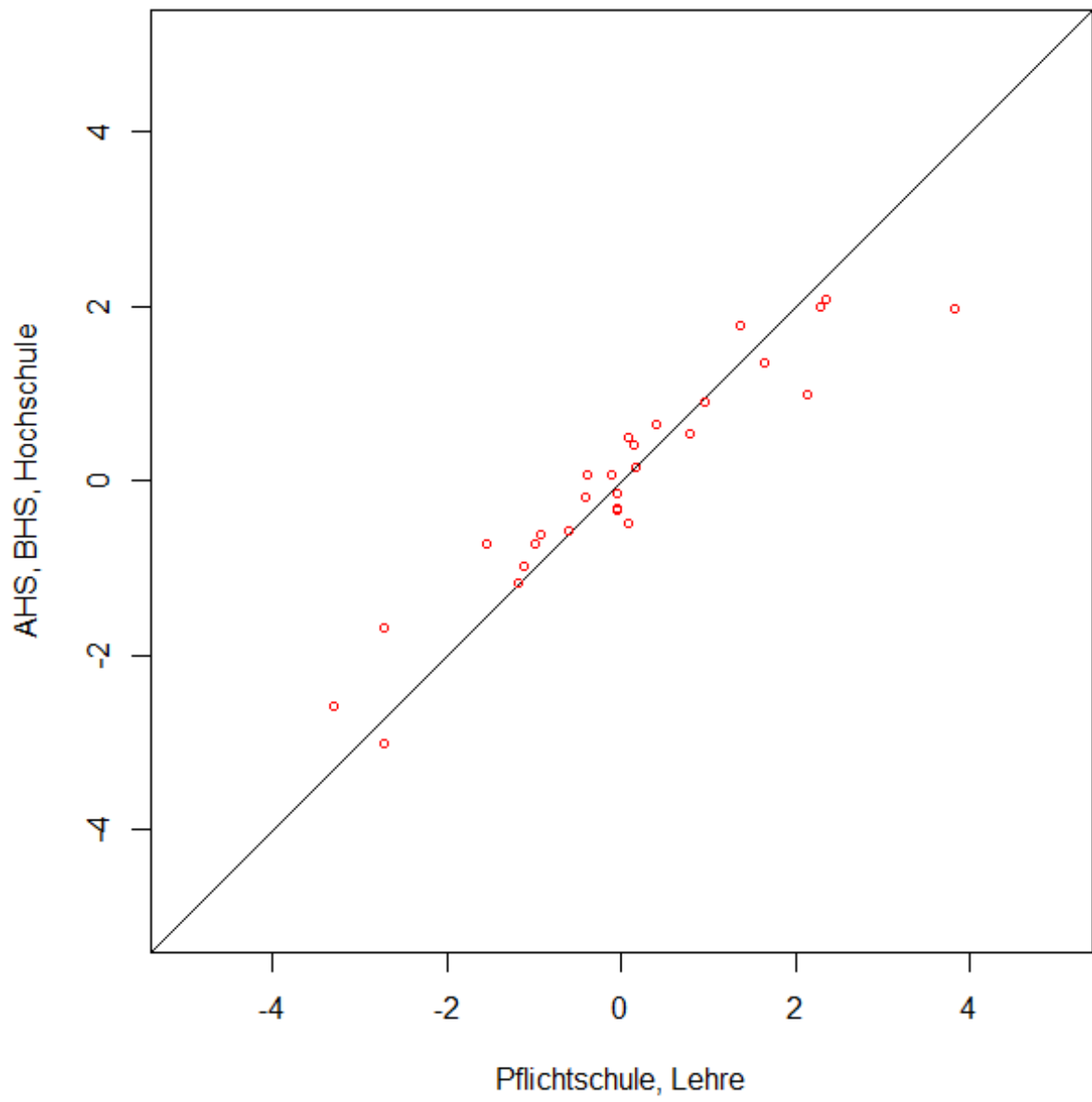


Abbildung 14. Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (*FRRT*, alle Items)

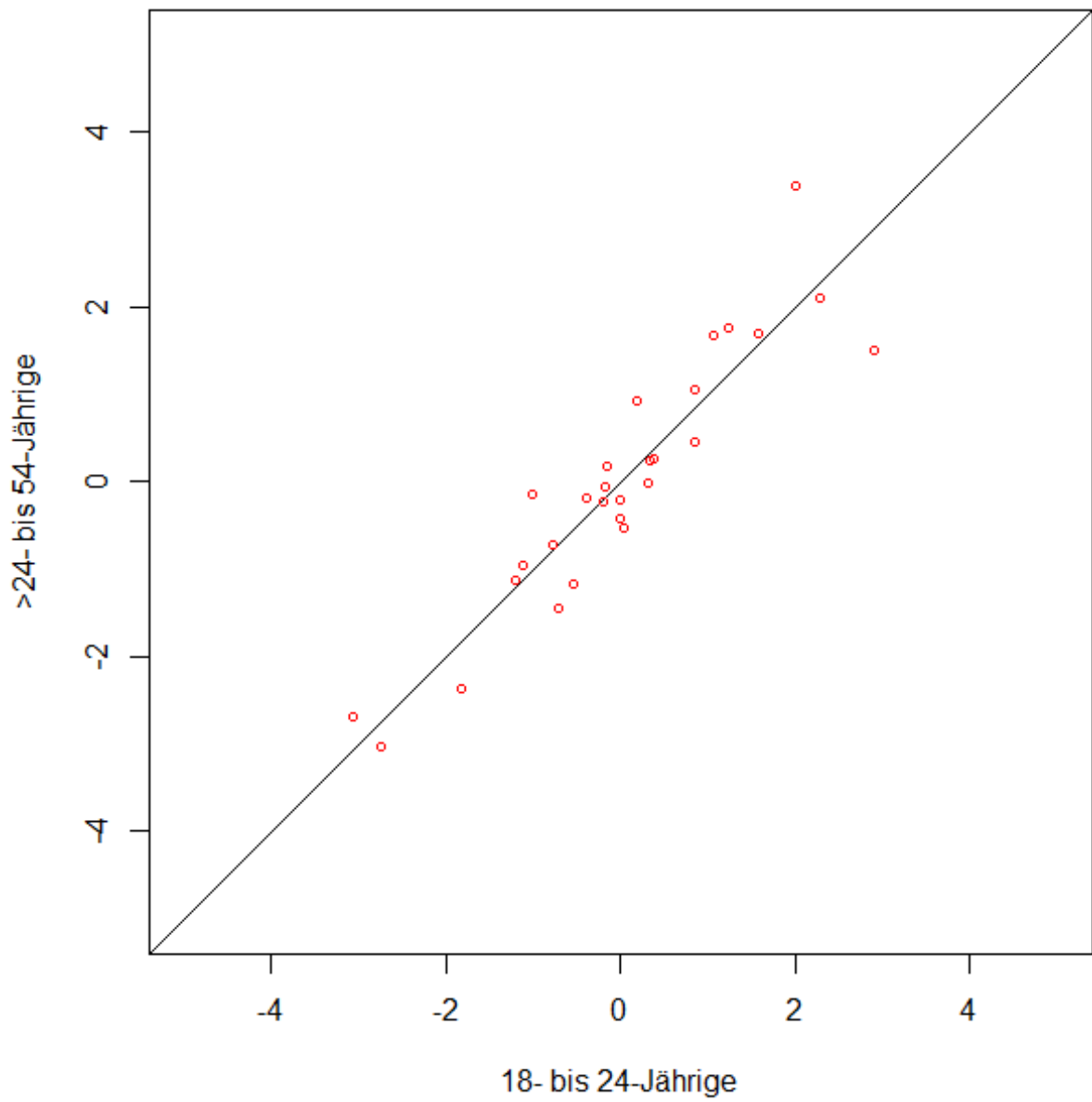


Abbildung 15. Grafische Modellkontrolle für das Teilungskriterium „Alter“ (FRRT, alle Items)

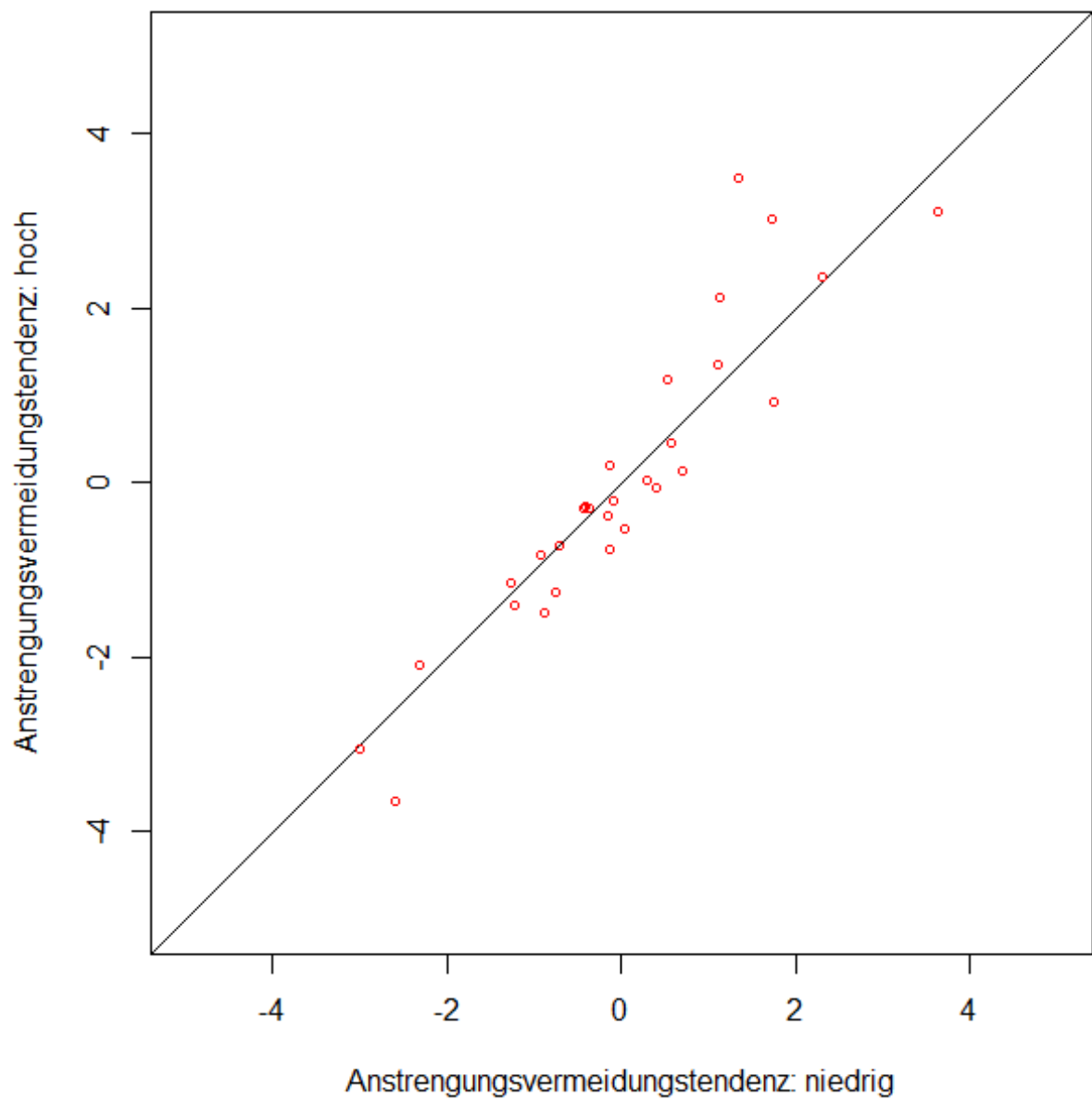


Abbildung 16. Grafische Modellkontrolle für das Teilungskriterium „Anstrengungsvermeidungstendenz“ (*FRRT*, alle Items)

Ergänzend werden im Folgenden die Ergebnisse der z -Tests nach Fischer und Scheiblechner für das Teilungskriterium Anstrengungsvermeidungstendenz für den *FRRT* in der Tabelle 10 präsentiert. Bezüglich der anderen Teilungskriterien zeigten sich keine Auffälligkeiten.

Tabelle 10

Ergebnisse des z-Tests nach Andersen für das Teilungskriterium „Anstrengungsvermeidungstendenz“ (FRRT, alle Items)

Itemnummer	z-Wert	p
1	-0,084	0,933
3	0,541	0,589
4	0,302	0,763
5	0,385	0,700
8	-1,739	0,082
10	-1,648	0,099
11	-0,503	0,615
15	0,231	0,817
16	-1,407	0,159
17	-0,345	0,730
18	0,409	0,682
19	-1,853	0,064
23	-1,621	0,105
27	1,174	0,240
29	-0,647	0,518
31	0,425	0,674
32	-0,042	0,966
33	-1,256	0,209
38	1,371	0,170
41	-0,305	0,760
42	-1,100	0,271

43	0,488	0,626
46	1,202	0,229
48	-0,561	0,575
53	1,534	0,125
56	2,626	0,009*
57	-1,749	0,080
61	0,084	0,933
64	-0,360	0,719

Anmerkung. (*) = p -Wert < 0,01

8.2 Korrelative Zusammenhänge zwischen den Reasoning-Tests

Die untersuchten Verfahren *N-TOP*, *FRRT* und *WMT 2* erheben den Anspruch unterschiedliche Konstrukte des Faktors Reasoning zu erheben. Daher wurden auf Basis der Personenparameter die linearen Zusammenhänge zwischen den Verfahren und somit die diskriminante Validität untersucht. Bekanntlich haben die Fähigkeitsparameter des Rasch-Modells (mindestens) Intervallskalenniveau (Kubinger, 1989). Daher kann die Produkt-Moment-Korrelation nach Pearson verwendet werden.

Bei der Analyse der korrelativen Zusammenhängen wurde entsprechend der Nullhypothese (3) davon ausgegangen, dass zwischen den Reasoning-Tests, die denselben Intelligenzbereich (fluid vs. kristallisiert) erheben, wie der *N-TOP* und der *WMT 2*, der Zusammenhang größer sein muss als ein a priori festgelegter Grenzwert; während im Vergleich dazu bei Verfahren, welche unterschiedliche Konstrukte erfassen, der Zusammenhang nur maximal so groß wie der Grenzwert ist. Diese theoretischen

Überlegungen orientieren sich an Studien von Cattell (1987). Im Zuge von Untersuchungen und Vergleichen unterschiedlicher Intelligenzkonzepte und -tests stellte Cattell fest, dass die Faktoren fluide und kristallisierten Intelligenz einen positiven Zusammenhang von 0,4 bis 0,5 aufweisen (Cattell, 1987, S. 116). Cattell vertritt dabei die Annahme, dass die Entwicklung der kristallisierten Intelligenz zu einem gewissen Teil von der fluiden abhängig ist und sich somit der Zusammenhang erklären lässt. In Anlehnung an diese Ergebnisse sowie eine der vorliegenden Studie vorhergehende Diplomarbeit (siehe dazu Ünal, 2014) wird hier ein Grenzwert von $r = 0,4$ festgelegt.

In Tabelle 11 wird die Interkorrelationsmatrix schematisch dargestellt. Tabelle 17 zeigt eine Matrix der Bestimmtheitsmaße.

Tabelle 11

Interkorrelationsmatrix der Reasoning-Tests

	<i>N-TOP</i>	<i>FRRT</i>	<i>WMT 2</i>
<i>N-TOP</i>	1,000	0,379	0,357
<i>FRRT</i>	0,379	1,000	0,278
<i>WMT 2</i>	0,357	0,278	1,000

Tabelle 12

Matrix der Bestimmtheitsmaße der Reasoning-Tests

	<i>N-TOP</i>	<i>FRRT</i>	<i>WMT 2</i>
<i>N-TOP</i>	1,000	0,143	0,128
<i>FRRT</i>	0,143	1,000	0,077
<i>WMT 2</i>	0,128	0,077	1,000

Die Korrelationen zwischen den verschiedenen Reasoning-Tests weisen alle positive Zusammenhänge auf. Diese befinden sich in einem Bereich von $r = 0,278$ bis $r = 0,379$.

Der Vergleich zwischen *N-TOP* und *FRRT* ergibt die höchste Korrelation. Das Bestimmtheitsmaß mit $B = 0,143$ zeigt, dass ca. 14% der Varianz der Personenparameter des *N-TOP* durch den *FRRT* erklärt werden. Dies gilt ebenfalls für den umgekehrten Fall. Entsprechend der Hypothese (3) überschreitet die Korrelation den festgelegten Grenzwert von $r = 0,4$ nicht. Die Verfahren *N-TOP* und *FRRT* erheben also Fähigkeiten, die sich wesentlich voneinander unterscheiden.

Auch die berechnete Korrelation zwischen *WMT 2* und *FRRT* fällt mit $r = 0,278$ kleiner als der festgelegte Grenzwert aus. Beide Verfahren erheben somit unterschiedliche Konstrukte.

Der positive Zusammenhang zwischen den Tests *N-TOP* und *WMT 2*, die beide auf dem Konstrukt der fluiden Intelligenz basieren, fällt mit $r = 0,357$ entgegen der Hypothese (3) kleiner als der festgelegte Grenzwert aus. Das Bestimmtheitsmaß mit $B = 0,128$ gibt an, dass lediglich 12% gemeinsam erklärter Varianz besteht.

IV. DISKUSSION UND AUSBLICK

Ziel der vorliegenden Diplomarbeit war eine Analyse und ein Vergleich der Reasoning-Tests *Numerische Topologien*, *Family-Relation-Reasoning-Test* und *Wiener Matrizen-Test 2*. Der *N-TOP* und der *FRRT*, die sich beide noch in Entwicklung befinden, wurden zunächst auf Rasch-Modell-Konformität hin überprüft. Anschließend erfolgte ein Vergleich der zugrunde liegenden Konstrukte aller Verfahren.

Konkret wurden die Items des *N-TOP* sowie des *FRRT* a-posteriori dahingehend überprüft, ob sie mit dem dichotomen logistischen Modell von Rasch konform sind. Gleichzeitig wurde festgestellt, ob die Tests die Fähigkeit zum schlussfolgernden Denken tatsächlich eindimensional erheben.

Bei den diesbezüglichen Analysen zeigte sich, dass beim *N-TOP* ein Item (entspricht ca. 6.25% aller Items) ausgeschlossen werden musste, um a-posteriori Modellgültigkeit zu erreichen. Das entsprechende Item konnten 247 von insgesamt 292 Testpersonen lösen. Lediglich 45 Personen gelang dies nicht.

Ein inhaltlicher Grund für die Auffälligkeit dieses Items könnte die Formatierung des Aufgabenmaterials sein, da diese teilweise uneinheitlich ausfällt. Konkret variieren die Abstände zwischen den vorgegebenen Ziffern und dem freien Kästchen, in welches die Lösung eingetragen werden soll. Hierdurch könnte bei zweistelligen Lösungen potentiell eine Unklarheit dahingehend entstehen, ob das freie Kästchen isoliert zu betrachten ist – also eine aus mehreren Ziffern bestehende Lösung zur Gänze in das Kästchen eingetragen werden muss – oder ob die vorangestellte Zahl bereits als Teil der Lösung zu betrachten ist. Im Fall des ausgeschlossenen Items besteht die Lösung aus einer Zahl mit zwei Ziffern, die jedoch nicht beide in das leere Kästchen einzutragen sind. Die erste Ziffer der Lösung steht bereits vor dem leeren Kästchen. Folglich sollte lediglich die zweite Ziffer der Lösungszahl eingetragen

werden. Von besagten 45 Testpersonen, die das Item nicht wie vorgesehen lösen konnten, trugen 28 die Lösung, bestehend aus zwei Ziffern, zur Gänze in das freie Kästchen ein. Es liegt also die Vermutung nahe, dass trotz der hohen Anzahl an Personen, die das Item lösen konnten, die beschriebene Auffälligkeit aus Formatierungsgründen besteht.

Bei der Analyse bezüglich des *FRRT* mussten keine Items ausgeschlossen werden. Es bleibt jedoch festzuhalten, dass die Bearbeitungszeit von 15 Minuten des *FRRT* sehr knapp bemessen scheint. Betrachtet man die Items, die nach dem Strich folgen, den die Testpersonen unter das von ihnen zuletzt bearbeitete Item nach Ablauf der Bearbeitungszeit ziehen sollten, zeigt sich, dass ungefähr ab dem neunten Item, unabhängig von der Testform A, B oder C, ca. 200 Testpersonen nicht mehr zu deren Bearbeitung gelangt sind.

Das Item 64 wurde beispielsweise lediglich von 50 Personen bearbeitet, wobei nur von zwei von ihnen die korrekte Antwort gegeben wurde. Dieses Item mit dem höchsten Schwierigkeitsparameter konnte nicht in die Überprüfung mit einbezogen werden, da jene zwei Personen zudem keine Angaben zu den, für die Modell-Analyse obligatorischen, Teilungskriterien Alter sowie Bildungsniveau gemacht hatten.

Um alle Items genauer untersuchen zu können, empfiehlt sich in zukünftigen Studien die Bearbeitungszeit des *FRRT* zu erhöhen.

Die Untersuchung bezüglich der korrelativen Zusammenhänge zwischen den Verfahren diente der Bestimmung der diskriminanten Validität und gründete auf der Hypothese, dass der Zusammenhang zwischen den Reasoning-Tests, die denselben Intelligenzbereich (fluid vs. kristallisiert) erheben, größer ist als ein anhand von theoretischen Überlegungen festgelegter Grenzwert. Demgegenüber weisen Reasoning-Tests, die unterschiedliche Konstrukte erfassen einen maximal dem Grenzwert entsprechenden Zusammenhang auf. Diese Annahmen ließen sich teilweise bestätigen.

Entsprechend der Hypothese wiesen der *N-TOP* und der *FRRT* sowie der *WMT 2* und der *FRRT* jeweils einen Korrelationskoeffizienten kleiner dem festgelegten Grenzwert auf. Daraus kann geschlussfolgert werden, dass der *N-TOP* und der *FRRT*, bzw. der *WMT 2* und der *FRRT* auf unterschiedlichen Konstrukten basieren.

Entgegen der Hypothese fiel der Zusammenhang der Verfahren *N-TOP* und *WMT 2*, die beide nach dem Konzept der fluiden Intelligenz sensu Cattell (1957; 1963) konstruiert wurden, nicht größer als der festgelegte Grenzwert aus. Mit einem Korrelationskoeffizienten von $r = 0,357$ besteht lediglich ein relativ geringer positiver Zusammenhang in Bezug auf die Hypothese.

Ein Grund für den geringen Korrelationskoeffizienten könnte die Verwendung unterschiedlicher Materialformen sein (numerisch, figural). Innerhalb der Untersuchungssituation ließ sich bei der Instruktion des *N-TOP* häufig eine gewisse Ablehnung des Testmaterials beobachten, was sich beispielsweise in lautem Seufzen, Stöhnen und allgemeiner Unruhe der Gruppe äußerte. Dies könnte dahingehend interpretiert werden, dass Mathematik allgemein als polarisierendes Fachgebiet aufgefasst wird. Die subjektiv wahrgenommene mathematische Kompetenz der Testpersonen könnte sich bei numerischem Testmaterial auf die Motivation zur gewissenhaften Testbearbeitung auswirken.

Da das verwendete Testmaterial überdies gewisse mathematische Grundkompetenzen voraussetzt, kann ein Einfluss von kristallisiertem Wissen auf die Testergebnisse nicht ausgeschlossen werden. Ob das Verfahren *Numerische Topologien* eindeutig einem der beiden Intelligenzbereiche sensu Cattell (1957; 1963) zuzuordnen ist, könnte dementsprechend Gegenstand zukünftiger Untersuchungen sein.

Um Material-abhängige Fähigkeiten ausschließen zu können, empfiehlt es sich in weiterführenden Studien die Materialform (numerisch, lexikalisch, figural) als Faktor zu berücksichtigen. Insbesondere sollten Tests aus demselben Intelligenzbereich, die aber unterschiedliche Materialformen verwenden, miteinander verglichen werden und umgekehrt

Tests aus unterschiedlichen Intelligenzbereichen unter Verwendung ähnlicher Testmaterialien.

Aufgrund von Äußerungen der Teilnehmerinnen und Teilnehmer nach den Testungen, kann dem Verfahren *Family-Relation-Reasoning-Test* keine Augenscheinvalidität zugesprochen werden. Dies bedeutet, dass es den Testpersonen „augenscheinlich“ nicht klar war, was mit diesem Test erfasst werden sollte (siehe dazu beispielsweise Kubinger, 2009). In der Instruktion des FRRT wird nicht explizit darauf hingewiesen, dass es sich um einen Test zur Messung des logisch schlussfolgernden Denkens mit Hilfe von lexikalischem Material handelt. Die Teilnehmer/innen empfanden diesen Tests rückblickend als am anstrengendsten.

Wie erwähnt rief bereits die Vorgabe des Tests *Numerische Topologien* aufgrund seines numerischen Materials anfänglich bei allen Gruppen Ablehnung hervor. Wie sich jedoch herausstellte, bearbeiteten alle Gruppen, in Relation der zur Verfügung stehenden Bearbeitungszeit, diesen Test im Durchschnitt am schnellsten.

Der *Wiener Matrizen-Test 2* beziehungsweise ähnlich konzipierte Verfahren, die schlussfolgerndes Denken mit Hilfe von figuralem Material erheben, waren den Testpersonen bereits aus den von ihnen durchlaufenen Auswahlverfahren bekannt.

Obgleich sich die Einteilung der untersuchten Testverfahren in a-priori angenommene Intelligenzbereiche nicht vollständig bestätigen ließ, zeigten die korrelativen Zusammenhänge, dass zwar allen Verfahren ein gemeinsames Konzept zugrunde liegt – das des logisch-schlussfolgernden Denkens – sie jedoch jeweils unterschiedliche Teilaspekte hiervon zu erfassen scheinen. Dementsprechend bietet jedes dieser Verfahren die Möglichkeit eines spezifischen Informationsgewinns bezüglich der Fähigkeit zum schlussfolgernden Denken.

V. ZUSAMMENFASSUNG

Im Rahmen dieser Diplomarbeit wurden zwei sich in Entwicklung befindliche Reasoning-Tests auf Rasch-Modell-Konformität überprüft. Konkret wurde die Dimensionalität der psychologisch-diagnostischen Verfahren *Numerische Topologien* (kurz *N-TOP*) (Kubinger & Heuberger, in Vorb.) und *Family-Relation-Reasoning-Test* (kurz *FRRT*) (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.) untersucht. Der *N-TOP* basiert auf dem theoretischen Konzept der fluiden Intelligenz sensu Cattell (z. B. 1957; 1963) und erhebt schlussfolgerndes Denken mit Hilfe numerischen Testmaterials. Der *FRRT* wird dem kristallisierten Intelligenzbereich nach Cattell (1957; 1963) zugeordnet und erhebt die Fähigkeit zum schlussfolgernden Denken mittels lexikalischem Aufgabenmaterial. Überdies wurden korrelative Zusammenhänge der beiden Verfahren untereinander sowie mit dem im deutschsprachigen Raum bereits etablierten *Wiener Matrizen-Test 2* (Formann, Waldherr, & Piswanger, 2011), der Reasoning im figuralen Bereich erfasst, analysiert.

Die Überprüfung auf Rasch-Modell-Konformität der Items des *N-TOP* und des *FRRT* und die Untersuchung der korrelativen Zusammenhänge und damit einhergehend der diskriminanten Validität erfolgte an einer Stichprobe bestehend aus 292 Schülerinnen und Schülern der Sicherheitsakademie Traiskirchen sowie der Krankenpflegeschulen Baumgartner Höhe und Ybbs. Die Testungen fanden im Rahmen von Gruppentestungen in den Schulen statt. Neben den drei Reasoning-Tests bestand die Testbatterie aus einem Datenblatt zur Erhebung der demographischen Daten sowie dem Fragebogen *Anstrengungsvermeidungstest* (kurz *AVT-Sch*) (Rollett & Rollett, 2000, zitiert nach Gashi, 2011, S. 41), der als externes Teilungskriterium innerhalb der Rasch-Modell-Analysen eingesetzt wurde.

Die Überprüfung der Items auf Rasch-Modell-Konformität erfolgte mit dem Likelihood-Ratio-Test nach Andersen (1973), graphischen Modellkontrollen und dem *z*-Test

nach Fischer und Scheiblechner (1970, zitiert nach Kubinger et al., 2011, S. 556-557). Als externe Teilungskriterien dienten das Bildungsniveau (Pflichtschule, Lehre vs. AHS, BHS, Hochschule), das Alter (18- bis 24-Jährige vs. > 24- bis 54-Jährige) sowie der Testwert des *AVT-Sch* (niedrig (Testwert $\leq Md_{AVT-Sch} = 7$) vs. hoch (Testwert $> Md_{AVT-Sch} = 7$)). Um Modellgültigkeit zu erreichen musste ein auffälliges Item (Item 7b) aus dem *N-TOP* ausgeschlossen werden. Bei den Rasch-Modell-Analysen den *FRRT* betreffend konnte ohne Ausschluss von Items Modellgeltung festgestellt werden. Anzumerken ist, dass das Item 64 aufgrund fehlender Daten nicht in die Berechnungen miteinbezogen werden konnte. Nach Ausschluss der auffälligen Items kann angenommen werden, dass beide Verfahren logisch-schlussfolgerndes Denken eindimensional erheben.

Die korrelativen Zusammenhänge zwischen den drei Reasoning-Tests wurden mit Hilfe der Produkt-Moment-Korrelation nach Pearson berechnet. Entsprechend der Annahme zeigte sich ein positiver Zusammenhang kleiner dem anhand von theoretischen Überlegungen festgelegten Grenzwert von $r = 0,4$ zwischen *FRRT* und *N-TOP* ($r = 0,379$) sowie *FRRT* und *WMT 2* ($r = 0,278$). Diese Verfahren beruhen jeweils auf unterschiedlichen Intelligenzkonzepten (fluid vs. kristallisiert). Entgegen der Hypothese zeigte sich bei den Tests, die demselben Intelligenzbereich zugeordnet werden, kein Korrelationskoeffizient größer dem festgelegten Grenzwert. Es konnte lediglich ein Zusammenhang von $r = 0,357$ festgestellt werden. Die insgesamt eher niedrigen Interkorrelationen zwischen den Personenparametern der Verfahren weisen darauf hin, dass der *N-TOP*, der *FRRT* und der *WMT 2* jeweils unterschiedliche Aspekte der Fähigkeit Reasoning erfassen.

Vor diesem Hintergrund erscheint es sinnvoll für zukünftige Untersuchungen die Geltung des Rasch-Modells anhand neuer und umfangreicherer Daten zu erheben sowie eine größere Anzahl an Reasoning-Tests miteinander zu vergleichen.

VI. LITERATURVERZEICHNIS

- Adler, J. (2012). *R in a nutshell. A Desktop Quick Reference*. 2. Aufl. Sebastopol: O'Reilly.
- Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, 38 (1), 123-140.
- Beckmann, J. F., & Guthke, J. (1999). *Psychodiagnostik des schlußfolgernden Denkens: Handbuch zur Adaptiven Computergestützten Intelligenz-Lerntestbatterie für Schlußfolgerndes Denken (ACIL)*. Göttingen: Hogrefe.
- Brentano, F. (1956). *Die Lehre vom richtigen Urteil: nach den Vorlesungen über Logik mit Benützung anderer Manuskripte aus dem Gebiete der Erkenntnistheorie*. Bern: Francke.
- 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 Science Publishers .
- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. New York, Cambridge University Press.
- Dimitz, E., & Formann, A. (1979). *Programm zur Schaetzung der Item- und Personenparameter im dichotomen logistischen Modell von RASCH sowie Modellkontrollen und ein interaktives System zur Eingabe und Aenderung der Programmparameter, Steuerbefehle und Daten*. Institutsarbeit, Institut für Höhere Studien, Wien.

- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen*. Bern: Hans Huber.
- Funke, J., & Vaterrodt-Plünnecke, B. (1998). *Was ist Intelligenz?*. München: Beck.
- Gamsjäger, C. (2012). *Entwicklung des Tests „Gleichungen“ zur Erfassung von Reasoning*.
Unveröffentlichte Diplomarbeit, Universität Wien, Wien.
- Gashi, A. (2011). *Untersuchungen über Beziehungen zwischen Anstrengungsvermeidung, Selbststeuerung und Lernemotionen bei Jugendlichen aus psychologischer Sicht*.
Unveröffentlichte Diplomarbeit, Universität Wien, Wien.
- Guilford, J. P. (1971). *The nature of human intelligence*. (International student edition). New York: McGraw-Hill.
- Guilford, J. P., & Hoepfner, R. (1976). *Analyse der Intelligenz*. Weinheim: Beltz.
- Hofstätter, P. R. (1977). *Persönlichkeitsforschung*. Stuttgart: Alfred Kröner.
- Holling, H., Preckel, F., & Vock, M. (2004). *Intelligenzdiagnostik*. Göttingen: Hogrefe.
- Horn, J. L. (1968). Organization of abilities and the development of intelligence.
Psychological Review, 75 (3), 242-259.
- 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. (1989). *Moderne Testtheorie. Ein Abriß samt neuesten Beiträgen*. 2. verb. Aufl. Weinheim: Beltz.
- Kubinger, K. D. (2003a). Gütekriterien. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 195-204). Weinheim: Beltz.
- Kubinger, K. D. (2003b). Probabilistische Testtheorie. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415-423). Weinheim: Beltz.

- Kubinger, K. D. (2005). Psychological test calibration using the rasch model - Some critical suggestions on traditional approaches. *International Journal of Testing*, 5 (4), 377-394.
doi: 10.1207/s15327574ijt0504_3
- Kubinger, K. D. (2009). Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens. 2. überarb. und erweiterte Aufl. Göttingen: Hogrefe.
- Kubinger, K. D., & Heuberger, N. (in Vorb.). *Numerische Topologien (N-TOP)*.
- Kubinger, K. D., Poinstingl, H., Weidinger, G., & Skoda, S. (in Vorb.). *Family-Relation-Reasoning-Test (FRRT)*.
- Kubinger, K. D., Rasch, D., & Yanagida, T. (2011). *Statistik in der Psychologie: Vom Einführungskurs bis zur Dissertation*. Göttingen: Hogrefe.
- Kutschera, F. v., & Breitkopf, A. (2000). *Einführung in die moderne Logik*. 7. überarb. Aufl. Freiburg: Alber.
- Lohman, D. F. (2005). Reasoning Abilities. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition & Intelligence. Identifying the Mechanisms of the Mind* (pp. 225-250). Cambridge, USA: Cambridge Press.
- 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.
- Molenaar, I. W. (1995). Some Background for Item Response Theory and the Rasch Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models – Foundations, recent developements, and applications* (pp. 3-14). New York: Springer.
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. 2. Aufl. Berlin: Springer.
- Putz-Osterloh, W. (1981). *Problemlöseprozesse und Intelligenzleistung*. Stuttgart: Huber.

- Poinstingl, H., Mair, P., & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modeling) Anwendung des Rasch-Modells (1-PL-Modell)*. Lengerich: Pabst Science Publishers.
- Rollett, B., & Bartram, M. (1998). *Anstrengungsvermeidungstest (AVT). Manual*. 3. überarb. Aufl. Göttingen: Hogrefe.
- Schüling, H. (2002). *Die Genesis des schlußfolgernden Denkens*. Hildesheim: Georg Olms.
- Skoda, S. (2005). *Der Verwandtschaften-Reasoning-Test*. Unveröffentlichte Diplomarbeit, Universität Wien, 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.
- 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.
- 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. (1961). The theory of multiple factors. In J. J. Jenkins & D. G. Paterson (Eds.), *Studies in individual differences: The search for intelligence* (pp. 351-367). London: Methuen & Co.
- Ünal, S. (2014). *Figurale Analogien: Ein Test zur Messung des kristallisierten schlussfolgernden Denkens mit Hilfe figuralen Materials*. Unveröffentlichte Diplomarbeit, Universität Wien, Wien.
- Formann, A. K., Waldherr, K., & Piswanger, K. (2011). *Wiener Matrizen-Test 2. Manual*. Göttingen: Beltz Test.

- Waldherr, K. (2001). *Differential Item Functioning-Analysen mithilfe der Familie der Rasch-Modelle*. Diss., Universität Wien, Wien.
- 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.

VII. ANHANG

A. Itemanordnung

In der ersten Zeile der Tabelle sind jeweils die original Itemnummerierungen angeführt, wie sie in den Verfahren verwendet sind. In den darunterliegenden Zeilen sind die Nummerierungen entsprechend der Testformen A und B für den *N-TOP* sowie A, B und C für den *FRRT*, wie sie im Rahmen dieser Diplomarbeit vorgegeben wurden, zu finden.

Itemanordnung von *N-TOP* (Kubinger & Heuberger, in Vorb.)

Itemnummer								
	7b	9a	11a	12b	14a	16a	17b	18a
Testheft A	1	7	8	3	5	4	6	2
Testheft B	6	1	13	5	2	7	14	4

Itemnummer								
	19b	21b	22a	24b	25b	26a	27a	28a
Testheft A	11	9	13	14	10	15	16	12
Testheft B	3	12	10	11	16	8	9	15

Itemanordnung von *FRRT* (Kubinger, Poinstingl, Weidinger, & Skoda, in Vorb.)

Itemnummer								
	1	3	4	5	8	10	11	15
Testheft A	1	4	-	-	2	9	-	3
Testheft B	1	-	9	3	-	7	2	4
Testheft C	1	5	7	4	2	-	3	-

Itemnummer								
	16	17	18	19	23	27	29	31
Testheft A	6	12	-	-	5	10	8	7
Testheft B	-	12	13	6	8	10	-	5
Testheft C	8	-	13	6	-	10	9	-

Itemnummer								
	32	33	38	41	42	43	46	48
Testheft A	-	11	-	13	16	19	15	18
Testheft B	11	-	18	16	-	14	-	-
Testheft C	12	11	17	-	19	-	20	14

Itemnummer					
	53	56	57	61	64
Testheft A	17	-	14	-	20
Testheft B	-	20	15	19	17
Testheft C	18	15	-	16	-

B. Ansuchen für die Datenerhebung in Sicherheitsakademien

Josephine Kresnik, B.A.

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

E-Mail: - - -

Tel.: - - -

An Herrn Mag. Claus Polndorfer
Bundesministerium für Inneres
Referat II/5-Psychologischer Dienst
Herrengasse 7
A-1014 Wien

Wien, am 04.12.2014

Datenerhebung im Rahmen meiner Diplomarbeit:

Arbeitstitel „Tripelweiser Dimensionalitätsvergleich von Reasoningtests“

Sehr geehrter Herr Mag. Polndorfer,

im Rahmen meiner Diplomarbeit am Arbeitsbereich Psychologische Diagnostik an der Fakultät für Psychologie der Universität Wien beschäftige ich mich mit der Analyse und dem Vergleich psychologischer Testverfahren, die schlussfolgerndes Denken erfassen.

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

Um die psychologischen Tests analysieren und überprüfen zu können, sind die Daten von rund 300 Personen ab dem Alter von 16 Jahren notwendig.

Einige der Verfahren wurden im Rahmen der Diplomarbeit von Frau Mag. Gamsjäger bereits an Sicherheitsakademien (Traiskirchen und Ybbs) getestet. Daher wäre für deren Weiterentwicklung eine ähnliche Stichprobe sehr wichtig.

Ich ersuche daher Sicherheitsakademien um deren Unterstützung bzw. Mitwirkung an meinem Projekt. Die Studie koordiniere ich von Wien aus (Betreuung: Univ.-Prof. Mag. Dr.

Klaus D. Kubinger).

Die Testungen führe ich als Gruppentestungen in Räumlichkeiten der Sicherheitsakademie durch. Ich bin dabei sehr bemüht, den Aufwand für die jeweilige Akademie so gering wie möglich zu halten.

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

Es besteht die Möglichkeit – sollte dies gewünscht sein – den Teilnehmern als Dankeschön für die Teilnahme einen kurzen schriftlichen Ergebnisbericht über ihre Testergebnisse auszuhändigen.

Eine genaue Beschreibung des Testmaterials sowie des Untersuchungsablaufs finden Sie auf den folgenden Seiten.

Ich bitte höflichst um Genehmigung und Unterstützung der Studie.

Mit freundlichen Grüßen

Josephine Kresnik
Diplomandin an der Fakultät für Psychologie
Universität Wien

Anlagen:

Ansprechpartnerin
Untersuchungsziel und -ablauf
Informationen zu den eingesetzten Leistungstests
Stichprobenzusammensetzung

Ansprechpartnerin

Als Ansprechpartnerin stehe ich gerne unter folgenden Kontaktdaten zur Verfügung:

Josephine Kresnik, B.A.

E-Mail: - - -

Tel.: - - -

Diplomandin an der Fakultät für Psychologie
Institut für Angewandte Psychologie: Gesundheit, Entwicklung und Förderung
Universität Wien

Untersuchungsziel und -ablauf

Da die Testbatterie jeder zu testenden Person schriftlich vorgegeben wird, ist es möglich, die Aufgaben gleichzeitig einer Gruppe von Testpersonen vorzulegen; es handelt sich also um Gruppentestungen.

Um adäquate Aussagen hinsichtlich der Güte der zu analysierenden Tests treffen zu können, ist es nötig, dass jede Testperson einen kurzen Fragebogen und die unten angeführten Verfahren bearbeitet, wodurch sich pro Gruppe eine Gesamtdauer von etwa 60 Minuten ergibt; es ist dafür also ca. 1 Stunde notwendig. Für Testungen kommen sowohl Vormittags- als auch Nachmittagsunterrichtsstunden in Frage.

Die Stichprobe soll aus Personen bestehen, die zum Untersuchungszeitpunkt mindestens 16 Jahre alt sind.

Natürlich werden nur Jugendliche/Erwachsene getestet, die sich freiwillig dazu bereit erklären und – im Falle von Minderjährigen – deren Eltern ihr Einverständnis gegeben haben. Wenn klar ist, wie viele Personen an der Testung teilnehmen, werden Termine fixiert, an denen ich zum Testen an die Sicherheitsakademie kommen darf. Bei der Terminvereinbarung lege ich großen Wert darauf, den Ablauf des regulären Unterrichts so wenig wie möglich zu stören.

Die Erhebung soll ab Januar 2015 stattfinden.

Die gewonnenen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke verwendet und anonymisiert.

Informationen zur Testbatterie

Im Rahmen der Untersuchung kommen folgende Verfahren zur Anwendung:

- **N-TOP (Numerische Topologien; Kubinger & Heuberger, in Vorb.):** Dieser in Entwicklung befindliche Test erfasst schlussfolgerndes Denken im figuralen Bereich mithilfe von abstraktem Material. Die Teilnehmer sollen dabei Gesetzmäßigkeiten zwischen den Symbolen der ersten zwei Kästchen herausfinden und das fehlende Symbol im dritten Kästchen eintragen.
- **Family-Relation-Reasoning-Test (Kubinger, Poinstingl, Weidinger & Skoda, in Vorb.):**
Dieser Test dient zur Messung des kristallisierten schlussfolgernden Denkens mit Hilfe lexikalischen Materials.
- **WMT 2 (Formann, Waldherr & Pischwanger):**
Mit diesem Verfahren wird schlussfolgerndes Denken mithilfe von figuralen Darstellungen erhoben. Erfasst wird dabei die Fähigkeit, Regeln und Zusammenhänge bei figuralen Problemstellungen zu erkennen.
- **Fragebogen zum Anstrengungsvermeidungsverhalten (Rollett & Bertram)**

Stichprobenzusammensetzung

In dieser Studie sollen insgesamt 300 Personen ab dem Alter von 16 Jahren getestet werden.

C. Ansuchen für die Datenerhebung in Krankenpflegeschulen

Josephine Kresnik, B.A.

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

E-Mail: - - -

Tel.: - - -

An Gabriele Fuchs-Hlinka, MSc.
Geschäftsbereich Pflegemanagement
Generaldirektion KAV
Thomas Klestil Platz, 5. Stock, Top 143. 514
1030 Wien

Wien, am 12.03.2015

Datenerhebung im Rahmen unserer Diplomarbeit:

Arbeitstitel „Tripelweiser Dimensionalitätsvergleich von Reasoningtests“

Sehr geehrte Frau Fuchs-Hlinka,

im Rahmen meiner Diplomarbeit am Arbeitsbereich Psychologische Diagnostik an der Fakultät für Psychologie der Universität Wien beschäftige ich mich mit der Analyse und dem Vergleich psychologischer Testverfahren, die schlussfolgerndes Denken erfassen.

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

Um die psychologischen Tests analysieren und überprüfen zu können, sind die Daten von rund 300 Personen ab dem Alter von 16 Jahren notwendig.

Die Testungen führe ich als Gruppentestungen in Räumlichkeiten der Schule durch. Ich bin dabei sehr bemüht, den Aufwand für die jeweilige Schule so gering wie möglich zu halten.

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

Es besteht die Möglichkeit – sollte dies gewünscht sein – den Teilnehmern als Dankeschön für die Teilnahme einen kurzen schriftlichen Ergebnisbericht über ihre Testergebnisse auszuhändigen.

Eine genaue Beschreibung des Testmaterials sowie des Untersuchungsablaufs finden Sie auf den folgenden Seiten.

Ich bitte höflichst um Genehmigung und Unterstützung der Studie.

Mit freundlichen Grüßen

Josephine Kresnik
Diplomandin an der Fakultät für Psychologie
Universität Wien

Anlagen:

Ansprechpartnerin
Untersuchungsziel und -ablauf
Informationen zu den eingesetzten Leistungstests
Stichprobenzusammensetzung

Ansprechpartner

Als Ansprechpartnerin stehe ich gerne unter folgenden Kontaktdaten zur Verfügung:

Josephine Kresnik, B.A.

E-Mail: - - -

Tel.: - - -

Diplomandin an der Fakultät für Psychologie
Institut für Angewandte Psychologie: Gesundheit, Entwicklung und Förderung
Universität Wien

Untersuchungsziel und -ablauf

Da die Testbatterie jeder zu testenden Person schriftlich vorgegeben wird, ist es möglich, die Aufgaben gleichzeitig einer Gruppe von Testpersonen vorzulegen; es handelt sich also um Gruppentestungen.

Um adäquate Aussagen hinsichtlich der Güte der zu analysierenden Tests treffen zu können, ist es nötig, dass jede Testperson einen kurzen Fragebogen und drei Verfahren bearbeitet, wodurch sich pro Gruppe eine Gesamtdauer von ca. 60 Minuten ergibt. Für Testungen kommen sowohl Vormittags- als auch Nachmittagsunterrichtsstunden in Frage.

Die Stichprobe soll aus Personen bestehen, die zum Untersuchungszeitpunkt mindestens 16 Jahre alt sind.

Natürlich werden nur Jugendliche/Erwachsene getestet, die sich freiwillig dazu bereit erklären und – im Falle von Minderjährigen – deren Eltern ihr Einverständnis gegeben haben. Wenn klar ist, wie viele Personen an der Testung teilnehmen, werden Termine fixiert, an denen ich zum Testen an die Schule kommen dürfen. Bei der Terminvereinbarung lege ich großen Wert darauf, den Ablauf des regulären Unterrichts so wenig wie möglich zu stören.

Die Erhebung soll ab Februar 2015 stattfinden.

Die gewonnenen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke verwendet und anonymisiert.

Informationen zur Testbatterie

Im Rahmen der Untersuchung kommen folgende Verfahren zur Anwendung:

- **AN-TOP (Alphanumerische Topologien; Kubinger & Heuberger, in Vorb.):**
Dieser in Entwicklung befindliche Test erfasst schlussfolgerndes Denken im figuralen Bereich mithilfe von abstraktem Material. Die Teilnehmer sollen dabei Gesetzmäßigkeiten zwischen den Symbolen der ersten zwei Kästchen herausfinden und das fehlende Symbol im dritten Kästchen eintragen.
- **Family-Relation-Reasoning-Test (Kubinger, Poinstingl, Weidinger & Skoda, in Vorb.):**
Dieser Test dient zur Messung des kristallisierten schlussfolgernden Denkens mit Hilfe lexikalischen Materials.
- **WMT 2 (Formann, Waldherr & Piswanger):**
Mit diesem Verfahren wird schlussfolgerndes Denken mithilfe von figuralen Darstellungen erhoben. Erfasst wird dabei die Fähigkeit, Regeln und Zusammenhänge bei figuralen Problemstellungen zu erkennen.
- **Fragebogen zum Anstrengungsvermeidungsverhalten (Rollett & Bertram)**

Stichprobenzusammensetzung

In dieser Studie sollen insgesamt 300 Personen ab dem Alter von 16 Jahren getestet werden.

D. Informationsschreiben für Lehrkräfte

Josephine Kresnik, B.A.

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

LehrerInnenbrief

Liebe Lehrende!

im Rahmen meiner Diplomarbeit an der Fakultät für Psychologie der Universität Wien (Betreuung: Univ.-Prof. Mag. Dr. Klaus D. Kubinger) beschäftige ich mich mit der Analyse und dem Vergleich psychologischer Verfahren zur Erfassung der Fähigkeit zum schlussfolgernden Denken. Neben bereits etablierten psychologischen Tests werden in meiner Untersuchung auch einige neu entwickelte Verfahren einbezogen, die hinsichtlich ihrer Qualität und des erfassten Fähigkeitsbereichs im Vergleich zu anderen Tests beurteilt und eingeordnet werden sollen. Forschungen dieser Art sind wichtig, um den Standard psychologischer Diagnostik laufend zu verbessern und dem neuesten Stand der Wissenschaft anpassen zu können.

Zum Zweck meiner Untersuchung werden Schülerinnen und Schülern ab dem (Mindest-)Alter von 16 Jahren verschiedene psychologische Tests zur Bearbeitung vorgegeben. Beim Aufgabenmaterial, das im Rahmen meiner Studie zur Anwendung kommt, handelt es sich um drei verschiedene psychologische Tests, die die Fähigkeit zum logischen Schlussfolgern auf unterschiedliche Weise erheben und einen Fragebogen. Die Untersuchungen finden während der Schulzeit statt und dauert ca. 50 Minuten. Die Testungen werden im Klassenverband durchgeführt, wobei die SchülerInnen die Tests jede/r für sich schriftlich bearbeiten werden.

Als Dankeschön für die Teilnahme besteht die Möglichkeit, den SchülerInnen einen kurzen schriftlichen Ergebnisbericht über seine/ihre Stärken und Schwächen zukommen zu lassen. Die gewonnenen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke verwendet und **anonymisiert**.

Für Rückfragen stehe ich gerne telefonisch oder per E-Mail zur Verfügung:

- Josephine Kresnik; E-Mail: - - -

Mit freundlichen Grüßen und herzlichem Dank im Voraus!

Josephine Kresnik

Diplomandin an der Fakultät für Psychologie Universität Wien

E. Testleiter/innen-anweisung für die Durchführung der Testungen

Testleiteranweisung

Testungen: Zoé Josephine Kresnik

Informationen zum Ablauf der Testungen

1) Einleitungstext/Testleiterinstruktion

2) Es ist jedem/jeder Teilnehmer/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. Die Testbatterie besteht aus 3 Untertest, dem AVT (Fragebogen) und einem Blatt für die demographischen Daten in folgender **Reihenfolge**:

- ⑩ Demographisches Datenblatt
- ⑩ Family Realtion Reasoning Test
- ⑩ Numerische Topologien
- ⑩ Wiener Matrizen-Test 2
- ⑩ Fragebogen: AVT

Das Testheft Family Realtion Reasoning Test liegt in Form A, B und C vor; das Testheft N-Top in Form A und B. Der Test WMT 2 und der Fragebogen AVT liegen in einer Form vor. Testpersonen, die nebeneinander sitzen, sollten jeweils eine unterschiedliche Testform erhalten.

Das heißt die **Testhefte werden abwechselnd an die Teilnehmer/innen 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.

Testperson	Reihenfolge der Testheft
1.	0 – 1A – 2A – 3 – AVT
2.	0 – 1B – 2B – 3 – AVT
3.	0 – 1C – 2A – 3 – AVT
4	0 – 1A – 2B – 3 – AVT
5.	0 – 1B – 2A – 3 – AVT
6.	0 – 1C – 2B – 3 – AVT
7.	0 – 1A – 2A – 3 – AVT
8.	usw. ...

3) **Beginn der Testung** (ab Einleitungstext) und **Ende der Testung** (sobald die Bearbeitung des letzten Tests vorbei ist) sollten bitte am Testleiterprotokoll notiert werden, wie auch Datum, Schule/Ort und Klasse sowie Namen der Testleiter.

4) Um den Testpersonen Rückmeldung über ihre Testergebnisse geben zu können, erhält jeder zusammen mit dem Testheft einen Zettel mit seinem individuellen Probandencode und meinen Kontaktdaten. Dieser Zettel darf mitgenommen werden.

5) Die Teilnehmer/innen sollen einen Stift bereit legen. 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).

6) Wenn eine Instruktion nicht gleich verstanden wird (z.B.: Beispielaufgabe) oder es noch Fragen gibt, so werden diese bitte nochmals langsam mit allen gemeinsam wiederholt oder mit anderen Worten erklärt.

7) 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 **großen Strich unter die zuletzt bearbeitete Aufgabe** zu ziehen.

8) Sollten vor Ablauf der Bearbeitungszeit alle Testpersonen fertig sein, so kann 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 auf dem Testleiterprotokoll notiert.

9) Sollten Teilnehmer/innen nachfragen, ob sie bei Unsicherheit **raten** sollen oder die Aufgabe auslassen sollen, so antworten Sie entsprechend der jeweiligen Testinstruktion! (Beispielsweise wird beim WMT 2 explizit darauf hingewiesen, bei Unsicherheit Seitens der Testpersonen nichts an zu kreuzen.)

10) Bitte die Teilnehmer/innen animieren, die **Aufgaben der Reihe nach zu bearbeiten**. Wenn eine Aufgabe zu schwierig sein sollte, dann kann mitgeteilt werden, dass sie die

Testleiteranweisung

Testungen: Zoé Josephine Kresnik

nächste Aufgabe versuchen möchten, weil auch wieder einfachere kommen. Es sollte jedoch nicht dazu animiert werden, zwischen den Testaufgaben „herumzuspringen“.

11) Bitte darauf achten, dass die Teilnehmer/innen auch **umblättern**, wenn „BITTE UMBLÄTTERN“ am Ende der Seite steht.

12) Am **Ende eines Tests** müssen die Testpersonen 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.

13) Es ist darauf zu achten, dass die Testpersonen nicht vom Nachbarn **abschreiben oder miteinander reden**. Liegt der Verdacht nahe, dass ein/e Teilnehmer/in abgeschrieben hat, so ist das beim Absammeln am Testheft zu notieren.

14) Bitte eventuelle **Auffälligkeiten und Probleme** pro Untertest am Testleiterprotokoll notieren!

15) Den Testpersonen darf bei der Bearbeitung der Aufgaben **nicht geholfen** oder Hinweise zur Lösung gegeben werden. Bitte auch **keine inhaltlichen Fragen beantworten**. Bei derartigen Fragen bitte sagen, dass man diese leider nicht beantworten darf.

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

Zur Informationen:

- *Kursiv* geschriebener Text ist der vom/von der Testleiter/in gesprochene Text, der an die Testpersonen gerichtet ist.
- „Normal“ geschriebener Text ist Erklärung für die Testleiter/innen.

F. Testinstruktionen

Testleiteranweisung

Testungen: Zoé Josephine Kresnik

Testinstruktionen

1) Einleitung und Austeilen der Testhefte:

*„Herzlich willkommen und vielen Dank für Ihre Teilnahme an dieser Erhebung.
Gleich werden meine Kollegen . . . jedem von Ihnen ein Testheft austeilen.*

*Ich bitte Sie die Testhefte **erst aufzuschlagen** und auch immer **erst umzublättern**, wenn ich Sie dazu auffordere! Steht am Ende einer Seite „Bitte Umblättern“ - dann dürfen Sie nicht warten und sollen **sofort** umblättern.*

*Die Erhebung ist **freiwillig** und vollkommen **anonym**, d.h. Sie müssen und sollen Ihren Namen nicht angeben.*

Jeder Teilnehmer erhält jedoch die Möglichkeit ein Feedback über die eigenen Testergebnisse zu erhalten. Dafür benötigen Sie ihren individuellen Personencode, der auf diesem kleinen Zettel steht. Wenn Sie Ihre Testergebnisse erfahren möchten, dann senden Sie mir bitte eine Email mit „Personencode“ im Betreff. Sollten Sie ihren Personen-code verlieren, gibt es leider keine Möglichkeit mehr die Ergebnisse richtig zu zuordnen und dementsprechend zu erfahren.

Dieses Testheft besteht aus vier verschiedenen Aufgabengruppen. Zu jeder wird Ihnen anhand von Beispielen genau erklärt, was Sie zu tun haben.

Einige Aufgaben werden Ihnen vielleicht leichter fallen, andere schwerer. Wenn Sie eine Aufgabe nicht lösen können, gehen Sie einfach zur nächsten weiter. Versuchen Sie aber bitte die Aufgaben der Reihe nach zu lösen.

– Strich nach der zuletzt bearbeiteten Aufgabe!

– Bitte nicht Schummeln und miteinander während der Testung sprechen!

Für jede Aufgabengruppe steht Ihnen eine bestimmte Zeit zur Verfügung. Sollten Sie früher fertig sein, bleiben Sie bitte ruhig sitzen und stören Sie die anderen nicht. Sie können die Zeit auch nutzen, um Ihre Antworten zu kontrollieren oder ausgelassene Aufgaben noch einmal zu versuchen.

Testleiteranweisung

Testungen: Zoé Josephine Kresnik

Sollten Sie während der Testbearbeitung Fragen haben, melden Sie sich bitte per Handzeichen.

Bitte füllen Sie nun auf der Vorderseite Ihre persönlichen Daten aus!“

2. Family Realtion Reasoning Test

Der/die Testleiter/in liest folgendes vor, während die Testpersonen in ihrem Testheft mitlesen können:

Auf den nächsten Seiten werden Ihnen Sätze über Familienbeziehungen präsentiert. Sie sollen mit Hilfe der gegebenen Informationen die fehlende Beziehung finden.

Sehen Sie sich bitte zuerst folgende Beispielaufgaben an:

Die Beispielaufgaben sind laut vorzulesen. Die Lösung soll von einer Testperson genannt und bei Fragen vom/von der Testleiter/in näher erklärt werden.

Beispielaufgabe 1: Karin hat die Tochter Tina. Peter ist der Bruder von Tina.

Karin ist Peters _____.

Beispielaufgabe 2: Harry ist der Sohn von Edith Edith hat die Tochter Lisa.

Lisa ist Harrys _____.

Der/die Testleiter/in liest nun die Instruktion weiter laut vor, während die Teilnehmer/innen im Testheft mitlesen können:

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 zur Lösung benötigten Informationen sind im Text

Testleiteranweisung

Testungen: Zoé Josephine Kresnik

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 Aufgabe fertig sind, warten Sie bitte auf ein Zeichen des Testleiters.

Gibt es von Ihrer Seite noch Fragen?

Es ist besonders darauf zu achten, dass während der Bearbeitungszeit **keinerlei Notizen** gemacht werden!

Der/die Testleiter/in gibt ein klares Startzeichen zum Umblättern er/sie instruiert:

Los geht's. Bitte alle umblättern!

Nach Ablauf der Bearbeitungszeit von **15 Minuten** ist mit dem Wort „Stopp!“ abzubrechen.

Die Teilnehmer/innen werden nochmals darauf hingewiesen, unter die zuletzt bearbeitete Aufgabe einen Strich zu ziehen.

Bitte ziehen Sie nun einen Strich unter die zuletzt bearbeitete Aufgabe und blättern Sie nun zum nächsten Abschnitt weiter!

3. Numerische Topologien

Der/die Testleiter/in liest folgendes vor, während die Testpersonen in ihrem Testheft mitlesen können:

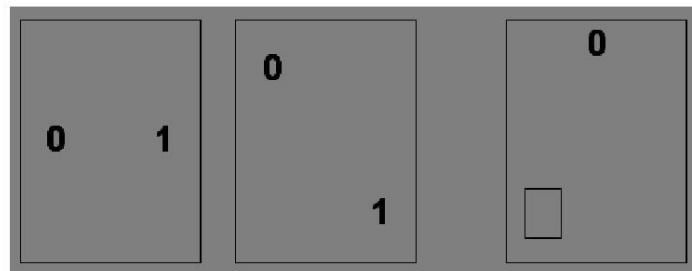
In diesem Test geht es darum, Gesetzmäßigkeiten zu erkennen und entsprechend anzuwenden. Die dabei jeweils gesuchte Zahl ist in das vorgegebene Kästchen einzutragen Fünf Beispielaufgaben illustriert das.

Testleiteranweisung

Testungen: Zoé Josephine Kresnik

Die Beispielaufgaben sind laut vorzulesen. Bei Fragen sollen die Beispiele vom/von der Testleiter/in in anderen Worten nochmals erklärt werden.

Beispielaufgabe 1:



Auch bei allen folgenden Aufgaben gibt es immer eine bestimmte Gesetzmäßigkeit, die die fehlende Zahl eindeutig festlegt.

Für die Testbearbeitung stehen Ihnen 10 Minuten zur Verfügung.

Falls Sie vor Ablauf der Zeit mit der Bearbeitung der 16 Aufgaben fertig sind, warten Sie bitte auf ein Zeichen des Testleiters.

Gibt es von Ihrer Seite noch Fragen?

Der/die Testleiter/in gibt ein klares Startzeichen zum Umblättern er/sie instruiert:

Los geht's. Bitte alle umblättern!

Nach Ablauf der Bearbeitungszeit von **10 Minuten** ist mit dem Wort „Stopp!“ abzubrechen.

Die Teilnehmer/innen werden nochmals darauf hingewiesen, unter die zuletzt bearbeitete Aufgabe einen Strich zu ziehen.

Bitte ziehen Sie nun einen Strich unter die zuletzt bearbeitete Aufgabe und blättern Sie nun zum nächsten Abschnitt weiter!

Testleiteranweisung

Testungen: Zoé Josephine Kresnik

4. Wiener Matrizen-Test 2

Der/die Testleiter/in liest die Testinstruktion laut vor, während die Testpersonen in ihrem Testheft mitlesen können:

*Für die Testinstruktion sowie die Beispielitems wird auf das Manual des *WMT 2* verwiesen.

Der/die Testleiter/in gibt ein klares Startzeichen zum Umblättern er/sie instruiert:

Los geht's. Bitte alle umblättern!

Nach Ablauf der Bearbeitungszeit von **15 Minuten** ist mit dem Wort „*Stopp!*“ abzubrechen.

Die Teilnehmer/innen werden nochmals darauf hingewiesen, unter die zuletzt bearbeitete Aufgabe einen Strich zu ziehen.

Bitte ziehen Sie nun einen Strich unter die zuletzt bearbeitete Aufgabe und blättern Sie nun zum nächsten Abschnitt weiter!

5. Fragebogen AVT

Der/die Testleiter/in liest folgendes vor, während die Testpersonen in ihrem Testheft mitlesen können:

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“, „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?

Die Testpersonen sind darauf hinzuweisen, dass es keine Zeitbeschränkung für die

Testleiteranweisung

Testungen: Zoé Josephine Kresnik

Bearbeitung des Fragebogen gibt. Sie sollen jedoch nach Möglichkeit zügig arbeiten.

Wenn **ALLE** mit der Bearbeitung fertig sind, ist die Testung folgendermaßen abzuschließen:

Bitte schließen Sie nun alle ihre Testhefte. Meine Kollegen und ich werden die Hefte einsammeln. Ich bedanke mich ganz herzlich für Ihre Teilnahme.

Bitte darauf achten, dass alle Testhefte vollständig abgegeben/eingesammelt wurden!

Sollten demographische Daten am Deckblatt fehlen, ist bei den betreffenden Personen nachzufragen, ob diese vergessen oder aus einem bestimmten Grund nicht angegeben wurden.

Auch bei der Lehrkraft ist sich für die Testung zu bedanken.

G. Beispiel eines schriftlichen Ergebnisberichts für die Teilnehmer/innen

Personencode: x x x

Datum der Testung: x x x

Ergebnisbericht

Liebe Teilnehmerin, lieber Teilnehmer,

im Folgenden sind Ihre Ergebnisse der psychologischen Tests beschrieben, an denen Sie im Rahmen der Untersuchungen von Frau Josephine Kresnik teilgenommen haben.

Bei den Ergebnissen ist zu berücksichtigen, dass es sich bei den durchgeführten Tests um **in Entwicklung befindliche Verfahren** handelt, d. h. sie sind zum Teil noch nicht hinreichend überprüft und die Ergebnisse demnach nicht im selben Maße aussagekräftig wie dies bei bereits etablierten Tests der Fall ist.

Die **Aussagekraft der Ergebnisse** hängt auch damit zusammen, wie gewissenhaft und ernsthaft Sie die Tests bearbeitet haben. Je nach dem sind die Ergebnisse zuverlässige Darstellung der erhobenen Fähigkeiten.

Die Ergebnisse sind in Form von **Prozenträngen (PR)** angeführt, wobei Werte von 25 bis 75 als durchschnittlich gelten. In diesem Bereich liegen etwa 50% der Vergleichspersonen mit ihren Ergebnissen. Prozentränge unter 25 stellen eine unterdurchschnittliche Leistung dar; Prozentränge über 75 eine überdurchschnittliche Leistung.

Der Prozentrang gibt jeweils an, wie viel Prozent der Vergleichsgruppe eine gleich gute oder niedrigere Leistung erzielt haben.

Als Vergleichsgruppe dienen Schülerinnen und Schüler aus Krankenpflegeschulen und Sicherheitsakademien in Österreich.

Personencode: x x x

Datum der Testung: x x x

<u>Name des Tests und Aufgabenbeschreibung</u>	<u>Testergebnis</u>	<u>Ergebnisbeschreibung</u>
Family-Relation-Reasoning-Test (Kubinger, Poinstingl, Weidinger & Skoda, in Vorb.)		
Bei diesem Test werden Sätze über Verwandtschaftsbeziehungen präsentiert, wobei eine fehlende Beziehung durch logisches Kombinieren zu finden ist.	PR = 20	Die Testperson hat mit einem Prozentrang von 20 eine unterdurchschnittlich ausgeprägte Fähigkeit zum logisch schlussfolgernden Denken im lexikalischen Bereich.
Numerische Topologien (Kubinger & Heuberger, in Vorb.)		
Bei diesem Verfahren sollen Gesetzmäßigkeiten zwischen Zahlen erkannt werden, wobei die gesuchte Zahl in ein leeres Kästchen einzutragen ist.	PR = 40	Dieser Test erfasst die Fähigkeit zum schlussfolgernden Denken mithilfe von Zahlenmaterial. Die Leistung der Testperson liegt mit einem Prozentrang von 40 im durchschnittlichen Bereich.
Wiener Matrizen Test 2 (Formann, Waldherr & Pischwanger, 2011)		
Die Aufgabe bei diesem Test ist es, in einer Matrix von Figuren Regeln einer Anordnung zu finden und anstelle des Fragezeichens eine passende Figur auszuwählen.	PR = 80	Dieser Wert zeigt die Fähigkeit, aus vorgegebenen Elementen Gesetzmäßigkeiten herauszufinden und diese anzuwenden. Die Testperson liegt hier im überdurchschnittlichen Bereich.
Anstrengungsvermeidungstest Sch (Rollett & Rollett, 2000)		
Dieser Fragebogen dient zur Erhebung der Tendenz zur Anstrengungsvermeidung.	PR = 50	Mit einem Prozentrang von 50 liegt die Anstrengungsvermeidungstendenz im durchschnittlichen Bereich.

Curriculum Vitae

Name: Zoé Josephine Babett Kresnik, BA
Geburtsdatum/ -ort: 12.12.1987 in Mannheim-Neckarau
Staatsangehörigkeit: Deutsch

Schul- und Ausbildung

9/1994 – 7/1998 Friedrich-Ebert Grundschule, Heidelberg

9/1998 – 6/2008 Helmholtz – Gymnasium, Heidelberg, Abschluss: Abitur

10/2009 – 05/2013 Bachelorstudium der Theater-, Film- und Medienwissenschaft an der Universität Wien

seit 10/2009 Diplomstudium der Psychologie an der Universität Wien

seit 10/2013 Masterstudium der Theater-, Film- und Medienwissenschaft an der Universität Wien

Praktika

Schulzeit Sozialpraktikum „Diakonische Hausgemeinschaft e. V.“ in Heidelberg

10/2014 – 07/2015 6-Wochenpflichtpraktikum in der klinisch-psychologischen Praxis von Frau Mag. Krauland in Wien