



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

Zur Dimensionalität von Reasoning-Tests:

Ein Beitrag zur Validierung einer neuen Testbatterie
zur Erfassung des logisch-schlussfolgernden Denkens

Verfasserin

Bernadette Schnait, BA

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Juli 2015

Studienkennzahl: A 298

Studienrichtung: Psychologie

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

DANKSAGUNG

An dieser Stelle möchte ich mich bei jenen Menschen bedanken, die mich auf meinem Studienweg und beim Entstehungsprozess meiner Diplomarbeit begleitet und unterstützt haben:

In erster Linie danke ich meinem Betreuer, Herrn Univ.-Prof. Mag. Dr. Klaus D. Kubinger, für die Ermöglichung dieser Diplomarbeit und für seine hilfreichen Anregungen und die kritische Auseinandersetzung mit meiner Arbeit.

Ein besonderer Dank gebührt meiner Subbetreuerin, Frau Mag. Serra Ünal, die stets bei Fragen und Unsicherheiten zur Verfügung stand und mich bei der Erarbeitung des Themas und bei der Untersuchungsplanung mit ihren konstruktiven Vorschlägen und Hilfestellungen sehr unterstützte. Danke auch an Frau Mag. Bettina Hagenmüller, die mich vor allem in der Anfangsphase meiner Diplomarbeit fachlich begleitete.

Für die Ermöglichung der Datenerhebung an ihren Schulen bedanke ich mich herzlich bei den Direktorinnen bzw. Direktoren der HAK Tulln, der HAK Hollabrunn, der HTL Mödling, des BG/BRG Schwechat sowie des BG Wiener Neustadt Zehnergasse.

Meinen Studienkolleginnen, die mich bei der Durchführung der Testungen begleitet haben, danke ich für ihre wertvolle Mitarbeit. Besonders bedanken möchte ich mich dabei bei Kerstin Grafl und Veronika Necker, die mich durch fachlichen und freundschaftlichen Austausch auch moralisch bei der Erstellung dieser Arbeit unterstützt haben.

Ausdrücklicher Dank gilt auch meinem Freund, Michael Moldaschl, der mir in unzähligen Phasen meines Studiums und meiner Diplomarbeit eine unsagbar große Stütze war und mir in den schwierigsten und schönsten Momenten zur Seite stand. Nicht zuletzt möchte ich mich bei ihm auch für die Hilfe bei der statistischen Auswertung bedanken.

Meinen Freundinnen, allen voran Silvia Mühle und Bianca Schönhofer, danke ich herzlich für ihre emotionale Unterstützung und Bianca Schönhofer zudem auch für ihre geschätzten sprachlichen Anregungen beim Korrekturlesen dieser Arbeit.

Nicht weniger dankbar bin ich meinen Eltern, Renate und Erich Schnait, die mir meinen Berufswunsch ermöglicht und mich geduldig beim Verfolgen meiner Ziele unterstützt haben.

ABSTRACT (Deutsch)

Im Rahmen dieser Arbeit wurde eine Testbatterie aus sechs psychologisch-diagnostischen Verfahren zur Erfassung des schlussfolgernden Denkens (*Reasoning*) in Hinblick auf ihre Dimensionalität und Validität untersucht. Der Untertest *Formale Folgerichtigkeit* aus der Intelligenz-Testbatterie *AID-Gruppe* (Kubinger & Hagenmüller, in Vorb.), der *Färbige Matrizentest (FMT)* (Undeutsch, 2010a) sowie das Verfahren *N-TOP (Numerische Topologien)* (Kubinger & Heuberger, in Vorb.) messen logisch-schlussfolgerndes Denken im Bereich der fluiden Intelligenz sensu Cattell, während die Tests *Sprachliche Analogien* (Kubinger & Ünal, in Vorb.), *Figurale Analogien* (Ünal & Kubinger, in Vorb.) und *Gleichungen* (Kubinger & Gamsjäger, in Vorb.) kristallisiertes schlussfolgerndes Denken erfassen.

Eine ausgewählte Anzahl an Items der Reasoning-Tests wurde einer Überprüfung auf Konformität mit dem dichotomen logistischen Testmodell von Rasch sowie einer Konstruktvalidierung mittels Faktorenanalyse unterzogen. Dafür wurden Daten an insgesamt 265 Schülerinnen und Schülern aus Maturaklassen allgemeinbildender und berufsbildender höherer Schulen (12. bzw. 13. Schulstufe) in Niederösterreich erhoben, die die Verfahren im Rahmen von Gruppentestungen bearbeiteten.

Die Rasch-Modell-Analysen ergaben a-priori Modellgeltung der Items des *FMT* sowie jener von *N-TOP* und *Gleichungen*. Den Items von *Formale Folgerichtigkeit*, *Sprachliche Analogien* und *Figurale Analogien* ist nach Ausscheiden von ca. 5, 24 bzw. 21 Prozent der Items a-posteriori Modellgeltung zuzusprechen. Laut den Ergebnissen der Hauptkomponentenanalyse sind die Leistungen der Testpersonen in den sechs Reasoning-Tests entweder auf einen Generalfaktor („abstraktes Denken“) oder auf zwei gemeinsame Faktoren („logisch-schlussfolgerndes Denken“ und „kristallisiertes schlussfolgerndes Denken“) zurückzuführen. Die Korrelationskoeffizienten nach Pearson zwischen den Personenparametern der einzelnen Verfahren betragen zwischen $r = 0,14$ und $r = 0,38$.

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

ABSTRACT (English)

The aim of this paper was to analyze the dimensionality and validity of a battery consisting of six psychological tests measuring *reasoning* skills. Of these, the tests *Formale Folgerichtigkeit*, which is a subtest of the intelligence test battery AID-Group (Kubinger & Hagenmüller, in progress), *Färbiger Matrizentest (FMT)* (Undeutsch, 2010a) as well as *N-TOP (Numerische Topologien)* (Kubinger & Heuberger, in progress) are designed to assess logical reasoning as part of Cattell's fluid intelligence. The tests *Sprachliche Analogien* (Kubinger & Ünal, in progress), *Figurale Analogien* (Ünal & Kubinger, in progress) and *Gleichungen* (Kubinger & Gamsjäger, in progress), on the other hand, can be used to measure crystallized reasoning skills.

A selected number of items of the reasoning tests were analyzed with regard to fit of the Rasch model and construct validity using factor analysis. For this purpose, a total of 265 students attending their final year of upper secondary education in vocational secondary schools (BHS; 13th grade) or academic secondary schools (AHS; 12th grade) in Lower Austria were administered the six tests in a group setting.

The results of the Rasch model analyses showed that this model holds for all the examined items of *FMT*, *N-TOP* and *Gleichungen*. As for the tests *Formale Folgerichtigkeit*, *Sprachliche Analogien* and *Figurale Analogien*, some items (5, 24 and 21% resp.) had to be excluded from the item pool in order to fit the Rasch model a posteriori.

According to the results of the Principal Component Analysis, the performance variance of the testees is accounted for either by one general factor ("abstract thinking") or by two joint factors ("logical reasoning" and "crystallized reasoning"). The Pearson correlation coefficients between the person parameters of the individual tests lie between $r = 0.14$ and $r = 0.38$.

Keywords: reasoning, fluid, crystallized, Rasch model, construct validity

INHALTSVERZEICHNIS

I. EINLEITUNG.....	1
II. THEORETISCHER TEIL	3
1. Intelligenztheoretische Grundlagen	3
1.1. Die Zwei-Faktoren-Theorie von Spearman	3
1.2. Die Primary Mental Abilities nach Thurstone	4
1.3. Das Modell der fluiden und kristallisierten Intelligenz von Cattell.....	4
1.4. Das Berliner Intelligenzstrukturmodell von Jäger	6
2. Das Konstrukt <i>Reasoning</i>	7
3. Testtheoretische Grundlagen	10
3.1. Das Gütekriterium Validität.....	10
3.2. Das dichotome logistische Testmodell von Rasch.....	12
3.3. Die exploratorische Faktorenanalyse	15
III. EMPIRISCHER TEIL.....	19
4. Ziel der Untersuchung und Hypothesen	19
5. Beschreibung der Testbatterie	21
5.1. Der Untertest <i>Formale Folgerichtigkeit</i> aus AID-Gruppe	22
5.2. <i>Sprachliche Analogien</i>	23
5.3. Der <i>Färbige Matrizentest (FMT)</i>	24
5.4. <i>Numerische Topologien (N-TOP)</i>	25
5.5. <i>Figurale Analogien</i>	25
5.6. <i>Gleichungen</i>	26

6. Planung und Durchführung der Untersuchung	27
7. Stichprobe	30
8. Darstellung der Ergebnisse	31
8.1. Analysen zur Überprüfung der Geltung des Rasch-Modells	31
8.1.1. Ergebnisse der Rasch-Modell-Analysen bei <i>Formale Folgerichtigkeit</i>	33
8.1.2. Ergebnisse der Rasch-Modell-Analysen bei <i>Sprachliche Analogien</i>	39
8.1.3. Ergebnisse der Rasch-Modell-Analysen beim <i>Färbigen Matrizentest</i>	50
8.1.4. Ergebnisse der Rasch-Modell-Analysen bei <i>N-TOP</i>	52
8.1.5. Ergebnisse der Rasch-Modell-Analysen bei <i>Figurale Analogien</i>	55
8.1.6. Ergebnisse der Rasch-Modell-Analysen bei <i>Gleichungen</i>	66
8.2. Durchführung der exploratorischen Faktorenanalyse	69
8.2.1. Ergebnisse der Hauptkomponentenanalyse	69
8.3. Korrelative Zusammenhänge zwischen den Reasoning-Tests	73
 IV. DISKUSSION UND AUSBLICK	 75
 V. ZUSAMMENFASSUNG	 81
 VI. LITERATURVERZEICHNIS	 83
 VII. ANHANG	 89

ABBILDUNGSVERZEICHNIS

Abbildung 1: Definition des induktiven Denkens (nach Klauer, 2001b, S. 169).....	8
Abbildung 2: Itembeispiel des Untertests <i>Formale Folgerichtigkeit</i> aus AID-Gruppe (Kubinger & Hagenmüller, in Vorb.).....	23
Abbildung 3: Itembeispiel aus <i>Sprachliche Analogien</i> (Kubinger & Ünal, in Vorb.)	23
Abbildung 4: Itembeispiel aus dem <i>FMT</i> (Undeutsch, 2010a).....	24
Abbildung 5: Itembeispiel aus <i>N-TOP</i> (Kubinger & Heuberger, in Vorb.)	25
Abbildung 6: Itembeispiel aus <i>Figurale Analogien</i> (Ünal & Kubinger, in Vorb.).....	26
Abbildung 7: Itembeispiel aus <i>Gleichungen</i> (Kubinger & Gamsjäger, in Vorb.).....	26
Abbildung 8: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>Formale Folgerichtigkeit</i> , alle Items).....	34
Abbildung 9: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>Formale Folgerichtigkeit</i> , alle Items).....	34
Abbildung 10: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>Formale Folgerichtigkeit</i> , alle Items).....	35
Abbildung 11: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>Formale Folgerichtigkeit</i> , nach Ausschluss von Item 95).....	37
Abbildung 12: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>Formale Folgerichtigkeit</i> , nach Ausschluss von Item 95).....	37
Abbildung 13: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>Formale Folgerichtigkeit</i> , nach Ausschluss von Item 95).....	38
Abbildung 14: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>Sprachliche Analogien</i> , alle Items).....	40
Abbildung 15: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>Sprachliche Analogien</i> , alle Items).....	41
Abbildung 16: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>Sprachliche Analogien</i> , alle Items).....	43
Abbildung 17: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>Sprachliche Analogien</i> , nach Ausschluss der Items 27, 8, 41, 25 und 36)	48
Abbildung 18: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>Sprachliche Analogien</i> , nach Ausschluss der Items 27, 8, 41, 25 und 36)	49
Abbildung 19: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>Sprachliche Analogien</i> , nach Ausschluss der Items 27, 8, 41, 25 und 36)	49
Abbildung 20: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>FMT</i> , alle Items)	51
Abbildung 21: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>FMT</i> , alle Items).....	51

Abbildung 22: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>FMT</i> , alle Items).....	52
Abbildung 23: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>N-TOP</i> , alle Items)....	53
Abbildung 24: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>N-TOP</i> , alle Items).....	54
Abbildung 25: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>N-TOP</i> , alle Items).....	54
Abbildung 26: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>Figurale Analogien</i> , alle Items).....	57
Abbildung 27: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>Figurale Analogien</i> , alle Items).....	58
Abbildung 28: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>Figurale Analogien</i> , alle Items).....	60
Abbildung 29: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>Figurale Analogien</i> , nach Ausschluss der Items 2, 7, 24, 5 und 23)	65
Abbildung 30: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>Figurale Analogien</i> , nach Ausschluss der Items 2, 7, 24, 5 und 23).....	65
Abbildung 31: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>Figurale Analogien</i> , nach Ausschluss der Items 2, 7, 24, 5 und 23).....	66
Abbildung 32: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>Gleichungen</i> , alle Items).....	67
Abbildung 33: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“ (<i>Gleichungen</i> , alle Items).....	68
Abbildung 34: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“ (<i>Gleichungen</i> , alle Items).....	68
Abbildung 35: Scree-Plot der Eigenwerte (Hauptkomponentenanalyse)	70

TABELLENVERZEICHNIS

Tabelle 1: Überblick über die untersuchte Reasoning-Test-Batterie	22
Tabelle 2: Itemanzahl und Bearbeitungszeit pro Verfahren.....	29
Tabelle 3: Verteilung der Stichprobe in Bezug auf das Alter in Jahren ($n = 265$).....	30
Tabelle 4: Ergebnisse der Likelihood-Ratio-Tests (<i>Formale Folgerichtigkeit</i> , alle Items).....	33
Tabelle 5: Ergebnisse des z -Tests für das Teilungskriterium „Schultyp“ (<i>Formale Folgerichtigkeit</i> , alle Items)	35
Tabelle 6: Ergebnisse der Likelihood-Ratio-Tests (<i>Formale Folgerichtigkeit</i> , nach Ausschluss von Item 95)	36
Tabelle 7: Ergebnisse der Likelihood-Ratio-Tests (<i>Sprachliche Analogien</i> , alle Items).....	39
Tabelle 8: Ergebnisse des z -Tests für das Teilungskriterium „Score“ (<i>Sprachliche Analogien</i> , alle Items)	40
Tabelle 9: Ergebnisse des z -Tests für das Teilungskriterium „Geschlecht“ (<i>Sprachliche Analogien</i> , alle Items)	42
Tabelle 10: Ergebnisse des z -Tests für das Teilungskriterium „Schultyp“ (<i>Sprachliche Analogien</i> , alle Items)	43
Tabelle 11: Ergebnisse der Likelihood-Ratio-Tests (<i>Sprachliche Analogien</i> , nach Ausschluss von Item 27)	45
Tabelle 12: Ergebnisse der Likelihood-Ratio-Tests (<i>Sprachliche Analogien</i> , nach Ausschluss der Items 27 und 8).....	46
Tabelle 13: Ergebnisse der Likelihood-Ratio-Tests (<i>Sprachliche Analogien</i> , nach Ausschluss der Items 27, 8 und 41).....	46
Tabelle 14: Ergebnisse der Likelihood-Ratio-Tests (<i>Sprachliche Analogien</i> , nach Ausschluss der Items 27, 8, 41 und 25).....	47
Tabelle 15: Ergebnisse der Likelihood-Ratio-Tests (<i>Sprachliche Analogien</i> , nach Ausschluss der Items 27, 8, 41, 25 und 36).....	48
Tabelle 16: Ergebnisse der Likelihood-Ratio-Tests (<i>FMT</i> , alle Items).....	50
Tabelle 17: Ergebnisse der Likelihood-Ratio-Tests (<i>N-TOP</i> , alle Items).....	53
Tabelle 18: Bei den Rasch-Modell-Analysen nicht berücksichtigte Items von <i>Figurale Analogien</i>	55
Tabelle 19: Ergebnisse der Likelihood-Ratio-Tests (<i>Figurale Analogien</i> , alle Items).....	56
Tabelle 20: Ergebnisse des z -Tests für das Teilungskriterium „Score“ (<i>Figurale Analogien</i> , alle Items)	57
Tabelle 21: Ergebnisse des z -Tests für das Teilungskriterium „Geschlecht“ (<i>Figurale Analogien</i> , alle Items)	59
Tabelle 22: Ergebnisse des z -Tests für das Teilungskriterium „Schultyp“ (<i>Figurale Analogien</i> , alle Items)	60

Tabelle 23: Ergebnisse der Likelihood-Ratio-Tests (<i>Figurale Analogien</i> , nach Ausschluss von Item 2)	62
Tabelle 24: Ergebnisse der Likelihood-Ratio-Tests (<i>Figurale Analogien</i> , nach Ausschluss der Items 2 und 7).....	62
Tabelle 25: Ergebnisse der Likelihood-Ratio-Tests (<i>Figurale Analogien</i> , nach Ausschluss der Items 2, 7 und 24)	63
Tabelle 26: Ergebnisse der Likelihood-Ratio-Tests (<i>Figurale Analogien</i> , nach Ausschluss der Items 2, 7, 24 und 5)	63
Tabelle 27: Ergebnisse der Likelihood-Ratio-Tests (<i>Figurale Analogien</i> , nach Ausschluss der Items 2, 7, 24, 5 und 23)	64
Tabelle 28: Ergebnisse der Likelihood-Ratio-Tests (<i>Gleichungen</i> , alle Items).....	67
Tabelle 29: Ergebnis der Hauptkomponentenanalyse: Eigenwerte und erklärter Varianzanteil pro Faktor	70
Tabelle 30: Ladungen und Kommunalitäten der Generalfaktorlösung	71
Tabelle 31: Ladungen und Kommunalitäten der Varimax-rotierten Zwei-Faktoren-Lösung.....	72
Tabelle 32: Interkorrelationsmatrix der Reasoning-Tests	74

I. EINLEITUNG

Die Fähigkeit zum logisch-schlussfolgernden Denken, in der Fachliteratur kurz als *Reasoning* bezeichnet, zählt zu den wichtigsten Leistungsbereichen der menschlichen Intelligenz. In allen gängigen Intelligenz-Testbatterien spielen Untertests, die das Konstrukt Reasoning erfassen, eine wichtige Rolle, und auch außerhalb solcher Testbatterien existieren unzählige Reasoning-Tests, da die Überprüfung dieses Fähigkeitsbereichs bei vielen psychologisch-diagnostischen Fragestellungen von Relevanz ist. So werden beispielsweise im Rahmen von Eignungstests oder Auswahlverfahren häufig auch Tests zur Erfassung des schlussfolgernden Denkens eingesetzt, da angenommen wird, dass bestimmte Reasoning-Tests eine Abschätzung des allgemeinen Intelligenzniveaus eines Menschen ermöglichen. Aus diesem Grund ist es von großer Bedeutung, dass die zum Einsatz kommenden psychologisch-diagnostischen Verfahren ausreichend evaluiert und validiert wurden, um zuverlässige Aussagen über die Fähigkeit einer Person tätigen und darauf aufbauend gute Entscheidungen treffen zu können.

Am Arbeitsbereich Psychologische Diagnostik, Institut für Angewandte Psychologie: Gesundheit, Entwicklung und Förderung, der Fakultät für Psychologie der Universität Wien wurden unter der Leitung von Herrn Univ.-Prof. Mag. Dr. Klaus Kubinger in den letzten Jahren einige Reasoning-Tests entwickelt, die den Ansprüchen an eine dem aktuellen Stand der Wissenschaft entsprechende Diagnostik genügen sollen, d. h. vor allem den Kriterien der Modernen Testtheorie. In diesem Zusammenhang wurden einerseits Verfahren konstruiert, die Reasoning im Bereich der fluiden Intelligenz nach Cattell (z. B. 1943, 1963) erfassen sollen, und andererseits solche, die dem kristallisierten Intelligenzbereich sensu Cattell zuzuordnen sind. Sechs dieser Tests wurden im Rahmen der vorliegenden Diplomarbeit einer Validierung unterzogen, wobei sie vor allem auch dahingehend überprüft wurden, ob sie mit dem dichotomen logistischen Modell von Rasch konform sind und somit eindimensional die entsprechende Fähigkeit erfassen. Des Weiteren wurde eine faktorenanalytische Überprüfung durchgeführt.

Im theoretischen Teil dieser Arbeit wird zunächst Bezug auf die Intelligenz- und testtheoretischen Grundlagen genommen, die für die vorliegende Untersuchung von Bedeutung sind. Anschließend wird im empirischen Teil auf das Ziel dieser Diplomarbeit, die zu evaluierende Reasoning-Test-Batterie sowie die empirische Untersuchung und deren Ergebnisse eingegangen.

II. THEORETISCHER TEIL

1. Intelligenztheoretische Grundlagen

Da die Fähigkeit zum logisch-schlussfolgernden Denken als wesentlicher Bestandteil der menschlichen Intelligenz gilt (vgl. z. B. Beckmann & Guthke, 1999; Guthke, 2003; Waldmann & Weinert, 1990), sollen in diesem Kapitel einige ausgewählte, im Zusammenhang mit dieser Arbeit wichtig erscheinende Intelligenztheorien sowie die Einbettung der Reasoning-Fähigkeit im Rahmen dieser Modelle beschrieben werden.

1.1. Die Zwei-Faktoren-Theorie von Spearman

Charles Spearman, der als Begründer der faktorenanalytischen Intelligenzforschung gilt, postulierte in seiner *Zwei-Faktoren-Theorie der Intelligenz* (Spearman, 1904, 1914, 1927) einen allgemeinen Intelligenzfaktor „g“, der allen intellektuellen Fähigkeiten zugrunde liege und durch den sich Korrelationen zwischen verschiedenen Intelligenztests erklären ließen. Dabei sei dieser Faktor in unterschiedlichem Ausmaß an verschiedenen Intelligenzleistungen beteiligt. Neben diesem g-Faktor spiele bei jedem Test auch noch eine spezifische Komponente „s“ eine Rolle, die nur in diesem einen Test enthalten sei. Intelligenz im Sinne dieser Theorie entspricht also einer einzigen Fähigkeitsdimension, die auch als „Allgemeine Intelligenz“ oder „geistige Energie“ bezeichnet werden kann (vgl. dazu auch Conrad, 1983; Guilford & Hoepfner, 1976; Stemmler, Hagemann, Amelang, & Bartussek, 2011; Süß, 2003).

Laut Spearman (1927; 1946, zitiert nach Beckmann & Guthke, 1999, S. 6) weisen Tests, die die Fähigkeit zum induktiven Denken erfassen bzw. auf das Erkennen von Beziehungen und das Entdecken von Regeln abzielen, eine hohe Ladung auf diesem Generalfaktor der Intelligenz auf. Somit kann Reasoning als eine wesentlich an der allgemeinen Intelligenz sensu Spearman beteiligte Fähigkeit bezeichnet werden.

1.2. Die Primary Mental Abilities nach Thurstone

Basierend auf der Multiplen Faktorenanalyse entwickelte L. L. Thurstone (1932, zitiert nach Thurstone, 1961; 1938 bzw. gemeinsam mit T. G. Thurstone, 1941, beide zitiert nach Stemmler et al., 2011, S. 148 ff.) das *Modell mehrerer gemeinsamer Faktoren*, demzufolge sich Intelligenz aus sieben verschiedenen sogenannten Primärfähigkeiten („*primary abilities*“) zusammensetzt (Stemmler et al., 2011), die jeweils in unterschiedlichem Ausmaß und in unterschiedlicher Konstellation bei intelligenten Leistungen in verschiedenartigen Testaufgaben zum Tragen kommen und gleichwertig nebeneinander bestehen. Diese sieben Einzelfähigkeiten werden wie folgt bezeichnet (vgl. dazu Conrad, 1983; Jäger, 1973; 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)*: unmittelbare Merkfähigkeit
- *P (perceptual speed)*: Auffassungs- bzw. Wahrnehmungsgeschwindigkeit
- *I (induction)* oder *R (reasoning)*: Fähigkeit zum logisch-schlussfolgernden Denken bzw. zum Auffinden und Anwenden von Regeln

Im Rahmen dieses Intelligenzmodells wird Reasoning also von Thurstone als wichtiger Leistungsbereich innerhalb der menschlichen Intelligenz angesehen, der als gleichwertig neben den anderen sechs Primärfähigkeiten existierend zu verstehen ist.

1.3. Das Modell der fluiden und kristallisierten Intelligenz von Cattell

Beginnend im Jahr 1941 entwickelte R. B. Cattell ein Intelligenzmodell, nach dem die allgemeine Intelligenz des Menschen aus mehreren Faktoren besteht, die nicht ganz unabhängig voneinander sind und von denen die wichtigsten beiden als fluide (g_f) und kristallisierte Intelligenz (g_c) bezeichnet werden (Cattell, 1943, 1963, sowie 1957, zitiert nach Cattell, 1961; Horn, 1968; Horn & Cattell, 1966, 1982).

Zum Bereich der fluiden Intelligenz gehören Fähigkeiten, die es einem ermöglichen, sich neuartigen Situationen und Problemen anzupassen und adäquat mit diesen umzugehen. Diese kognitiven Leistungen betreffen die Basisprozesse des Denkens und Problemlösens, die zum Beispiel beim Erkennen von Beziehungen und Zusammenhängen eine Rolle spielen (Cattell, 1943, 1963; vgl. auch Süß, 2003). Sie werden als vorwiegend biologisch und genetisch determiniert angesehen, wodurch sie weitgehend unabhängig von früheren Lernerfahrungen sind (Cattell, 1957, zitiert nach Cattell, 1961; Cattell, 1963). Die fluide Intelligenz wird beispielsweise mit Tests oder Aufgaben erfasst, in denen Beziehungen und Regeln erkannt und angewandt werden müssen und dem induktiven und deduktiven Denken (siehe dazu Kapitel 2) zugeordnet werden können (Süß, 2003). Derartige Intelligenztestaufgaben werden häufig auch als „culture-fair“ oder „culture-free“ bezeichnet, da sie entweder abstraktes, neuartiges Material oder solches, mit dem Mitglieder verschiedener Kulturen gleichermaßen vertraut sind, verwenden und zudem meist zeitlich beschränkt sind. Daher werden sie als sprachfrei und kulturunabhängig angesehen (Cattell, 1957, zitiert nach Cattell, 1961; Cattell, 1963; vgl. auch Stemmler et al., 2011; Süß, 2003). Nach den Arbeiten von Cattell (1943, 1963, und 1957, zitiert nach Cattell, 1961) sowie von Horn (1968) entwickelt sich die fluide Intelligenz rascher im Laufe des Lebens und ist einem stärkeren Abbau über die Lebensspanne unterworfen, der schon im frühen Erwachsenenalter beginnt.

Die kristallisierte Intelligenz repräsentiert kognitive Fertigkeiten, in denen sich frühere Lernerfahrungen kumuliert und verfestigt haben. Hierunter fallen also beispielsweise durch Sozialisation und Schulbildung erworbenes Wissen sowie Kenntnisse, die auf vergangenen Erfahrungen beruhen. Damit wird auch deutlich, dass kristallisierte Intelligenzfähigkeiten kulturabhängig sind und beispielsweise mithilfe von Verfahren erfasst werden können, in denen Wortschatz, Sprachverständnis oder auch Allgemeinwissen gefragt ist (Horn, 1968; Horn & Cattell, 1982; vgl. auch Stemmler et al., 2011). Im Gegensatz zur fluiden Intelligenz erreicht die kristallisierte Intelligenz ihren Höhepunkt erst später im Leben und zeigt über die Lebensspanne hinweg ein relativ gleichbleibendes oder sogar ansteigendes Niveau (Cattell, 1957, zitiert nach Cattell, 1961; Cattell, 1963; Horn, 1968; Horn & Cattell, 1982).

Im Zusammenhang mit Cattells Theorie kann Reasoning also grundsätzlich, sofern die entsprechenden Testaufgaben kein sprachliches oder kulturabhängiges Material beinhalten, in den Bereich der fluiden Intelligenz eingeordnet werden. Wie in späteren Ab-

schnitten dieser Arbeit zu sehen sein wird, kann die Fähigkeit zum logisch-schlussfolgernden Denken aber auch mithilfe von Verfahren erfasst werden, in denen Allgemeinwissen oder andere erworbene Fertigkeiten bei der Lösung von Aufgaben, in denen Regeln oder Beziehungen erkannt werden müssen, zum Tragen kommen.

1.4. Das Berliner Intelligenzstrukturmodell von Jäger

Das *Berliner Intelligenzstrukturmodell* wurde von Jäger (1982, zitiert nach Süß, 2003, S. 219 f.) entwickelt und integriert verschiedene Strukturmodelle der Intelligenz zu einem Gesamtmodell. Jägers Modell ist ein hierarchisches Strukturmodell, wobei an der Spitze ein g-Faktor sensu Spearman (1904) steht, dem insgesamt sieben Fähigkeitsdimensionen untergeordnet sind, von denen vier als „Operationen“ und drei als „Inhalte“ bezeichnet werden. Zu den operativen Fähigkeiten zählen „Verarbeitungskapazität“, „Einfallsreichtum“, „Merkfähigkeit“ und „Bearbeitungsgeschwindigkeit“, während unter inhaltsgebundenen Fähigkeiten „sprachgebundenes“, „zahlengebundenes“ und „anschauungsgebundenes Denken“ verstanden wird (vgl. Stemmler et al., 2011, S. 160). Auf der untersten Ebene der Hierarchie resultieren also bei Kombination aller Operationen mit allen Inhalten insgesamt zwölf spezifische Fähigkeitsbereiche (z. B. Merkfähigkeit im figural-bildhaften Bereich).

Ordnet man das Konstrukt Reasoning auch in dieses Intelligenzmodell ein, so wäre diese Fähigkeit hier mit der Operation „Verarbeitungskapazität“ – von Jäger (1973) auch noch als „formallogisches Denken und Urteilsfähigkeit“ bezeichnet – gleichzusetzen. Diese Funktionseinheit spielt laut dem Autor bei solchen Aufgaben eine Rolle, bei denen Informationen beispielsweise zueinander in Beziehung gesetzt oder beurteilt werden müssen, und ist als zentraler Fähigkeitsbereich im Rahmen der menschlichen Intelligenz zu verstehen. Im Sinne dieser Theorie sind außerdem drei verschiedene Arten von schlussfolgerndem Denken zu unterscheiden, nämlich Reasoning im numerischen Bereich, im Zusammenhang mit verbal-lexikalischen Inhalten sowie mit figural-bildhaftem Material.

2. Das Konstrukt *Reasoning*

In diesem Abschnitt soll das Konstrukt *Reasoning* näher betrachtet werden. Zu diesem Zweck werden der Begriff definiert und die Formen des logisch-schlussfolgernden Denkens kurz erläutert, die in der Fachliteratur unterschieden werden.

Laut Kubinger (2009b) ist Reasoning „die Fähigkeit, Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge erkennen und zweckentsprechend verwerten zu können“ (S. 206).

Logisch-schlussfolgerndes Denken kann im Wesentlichen in zwei Arten oder Bereiche unterteilt werden, nämlich in induktives und deduktives Denken (vgl. z. B. Klauer, 2001a; Waldmann & Weinert, 1990). *Induktives Denken* beschreibt laut Waldmann und Weinert (1990) jene „Prozesse, die ein Individuum dazu befähigen, aus spezifischen Erfahrungen allgemeine Regeln, Muster, Konzepte oder Gesetzmäßigkeiten abzuleiten und auf neue Ereignisse anzuwenden“ (S. 25–26). Auch Beckmann und Guthke (1999) führen eine ähnliche Definition für Induktion an und sehen in diesem Konstrukt ein Schließen von einzelnen Fakten auf allgemeine Regeln, wobei sie betonen, dass die durch induktives Denken gewonnenen Erkenntnisse immer nur wahrscheinlich sind, nicht aber mit Sicherheit als wahr angesehen werden können.

In ähnlicher Weise definiert Klauer (2001b) „induktives Denken“ als jenes „Denken, das zur Entdeckung von Regelhaftigkeiten oder Gesetzmäßigkeiten führt“ (S. 168) und grenzt diese Prozesse insofern vom „induktiven Schließen“ ab, als Ersteres nicht über das im Hier und Jetzt Beobachteten hinausgeht und Letzteres sehr wohl auch Verallgemeinerungen auf eine Grundgesamtheit enthält. Beim induktiven Denken lassen sich laut Klauer (2001b, S. 169) verschiedene Varianten anhand des in Abbildung 1 dargestellten Schemas unterscheiden. Bei Kombination aller Ausprägungen der drei Facetten ergeben sich insgesamt 30 Formen des induktiven Denkens, wobei sechs davon, nämlich jene, die bei Kombination der Kernfacetten A und B zustande kommen, von Klauer (2001b) namentlich benannt wurden, beispielsweise mit „Diskrimination“ für die Feststellung der Verschiedenheit von Merkmalen oder „Beziehungserfassung“ für das Erkennen einer Gleichheit von Relationen (S. 170).

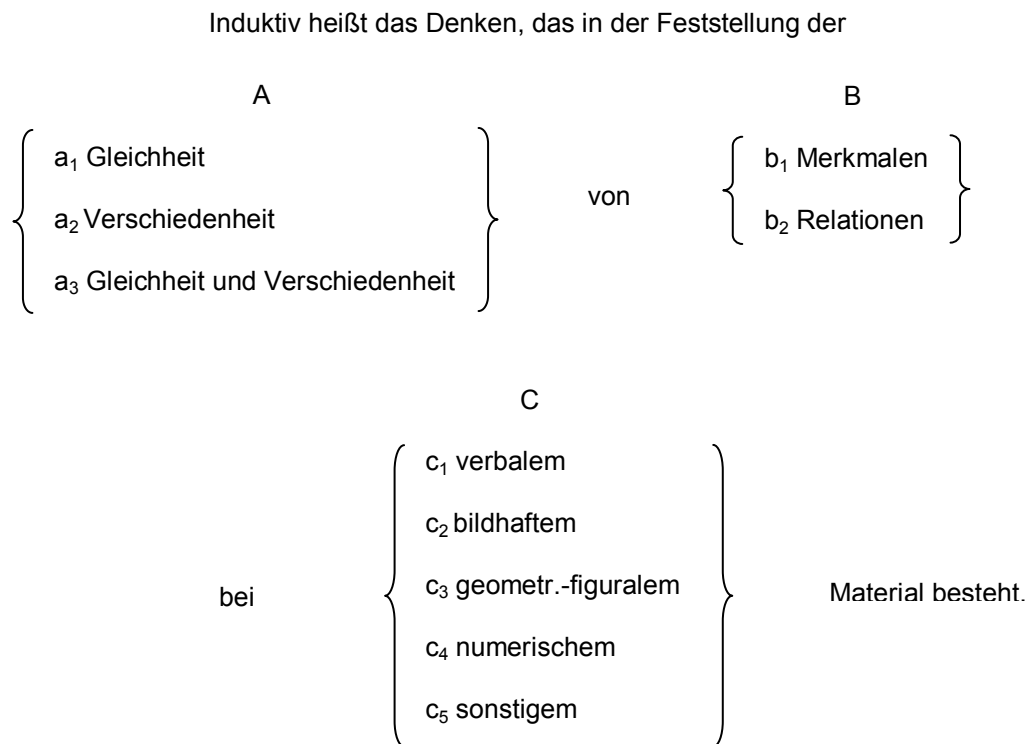


Abbildung 1: Definition des induktiven Denkens (nach Klauer, 2001b, S. 169)

Im Gegensatz zum induktiven Denken ist das *deduktive Denken* dadurch gekennzeichnet, dass man ausgehend von Prämissen unter Anwendung logischer Regeln zu bestimmten Schlussfolgerungen gelangt, wobei die Regeln so beschaffen sind, dass bei Gültigkeit der Prämissen auch die Gültigkeit bzw. Wahrheit der Schlüsse gesichert ist, und zwar unabhängig davon, ob der Inhalt der Prämissen auch realistisch und sinnvoll ist. Allein die richtige Anwendung der Regeln der Logik garantiert die Korrektheit der Schlussfolgerungen (Beckmann & Guthke, 1999; Meiser & Klauer, 2001; Waldmann & Weinert, 1990). Wichtig bei deduktiven Schlüssen ist vor allem auch, dass die Prämisseninformation bei der Konklusion nicht erweitert wird (Beckmann & Guthke, 1999).

Als typischer Repräsentant von Reasoning-Aufgaben, die auf die Erfassung des deduktiven Denkens abzielen, sind Syllogismen zu nennen (z. B. Waldmann & Weinert, 1990). Ein Syllogismus besteht aus zwei Prämissen, die Quantoren wie „alle“ oder „einige“ enthalten, und einer Schlussfolgerung, wobei bei derartigen Testauf-

gaben die beiden Prämissen vorgegeben sind und korrekte Konklusionen abgeleitet werden müssen (vgl. Kubinger, 2009b; Waldmann & Weinert, 1990).

Psychologisch-diagnostische Verfahren, die dem induktiven Denken zugeordnet werden, sind zum Beispiel Analogieaufgaben, Reihenfortsetzungsaufgaben, Klassifikationsaufgaben oder Matrizentests. Gemeinsam ist diesen Aufgabentypen, dass aus den jeweils vorgegebenen Elementen Regeln zu erkennen sind und unter Anwendung dieser zur korrekten Lösung gelangt werden muss (vgl. Klauer, 2001b; Waldmann & Weinert, 1990). Nach Beckmann und Guthke (1999) lassen sich induktives und deduktives Denken jedoch nicht eindeutig trennen; die Autoren weisen darauf hin, dass „jede Deduktion das Resultat einer vorausgegangenen Induktion [ist und] das Resultat induktiven Denkens . . . sich letztlich über einen deduktiven Denkakt [manifestiert]“ (S. 6). Somit ist es gerade bei Aufgaben zum induktiven Denken meist der Fall, dass die zu erkennenden Regeln auch auf das jeweils fehlende Element bzw. das nächste Item im Sinne einer Deduktion anzuwenden sind und eine strikte Unterscheidung der beiden Formen des schlussfolgernden Denkens in Hinblick auf die mit einem Verfahren erfasste Fähigkeit nicht immer sinnvoll und eindeutig möglich ist (vgl. auch Waldmann & Weinert, 1990).

Wie bereits in Abschnitt 1.3 erwähnt, wird schlussfolgerndes Denken traditionell dem fluiden Intelligenzbereich sensu Cattell zugeordnet, da bei Reasoning-Aufgaben üblicherweise entweder sprachfreies, d. h. figural-bildhaftes oder numerisches, Material (vgl. z. B. den *Wiener Matrizen-Test 2* von Formann, Waldherr, & Piswanger, 2011, oder den Subtest *Zahlenreihen* im *Intelligenz-Struktur-Test 2000 R* von Liepmann, Beauducel, Brocke, & Amthauer, 2007) oder aber wie im Falle von Syllogismen zwar sprachliches Material zur Anwendung kommt, jedoch hierbei kein Wissen über die in den Items vorkommenden Inhalte vonnöten ist, da die Konklusion rein auf der Basis logischer Regeln zustande kommen soll und kann (vgl. z. B. den Test *Syllogismen* von Srp, 1994). Dennoch sind einige Reasoning-Tests dem Bereich der kristallisierten Intelligenz sensu Cattell zuzuordnen, nämlich solche, bei denen Allgemeinwissen, sprachliche Kenntnisse oder sonstige im Laufe des Lebens erworbene Fertigkeiten zum Zwecke des Ableitens und Anwendens von Regeln oder Beziehungen eingesetzt werden müssen (vgl. z. B. den Untertest *Analogien* im *I-S-T 2000 R*, Liepmann et al., 2007; oder das im Rahmen dieser Diplomarbeit untersuchte Verfahren *Figurale Analogien* von Ünal & Kubinger, in Vorb.). In diesem Sinne kann kristallisiertes schlussfolgerndes Denken mit Ünal (2014) definiert

werden als die Fähigkeit, „in einem bestimmten Kulturkreis erworbenes Wissen durch das Erkennen von Zusammenhängen und die Anwendung der erfassten Regeln einsetzen zu können“ (S. 13).

3. Testtheoretische Grundlagen

Nachdem im vorhergehenden Abschnitt auf die intelligenztheoretischen Hintergründe eingegangen wurde, sollen in diesem Kapitel die testtheoretischen Grundlagen, auf denen diese Arbeit basiert, etwas näher beleuchtet werden. Dabei wird zunächst das Gütekriterium Validität behandelt und im Anschluss ein Überblick über das dichotome logistische Testmodell von Rasch sowie die Methode der Faktorenanalyse gegeben.

3.1. Das Gütekriterium Validität

Unter dem Gütekriterium *Validität* oder Gültigkeit eines psychologisch-diagnostischen Verfahrens versteht man nach Lienert und Raatz (1998) „den Grad der Genauigkeit . . . , mit dem [dieses Verfahren] dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) [es] messen oder vorhersagen soll, tatsächlich mißt oder vorhersagt“ (S. 10). Von Kubinger (2009b) kurz und prägnant formuliert bedeutet die Validität eines Tests, „dass er tatsächlich jenes psychische Merkmal misst, welches er zu messen behauptet“ (S. 55).

Grundsätzlich können verschiedene Arten oder Aspekte der Validität unterschieden werden (Kubinger, 2003a, 2009b; Lienert & Raatz, 1998):

- **Inhaltliche Gültigkeit (Kontentvalidität):** Ein Verfahren ist inhaltlich gültig, wenn es das zu erfassende Merkmal repräsentiert bzw. selbst das optimale Kriterium für das Merkmal ist (z. B. eine Arbeitsprobe).
- **Konstruktvalidität:** Diese ist dann gegeben, wenn ein Test gewisse theoretische Ansprüche in Hinblick auf ein bestimmtes Konstrukt erfüllt und er dieses Konstrukt (wie z. B. die Fähigkeit zum schlussfolgernden Denken) nachweislich erfasst.

- **Kriteriumsvalidität:** Diese kennzeichnet das Ausmaß, in dem ein Verfahren mit einem anderen Test, der dasselbe Merkmal erfasst, korreliert (*Übereinstimmungsvalidität*) bzw. in dem es ein als relevant angesehenes in der Zukunft liegendes Außenkriterium vorhersagen kann (*prognostische Validität, Vorhersagegültigkeit*).

Im Rahmen dieser Arbeit geht es vornehmlich um die Konstruktvalidierung der zu untersuchenden Verfahren, also darum, ob mit den unterschiedlichen Tests tatsächlich das Konstrukt Reasoning erfasst wird und ob diese auch dem theoretisch zugrundeliegenden Modell, nach dem die Verfahren entweder dem fluiden oder dem kristallisierten Intelligenzbereich zugeordnet sind, entsprechen.

Bei der Konstruktvalidität ist nach dem Ansatz der sogenannten „*multitrait-multimethod matrix*“ (*MTMM-Matrix*) von Campbell und Fiske (1959) zwischen konvergenter und diskriminanter Validität zu unterscheiden. Von *konvergenter Validität* spricht man, wenn ein Test mit einem anderen Test korreliert, der dasselbe Konstrukt mit einer anderen Methode bzw. einem anderen Verfahren erfasst, d. h. wenn eine Korrelation zwischen konstruktnahen Tests besteht. *Diskriminante Validität* hingegen bedeutet, dass ein Verfahren *keinen* Zusammenhang mit Tests aufweist, die ein anderes Merkmal als der zu evaluierende Test messen, selbst wenn die betreffenden Fähigkeiten oder Eigenschaften mit derselben Methode erfasst werden; d. h. es ist keine Korrelation mit konstruktfernen Tests zu beobachten. Entsprechend diesem Konzept sollten zwei psychologisch-diagnostische Verfahren, die dasselbe Konstrukt mit unterschiedlichen Erhebungsmethoden erfassen (z. B. jeweils die Fähigkeit Reasoning, einmal mit figuralem und einmal mit numerischem Material), wesentlich stärker miteinander korrelieren als zwei Verfahren, die mithilfe derselben Methode jeweils unterschiedliche Merkmale erfassen (z. B. ein Reasoning-Test und ein Merkfähigkeitstest, die beide bildhaftes Material verwenden), und natürlich auch stärker als zwei Verfahren, die nicht nur verschiedene Konstrukte erfassen, sondern dies auch noch auf unterschiedliche Art und Weise tun (z. B. ein numerischer Reasoning-Test und ein figuraler Merkfähigkeitstest) (vgl. auch Kubinger, 2003a, 2009b; Schermelleh-Engel & Schweizer, 2003, 2012).

Neben diesem Ansatz gilt unter anderem auch die Faktorenanalyse, deren Idee und Prinzip in Abschnitt 3.3 noch genauer beschrieben wird, als Methode der Konstruktvali-

dierung, genauer gesagt als Verfahren zur Bestimmung der sogenannten *faktoriellen Validität* (Lienert & Raatz, 1998).

Was das Validierungskonzept der vorliegenden Arbeit betrifft, so handelt es sich um die Überprüfung der Konstruktvalidität, einerseits mithilfe der exploratorischen Faktorenanalyse und andererseits auf Basis der Berechnung der Korrelationen zwischen den verschiedenen Reasoning-Tests, die konzeptionell dem fluiden oder kristallisierten Intelligenzbereich sensu Cattell zugeordnet sind, wodurch die konvergente und diskriminante Validität festgestellt werden soll.

3.2. Das dichotome logistische Testmodell von Rasch

Das dichotome logistische Modell von Rasch (kurz: Rasch-Modell) gilt als das Grundmodell der Item-Response-Theorie (IRT). Die Idee dieser Theorie, die auch als Probabilistische Testtheorie – im Gegensatz zur deterministischen oder Klassischen Testtheorie – bezeichnet wird, besteht in der Annahme, dass den beobachtbaren oder manifesten Reaktionen einer Testperson eine bestimmte nicht beobachtbare, sog. latente Eigenschaft zugrunde liegt. Die Beziehung zwischen diesen beiden kann durch eine Wahrscheinlichkeitsfunktion dargestellt werden (Kubinger, 1989, 2003b; Kubinger, Rasch, & Yanagida, 2011; Moosbrugger, 2012).

In Formel (3.1) ist die Wahrscheinlichkeitsfunktion aufgeführt, auf der das Rasch-Modell basiert. Hiermit wird die Wahrscheinlichkeit beschrieben, mit der eine Testperson v ein bestimmtes Item i löst („+“), in Abhängigkeit von der Fähigkeit der Person (= Personenparameter ξ_v) und der Schwierigkeit des Items (= Itemparameter σ_i). Die Komplementärwahrscheinlichkeit sagt entsprechend aus, wie wahrscheinlich es ist, dass Testperson v Item i nicht löst („–“). Entsprechend dieser Funktion ist die Wahrscheinlichkeit, dass eine Aufgabe gelöst wird, also umso höher, je größer die Personenfähigkeit und je kleiner die Itemschwierigkeit ist. Sind die beiden Parameter ξ_v und σ_i gleich groß, so beträgt die Lösungswahrscheinlichkeit 0,5 (Kubinger, 1989, 2003b, 2009b; Kubinger et al., 2011).

$$P(+|\xi_v, \sigma_i) = \frac{e^{\xi_v - \sigma_i}}{1 + e^{\xi_v - \sigma_i}} \quad (3.1)$$

Das Rasch-Modell kann, wie aus den bisherigen Ausführungen hervorgeht, nur bei dichotomen Daten angewendet werden, d. h. die Items müssen jeweils als „gelöst“ (+ bzw. 1) oder „nicht gelöst“ (– bzw. 0) verrechnet werden.

Charakteristisch für das Rasch-Modell ist, dass sowohl die interessierende Fähigkeit der Testperson als auch der Schwierigkeitsgrad eines Items durch jeweils einen einzigen Parameter ausgedrückt werden, d. h. beide werden *eindimensional* gemessen, was bedeutet, dass durch sie verschiedene Ausprägungsgrade auf jeweils derselben latenten Dimension dargestellt werden. Folglich erfassen alle Items eines bestimmten Tests oder Subtests eine einzige, und zwar dieselbe zugrunde liegende Fähigkeit oder Eigenschaft. Daraus ergibt sich auch, dass die Anzahl der gelösten Aufgaben, also der Rohwert oder Score, eine *erschöpfende Statistik* für die zu erfassende Fähigkeit darstellt, sodass die gesamte relevante Information über die Fähigkeit einer Person durch den Rohwert ausgedrückt werden kann. Dabei spielt es keine Rolle, welche Items aus dem vorgegebenen Itempool gelöst und welche nicht gelöst wurden (Kubinger, 1989, 2003b; Kubinger et al., 2011). Dies spiegelt bereits eine weitere Annahme des Rasch-Modells wider, nämlich die der *lokalen stochastischen Unabhängigkeit*: Ob eine bestimmte Person ein bestimmtes Item löst oder nicht, hängt – abgesehen vom Zufall – nur von ihrer Fähigkeit und der Schwierigkeit des Items ab, nicht aber davon, welche anderen Aufgaben sie zuvor bereits gelöst hat oder noch lösen wird. Dies bedeutet, dass die Lösung vorangegangener Aufgaben keine Voraussetzung für die Bearbeitung aller weiteren Items ist (Kubinger, 1989, 2003b, 2005, 2009b).

Damit also die Anzahl gelöster Aufgaben als fairer Testwert für die Fähigkeit einer Person verrechnet werden kann und garantiert ist, dass dieser die empirisch feststellbaren Verhaltensrelationen zwischen verschiedenen Personen adäquat abbildet und somit das Gütekriterium *Skalierung* erfüllt ist (Kubinger, 2009b), muss daher *notwendigerweise* das Rasch-Modell gelten. Sind die Voraussetzungen des Rasch-Modells nicht erfüllt und ist es für die Bestimmung des Fähigkeitsniveaus einer Person nicht gleichgültig, welche Aufgaben gelöst und welche nicht gelöst wurden, so kann der Rohwert nicht als faires Maß für die erbrachte Testleistung angesehen werden (Kubinger, 1989, 2003b, 2005, 2009b; Kubinger et al., 2011; für den mathematischen Beweis dieses Gesetzes siehe z. B. Fischer, 1974).

Eine weitere Besonderheit des Rasch-Modells besteht darin, dass es *spezifisch objektive* Vergleiche ermöglicht. Das bedeutet, dass bei Geltung des Rasch-Modells einerseits zwei Personen direkt anhand ihrer Fähigkeitsparameter ξ verglichen werden können, unabhängig davon, welche Aufgaben eines Tests sie jeweils bearbeitet haben. Andererseits ist auch ein Vergleich zweier Items bezüglich ihrer Schwierigkeiten σ unabhängig davon möglich, welche Stichprobe zur Ermittlung der Itemparameter herangezogen wurde, d. h. es besteht hierbei *Stichprobenunabhängigkeit* (Kubinger, 1989, 2003b, 2009b; Kubinger et al., 2011).

Einen besonderen Vorteil, der sich aus den soeben besprochenen Charakteristika des Rasch-Modells ergibt, stellt dessen Prüfbarkeit mithilfe von Modelltests dar. Sofern für die Items eines psychologisch-diagnostischen Verfahrens das Rasch-Modell gilt, müssen die Schätzungen der Itemschwierigkeitsparameter in verschiedenen Teilstichproben statistisch gleich sein. Per Umkehrschluss folgt somit, dass das Rasch-Modell nicht gilt, wenn für mindestens ein Item des Itempools die Parameterschätzungen in zwei willkürlich getrennten Teilstichproben der Gesamtstichprobe nicht gleich sind (Kubinger, 1989, 2003b, 2009b; Kubinger et al., 2011). Als inferenzstatistisches Verfahren kann für die Prüfung der Geltung des Rasch-Modells der Likelihood-Ratio-Test nach Andersen (1973) herangezogen werden. Dieser überprüft die empirischen Daten dahingehend, ob diese durch die geschätzten Parameter der beiden Subgruppen besser beschrieben werden können als durch jene der gesamten Stichprobe (Kubinger, 1989, 2003b; vgl. auch Kubinger et al., 2011). Als sogenanntes *internes* Teilungskriterium wird zur Teilung der Gesamtstichprobe in zwei Gruppen üblicherweise die Anzahl gelöster Aufgaben verwendet. Als weitere (*externe*) Teilungskriterien werden beispielsweise Geschlecht, Alter oder Muttersprache der Testpersonen herangezogen (Kubinger, 1989, 2003b; Kubinger et al., 2011).

Als Methode der Beschreibenden Statistik kann zur Überprüfung der Geltung des Rasch-Modells eine grafische Modellkontrolle durchgeführt werden. Hierbei werden die Parameterschätzungen für alle Items eines Tests für die jeweiligen zwei Teilstichproben in einem rechtwinkligen Koordinatensystem gegeneinander aufgetragen. Im Idealfall liegen alle Items auf einer 45°-Geraden, die durch den Ursprung verläuft, da dies bedeuten würde, dass die Itemschwierigkeiten in beiden Subgruppen identisch sind. Wenn einzelne oder alle Items deutlich von der 45°-Geraden entfernt liegen, so ist keine Stichprobenunabhängigkeit gegeben und die Items erfassen nicht eindimensional dasselbe Konstrukt.

In diesem Fall werden die nicht modellkonformen Aufgaben aus dem Itempool entfernt (Kubinger, 1989, 2003b, 2009b; Kubinger et al., 2011). Zur Beurteilung der Modellgeltung der einzelnen Items kann zusätzlich noch der z-Test nach Fischer und Scheiblechner (1970, zitiert nach Kubinger et al., 2011, S. 556 f.) verwendet werden (siehe auch Kubinger, 1989; Kubinger et al., 2011, S. 556 f.).

3.3. Die exploratorische Faktorenanalyse

Die Faktorenanalyse ist wohl eine der am häufigsten in der Psychologie eingesetzten Methoden zur Konstruktion und Validierung psychologisch-diagnostischer Verfahren sowie zu Zwecken der Grundlagenforschung, vor allem im Bereich der Differentiellen und Persönlichkeitspsychologie (Kubinger et al., 2011; Stemmler et al., 2011), und wurde erstmals von Spearman (1904) im Zusammenhang mit der Entwicklung seiner Intelligenztheorie beschrieben (vgl. Stemmler et al., 2011; siehe auch Kapitel 1.1).

Die *exploratorische Faktorenanalyse* stellt ein hypothesengenerierendes Verfahren der multivariaten Statistik dar, das zur „Dimensionsanalyse“ (Kubinger et al., 2011, S. 530) dient. Das Ziel dieser Methode besteht darin, eine Vielzahl von Variablen hinsichtlich ihrer Gemeinsamkeiten zu gruppieren und so die Anzahl und Art der ihnen zugrunde liegenden gemeinsamen *Faktoren* (auch *Dimensionen* genannt) zu bestimmen. Sie geht von der Annahme aus, dass sich die Ausprägungen von Personen in bestimmten beobachtbaren Verhaltensweisen (z. B. die Antworten oder Reaktionen auf einzelne Items) auf latente, d. h. nicht direkt beobachtbare, Persönlichkeits- oder Fähigkeitsdimensionen zurückführen lassen. Auf Basis der wechselseitigen Zusammenhänge der beobachtbaren Merkmale werden mithilfe der Faktorenanalyse die dahinter liegenden latenten Faktoren ermittelt. In der Form einer Linearkombination kann sodann dargestellt werden, welchen Beitrag die einzelnen Faktoren zur Ausprägung in der beobachteten Variable leisten (Kubinger, 2009b; Kubinger et al., 2011; Moosbrugger & Hartig, 2003).

Der erste Schritt bei der exploratorischen Faktorenanalyse besteht in der Bestimmung der Anzahl der Faktoren. Zur Faktorenextraktion stehen unterschiedliche Methoden zur Verfügung (vgl. Moosbrugger & Hartig, 2003; Moosbrugger & Schermelleh-Engel, 2012), wobei in der vorliegenden Arbeit die *Hauptkomponentenanalyse* zur An-

wendung kommt und deshalb hier nur auf diese kurz eingegangen werden soll. Bei der Hauptkomponentenanalyse (engl. principal components analysis, PCA) wird davon ausgegangen, dass die gesamte Varianz der einzelnen Variablen durch die gemeinsamen, zugrunde liegenden Faktoren – in diesem Fall auch *Hauptkomponenten* genannt – erklärt wird. Aus diesem Grund entspricht die Zahl der extrahierten Faktoren zu Beginn der Anzahl der untersuchten Variablen (Moosbrugger & Hartig, 2003; Moosbrugger & Schermelleh-Engel, 2012; vgl. auch Kubinger et al., 2011). Auf Basis mehrerer möglicher Kriterien kann nun die Anzahl der Faktoren festgelegt werden, die als hinreichend zur Erklärung der unterschiedlichen Ausprägungen bei verschiedenen Personen in den beobachteten Merkmalen bzw. zur Erklärung der Interkorrelationen zwischen den manifesten Variablen angenommen wird. Häufig zur Anwendung kommen hierbei das sogenannte *Kaiser-Kriterium* oder der *Scree-Test* (vgl. Moosbrugger & Hartig, 2003; Moosbrugger & Schermelleh-Engel, 2012).

Beim erstgenannten Kriterium werden alle Faktoren als relevant betrachtet, die einen Eigenwert größer 1 aufweisen. Der *Eigenwert* eines Faktors gibt an, welcher Anteil der Varianz aller untersuchten Variablen durch diesen Faktor erklärt wird, wobei ein Eigenwert größer 1 im Falle von standardisierten Variablen (Varianz = 1) bedeutet, dass dieser Faktor mehr Varianz erklärt als jede einzelne Variable (Moosbrugger & Hartig, 2003; Moosbrugger & Schermelleh-Engel, 2012). Der Eigenwert errechnet sich aus der Summe der quadrierten Faktorladungen über alle beobachteten Merkmale, wobei man unter der *Ladung* die Korrelation zwischen dem latenten Faktor und der manifesten Variable versteht. Eine hohe Ladung einer Variablen auf einem Faktor bedeutet, dass der Faktor (das latente Konstrukt) einen starken Einfluss auf die Ausprägung der jeweiligen Variablen hat. Der Anteil der Varianz einer Variablen, der durch alle zugrunde liegenden Faktoren gemeinsam erklärt wird, wird als *Kommunalität* bezeichnet (Moosbrugger & Hartig, 2003; Moosbrugger & Schermelleh-Engel, 2012).

Beim *Scree-Test* werden die Eigenwerte der extrahierten Faktoren grafisch untersucht, wobei diese der Größe nach angeordnet und durch eine Linie verbunden sind. In diesem Eigenwerteverlauf zeigt sich meist ein deutlicher Knick und es werden all jene Faktoren als relevant erachtet, die vor diesem Knick liegen und damit einen deutlich höheren Eigenwert aufweisen als alle anderen (Moosbrugger & Hartig, 2003; Moosbrugger & Schermelleh-Engel, 2012). Aufgrund der verschiedenen Möglichkeiten zur Be-

stimmung der Faktorenzahl sind hier natürlich verschiedene Lösungen denkbar und das Ergebnis mitunter recht subjektiv.

Wurde nun die Anzahl der Faktoren festgelegt, so geht es im nächsten Schritt darum, diese inhaltlich zu beschreiben bzw. zu benennen. Zur besseren Interpretierbarkeit wird hierfür meist eine *Faktorenrotation* durchgeführt, deren Ziel es ist, ein Ladungsmuster zu erreichen, bei dem jede Variable auf nur einem einzigen Faktor hoch und auf allen übrigen sehr niedrig lädt (Prinzip der *Einfachstruktur*). Als Methoden zur Faktorenrotation können *orthogonale (rechtwinklige)* und *oblique (schiefwinklige) Rotationsverfahren* unterschieden werden. Während bei obliquen Verfahren die Unkorreliertheit, also die Unabhängigkeit der Faktoren, aufgegeben wird, wird diese bei orthogonalen Techniken beibehalten, sodass die Faktoren auch nach Rotation noch unabhängig voneinander interpretierbar sind. Das wichtigste orthogonale Rotationsverfahren ist die *Varimax-Rotation*, bei der die Varianz der quadrierten Faktorladungen innerhalb der Faktoren maximiert wird, sodass sich idealerweise ein Ladungsmuster pro Faktor mit einigen wenigen sehr hohen und sonst sehr niedrigen Ladungen ergibt (Moosbrugger & Hartig, 2003; Moosbrugger & Schermelleh-Engel, 2012; vgl. auch Kubinger et al., 2011).

Nachdem nun die theoretischen Grundlagen der vorliegenden Arbeit hinreichend erläutert wurden, wird im nachfolgenden empirischen Teil auf die Ziele der Untersuchung, das Untersuchungsdesign und die Ergebnisse eingegangen.

III. EMPIRISCHER TEIL

4. Ziel der Untersuchung und Hypothesen

Ziel der vorliegenden Diplomarbeit war es, einen Beitrag zur Validierung mehrerer in Entwicklung befindlicher Reasoning-Tests zu leisten. Bei den Verfahren handelt es sich um die folgenden, welche in Abschnitt 5 auch noch ausführlicher beschrieben werden:

- Untertest *Formale Folgerichtigkeit* aus AID-Gruppe (Kubinger & Hagenmüller, in Vorb.)
- *Sprachliche Analogien* (Kubinger & Ünal, in Vorb.)
- *Färbiger Matrizentest (FMT)* (Undeutsch, 2010a)¹
- *Numerische Topologien (N-TOP)* (Kubinger & Heuberger, in Vorb.)
- *Figurale Analogien* (Ünal & Kubinger, in Vorb.)
- *Gleichungen* (Kubinger & Gamsjäger, in Vorb.)

Im Vordergrund stand dabei deren Überprüfung auf Konformität mit dem dichotomen logistischen Modell von Rasch, was zugleich auch eine Überprüfung der Eindimensionalität der Verfahren bedeutet. Dabei entsprachen die Analysen im Falle von fünf der sechs verwendeten Reasoning-Tests einer „Art Kreuzvalidierung“ (vgl. dazu Kubinger, 2005; Kubinger et al., 2011, S. 509, 561), da diese bereits in früheren Untersuchungen im Rahmen der Testentwicklung auf Rasch-Modell-Geltung überprüft worden waren (siehe z. B. Gamsjäger, 2012, für Untersuchungen zu *Gleichungen* und *N-TOP*; Hagenmüller, 2011, zum Test *Formale Folgerichtigkeit*; Ünal, 2014, zu *Figurale Analogien* und *N-TOP*; Undeutsch, 2010b, zum *Färbigen Matrizentest*). Sollten sich Items als nicht mit dem Rasch-Modell in Einklang stehend herausstellen, so sollten diese aus dem Itempool des jeweiligen Verfahrens entfernt werden.

Des Weiteren sollte die verwendete Reasoning-Test-Batterie einer Dimensionsanalyse unterzogen werden sowie ein Vergleich der Verfahren in Hinblick auf die jeweils

¹ Der *Färbige Matrizentest* wurde freundlicherweise vom P&T Prieler Tometich Verlag für diese Untersuchung zur Verfügung gestellt.

damit erfassten Fähigkeitsbereiche erfolgen. Hierfür wurde einerseits eine exploratorische Faktorenanalyse durchgeführt, die Aufschluss darüber geben sollte, wie viele bzw. welche latenten Fähigkeitsdimensionen die Tests gemeinsam erfassen; andererseits wurden die Verfahren auch auf lineare Zusammenhänge überprüft, um festzustellen, wie ähnlich bzw. unterschiedlich die damit erfassten Leistungsbereiche einander sind, und somit Aussagen zur konvergenten bzw. diskriminanten Validität machen zu können.

Für die Ziele im Zusammenhang mit den Rasch-Modell-Analysen lassen sich folgende Hypothesen formulieren:

H₀ (1):

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

H₀ (2):

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

H₀ (3):

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

H₀ (4):

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

H₀ (5):

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

H₀ (6):

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

Bezüglich der diskriminanten und konvergenten Validität ist von folgender Hypothese auszugehen:

H₀ (7):

Zwischen Verfahren, die Reasoning im selben Intelligenzbereich (fluid bzw. kristallisiert) erfassen, besteht ein positiver linearer Zusammenhang, der größer ist als jener zwischen Verfahren, die dem jeweils anderen Intelligenzbereich angehören, selbst wenn bei diesen dieselbe Art von Material (lexikalisch, numerisch bzw. figural) zum Einsatz kommt.

Was Hypothese 7 betrifft, so ist anzunehmen, dass zwischen Verfahren, die gleichartiges Material verwenden, dennoch eine geringe Korrelation besteht, auch wenn diese nicht dieselbe Art Reasoning (fluid bzw. kristallisiert) erfassen. Jedoch sollten Tests, die weder demselben Intelligenzbereich noch derselben Materialform zuzuteilen sind, am schwächsten miteinander korrelieren (siehe dazu Kapitel 3.1).

Da die Faktorenanalyse als hypothesengenerierende Methode eingesetzt wird, lassen sich hierfür keine klaren Hypothesen aufstellen. Auf Basis theoretischer Überlegungen und aufbauend auf der siebten Hypothese wird aber erwartet, dass sich die Reasoning-Test-Batterie entweder auf einen gemeinsamen Reasoning-Faktor zurückführen lässt oder auf zwei Faktoren, die den fluiden sowie den kristallisierten Intelligenzbereich nach Cattell (z. B. 1943, 1963) repräsentieren (siehe auch die genaue Beschreibung der Testbatterie im folgenden Abschnitt bzw. vgl. Kapitel 1.3 zur Intelligenztheorie von Cattell).

5. Beschreibung der Testbatterie

In diesem Abschnitt werden die Reasoning-Tests, die im Rahmen der vorliegenden Diplomarbeit untersucht wurden, näher vorgestellt. Dabei wird beschrieben, welche Fähigkeiten mit den Verfahren jeweils erfasst werden sollen und wie das Itemmaterial aussieht. Die Anzahl der Items, die im Rahmen der empirischen Erhebung pro Test vorgegeben wurden, sowie die vorgesehenen Bearbeitungszeiten werden in Kapitel 6 dargestellt.

Insgesamt wurden sechs verschiedene Reasoning-Tests untersucht, welche in Tabelle 1 überblicksmäßig samt Zuordnung zur fluiden bzw. kristallisierten Intelligenz sensu Cattell sowie verwendeter Materialform angeführt sind.

Tabelle 1: Überblick über die untersuchte Reasoning-Test-Batterie

Verfahren	Zuordnung sensu Cattell	Material
<i>Formale Folgerichtigkeit</i>	fluid	figural
<i>Sprachliche Analogien</i>	kristallisiert	lexikalisch
<i>Färbiger Matrizentest</i>	fluid	figural
<i>N-TOP</i>	fluid	numerisch
<i>Figurale Analogien</i>	kristallisiert	figural
<i>Gleichungen</i>	kristallisiert	numerisch

5.1. Der Untertest *Formale Folgerichtigkeit* aus AID-Gruppe

Das Verfahren *Formale Folgerichtigkeit* stellt den zwölften Untertest des *Adaptiven Intelligenz Diagnostikums 3* (AID 3, Version 3.1; Kubinger & Holocher-Ertl, 2014) dar. Der AID – in seiner ersten Version erschienen im Jahr 1985 (vgl. Kubinger & Wurst, 1985) – ist eine im deutschsprachigen Raum etablierte Intelligenz-Testbatterie für Kinder und Jugendliche im Alter von 6;0 bis 15;11 Jahren, die als Individualtest vorgegeben wird. In der aktuellen Version 3.1 besteht sie aus 12 Untertests und 5 fakultativ vorzugebenden Zusatzuntertests (Kubinger & Holocher-Ertl, 2014). Seit einigen Jahren wird an einer Gruppenversion der Intelligenz-Testbatterie gearbeitet (Arbeitstitel: AID-Gruppe; Kubinger & Hagenmüller, in Vorb.), die es ermöglichen soll, mehrere Kinder und Jugendliche gleichzeitig zu testen, und die dennoch dieselbe Fähigkeit je Untertest erfassen soll wie der AID (genauer: AID 2, Version 2.2; Kubinger, 2009a; bzw. AID 3, Version 3.1, Kubinger & Holocher-Ertl, 2014) im Rahmen von Einzeltestungen.

Der Untertest *Formale Folgerichtigkeit*, der von Hagenmüller (2011) für den AID 3 entwickelt wurde, dient der Erfassung von Reasoning im Sinne der fluiden Intelli-

genz nach Cattell mithilfe von figuralen Aufgabenmaterial. Im Rahmen der vorliegenden Untersuchung kam die Version aus dem *AID-Gruppe* (Kubinger & Hagenmüller, in Vorb.) zur Anwendung. Die Aufgaben dieses Untertests bestehen aus Reihen von geometrischen Formen, die von der Testperson logisch korrekt fortgesetzt werden sollen. Dabei stehen bei jedem Item dieselben 16 Antwortmöglichkeiten – das sind alle figuralen Elemente, die im Untertest insgesamt vorkommen – zur Verfügung, aus denen die Testperson auszuwählen hat (siehe Abbildung 2).



Abbildung 2: Itembeispiel des Untertests *Formale Folgerichtigkeit* aus *AID-Gruppe* (Kubinger & Hagenmüller, in Vorb.)

5.2. Sprachliche Analogien

Sprachliche Analogien (Kubinger & Ünal, in Vorb.) ist ein Verfahren zur Erfassung des kristallisierten schlussfolgernden Denkens mithilfe von lexikalischem Material. Mit diesem Test wird also mittels Analogieaufgaben nicht nur Reasoning, sondern vor allem auch Allgemeinwissen erhoben. Die Items bestehen aus jeweils drei Begriffen, wobei zwischen den ersten beiden eine Beziehung besteht, die jener zwischen dem dritten vorgegebenen und dem gesuchten Begriff ähnelt (siehe Abbildung 3).

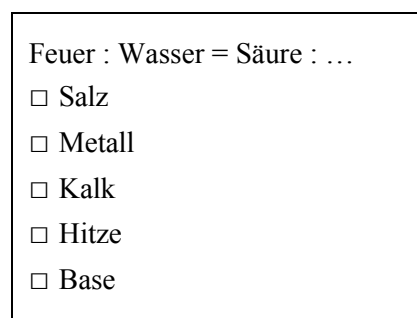


Abbildung 3: Itembeispiel aus *Sprachliche Analogien* (Kubinger & Ünal, in Vorb.)

Die Testperson ist also gefordert, die Art der Relation zwischen den ersten beiden Elementen zu erkennen, diese auf das dritte zu übertragen und so den fehlenden Begriff aus fünf vorgegebenen Antwortmöglichkeiten herauszufinden. Inhaltlich sind die Aufgaben sieben verschiedenen Themenbereichen zugeordnet.

5.3. Der Färbige Matrizentest (FMT)

Der *Färbige Matrizentest* (FMT; Undeutsch, 2010a) ist, wie der Name schon sagt, ein Matrizentest zur Erfassung der Fähigkeit zum logisch-schlussfolgernden Denken, der im Gegensatz zu vielen anderen Matrizentests farbiges Material verwendet und zudem so konzipiert ist, dass sich das gesuchte figurale Element in der Mitte der Matrix befindet. Die Items bestehen jeweils aus einer 3x3-Matrix mit einem Fragezeichen in der Mitte sowie aus acht Antwortmöglichkeiten, aus der die fehlende Figur auszuwählen ist (siehe Abbildung 4). Aufgabe der Testperson ist es also, die Regeln zu erkennen, nach denen die Elemente der Matrix aufgebaut und angeordnet sind und so die gesuchte Figur herauszufinden. Als Matrizentest ist das Verfahren dem fluiden Intelligenzbereich sensu Cattell zuzuordnen.

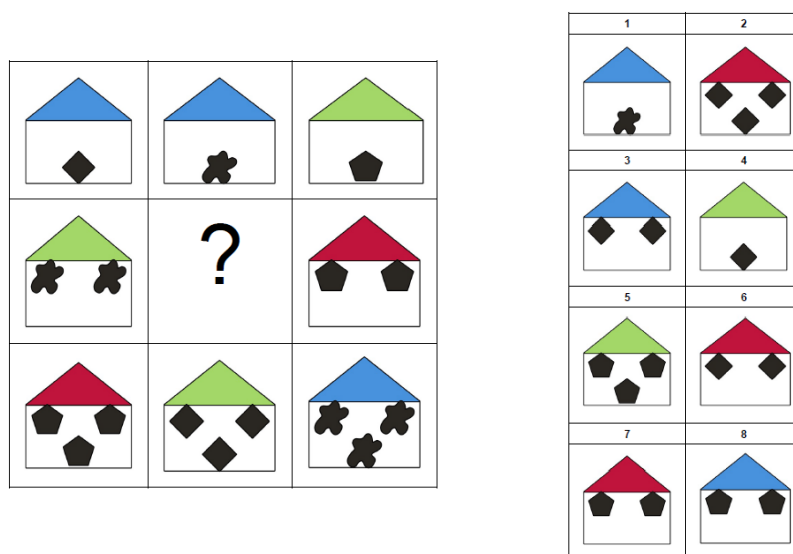


Abbildung 4: Itembeispiel aus dem FMT (Undeutsch, 2010a)

5.4. Numerische Topologien (N-TOP)

Das Verfahren *N-TOP* (Kubinger & Heuberger, in Vorb.) dient zur Erfassung von Reasoning im Sinne des fluiden Intelligenzbereichs mithilfe von numerischem Aufgabenmaterial. Der Testperson werden pro Item drei Rahmen mit Zahlen präsentiert, wobei im dritten Rahmen die jeweils fehlende Ziffer in ein leeres Kästchen einzutragen ist. Abbildung 5 zeigt eines der Instruktionsitems des Verfahrens.

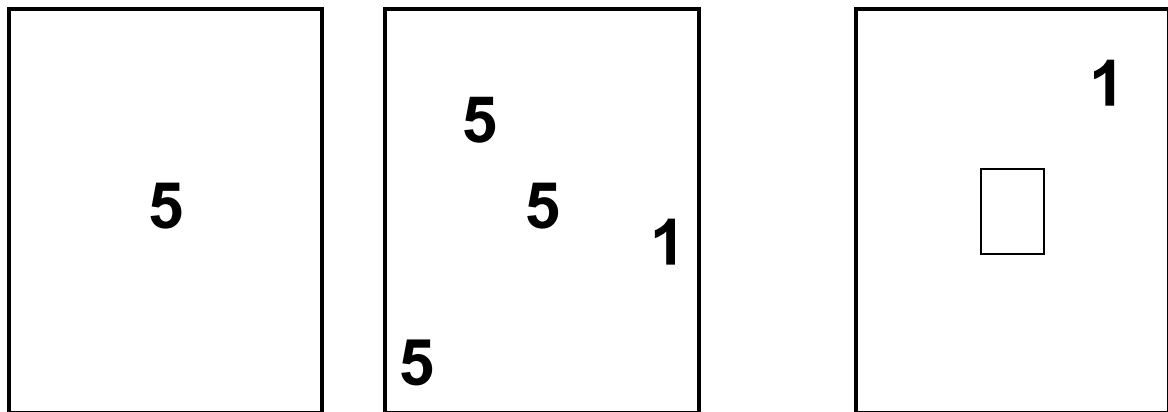


Abbildung 5: Itembeispiel aus *N-TOP* (Kubinger & Heuberger, in Vorb.)

5.5. Figurale Analogien

Der Test *Figurale Analogien* (Ünal & Kubinger, in Vorb.) ist ein Verfahren zur Erfassung des schlussfolgernden Denkens im Bereich der kristallisierten Intelligenz mithilfe von figuralen Analogieaufgaben. Die Testperson wird hier mit inhaltlich sinnvollen Bildern und Symbolen konfrontiert, die in Form einer Gleichung der Art „A : B = C : ?“ präsentiert werden. Wie im Verfahren *Sprachliche Analogien* muss die Testperson durch die Anwendung von Wissen aus verschiedenen Themengebieten zur logisch korrekten Lösung gelangen, indem sie die Beziehung zwischen den ersten beiden Bildern erkennt und auf das dritte Element, das eine ähnliche Beziehung zu einer der Antwortmöglichkeiten aufweist, überträgt. Jenes Bild der zur Verfügung stehenden fünf Auswahlbilder, das die Analogie richtig vervollständigt, soll herausgefunden werden. Zur Illustration ist in Abbildung 6 wiederum das Instruktionsitem als Beispielaufgabe angeführt.

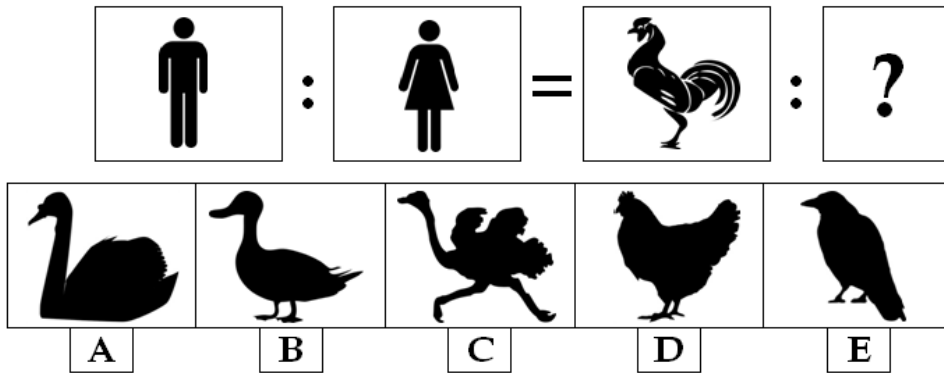


Abbildung 6: Itembeispiel aus *Figurale Analogien* (Ünal & Kubinger, in Vorb.)

5.6. Gleichungen

Gleichungen (Kubinger & Gamsjäger, in Vorb.) ist ein Verfahren zur Erfassung der Reasoning-Fähigkeit, das aus numerischem Aufgabenmaterial, nämlich mathematischen Gleichungen, besteht. Da für die Lösung der Aufgaben mathematische Kenntnisse, also das Wissen um Rechenzeichen und -operationen, notwendig sind, welche vor allem durch Schulbildung erworben werden, ist der Test der kristallisierten Intelligenz zuzuordnen. Pro Item wird der Testperson eine Gleichung vorgegeben, die die Variablen x und y enthält, sowie jeweils vier Antwortmöglichkeiten für x und y (siehe das Instruktionsitem in Abbildung 7). Die Testperson ist nun gefordert, jene x - y -Kombination aus den zur Auswahl stehenden Werten zu finden, die die Gleichung löst. Eine Aufgabe gilt also nur dann als richtig beantwortet, wenn für beide Variablen die korrekten Zahlen angekreuzt wurden. Zur Lösung der Items dürfen von der Testperson keine Hilfsmittel wie Taschenrechner verwendet und keine Notizen gemacht werden.

$x + y = 10$	x	2	8	5	10
	y	9	0	1	3

Abbildung 7: Itembeispiel aus *Gleichungen* (Kubinger & Gamsjäger, in Vorb.)

6. Planung und Durchführung der Untersuchung

Nachdem im vorangegangenen Abschnitt die Testbatterie vorgestellt wurde, soll nun in diesem Kapitel die Untersuchungsplanung sowie der Ablauf der Testungen beschrieben werden.

Für die Überprüfung der Reasoning-Tests in Hinblick auf Geltung des Rasch-Modells sowie die im Anschluss vorgesehene Faktorenanalyse wurde im Vorfeld der Stichprobenakquirierung festgelegt, dass als Testpersonen Schülerinnen und Schüler aus Maturaklassen² allgemeinbildender (AHS; 12. Schulstufe) und berufsbildender höherer Schulen (BHS; 13. Schulstufe) herangezogen werden sollten. Diese stellen eine Population dar, bei der psychologisch-diagnostische Verfahren, die logisch-schlussfolgerndes Denken erfassen, häufig, beispielsweise im Rahmen von Berufs- und Bildungsberatungen oder Aufnahmeverfahren an Universitäten und Fachhochschulen, zur Anwendung kommen. Dabei sollte ungefähr je ein Drittel der Stichprobe aus einer Handelsakademie (HAK), einer Höheren Technischen Lehranstalt (HTL) bzw. einem Gymnasium (AHS) stammen.

Nach der Kontaktaufnahme mit einigen Schulen in Niederösterreich und der Zusage mehrerer Direktorinnen und Direktoren, dass Testungen in ihren Schulen durchgeführt werden dürften, wurde ein Antrag auf Genehmigung der Untersuchung an den Landesschulrat für NÖ gestellt. Nach Erhalt der Genehmigung (siehe im Anhang unter B) wurden Informationsblätter an die Schüler/innen der Maturaklassen bzw. deren Eltern verteilt, in denen das Projekt vorgestellt wurde (siehe im Anhang unter C bzw. D). Im Falle von minderjährigen Schüler/innen wurden die Eltern um ihr Einverständnis mit der Teilnahme ihres Kindes an der Testung gebeten. Die über 18-Jährigen sollten ihre Bereitschaft zur Teilnahme durch ihre eigene Unterschrift in einer Teilnehmer/innenliste zum Ausdruck bringen. Danach wurden mit den Direktionen Termine für die Durchführung der Untersuchung vereinbart, wobei dabei die Vorstellungen der Schulen stark berücksichtigt wurden, um den regulären Ablauf des Unterrichts so wenig wie möglich zu stören.

Nachdem in einem Vortest mit einer 17-jährigen Schülerin einer Höheren Lehranstalt für wirtschaftliche Berufe (HLW) die festgelegten Bearbeitungszeiten (siehe weiter

² Der österreichischen Matura entspricht das Abitur in Deutschland.

unten) für die einzelnen Tests überprüft und der Ablauf geprobt worden war, fanden die Testungen zwischen dem 24.10. und 18.11.2014 an insgesamt fünf Schulen in Niederösterreich statt, und zwar an den Handelsakademien Tulln und Hollabrunn, an der Höheren Technischen Lehranstalt Mödling (Abteilungen Bautechnik, Holztechnik und Innenarchitektur), am Bundesgymnasium und Bundesrealgymnasium Schwechat sowie am Bundesgymnasium Wiener Neustadt Zehnergasse. Die einzelnen Maturaklassen wurden pro Schule entweder parallel, nacheinander oder in wenigen Fällen auch zu verschiedenen Terminen getestet. Die Untersuchung wurde während der regulären Unterrichtszeit in den Klassenräumen der Schüler/innen durchgeführt und dauerte durchschnittlich eine Stunde. Pro Klasse waren in den meisten Fällen zwei Testleiterinnen anwesend. Als solche fungierten neben der Diplomandin selbst einige weitere Studentinnen aus dem Arbeitsbereich Psychologische Diagnostik der Universität Wien, die über den Ablauf und ihre Aufgaben während der Testungen im Vorfeld geschult worden waren.

Zu Beginn der Testung wurden die Teilnehmer/innen umfassend über den Ablauf informiert. Sie wurden außerdem darauf hingewiesen, dass sie mittels ihres individuellen Testpersonencodes, der zum Zwecke der Anonymisierung auf jedem Testheft aufgedruckt war, bei Interesse einen schriftlichen Ergebnisbericht über ihre Testleistungen per E-Mail bei der Diplomandin anfordern konnten (siehe ein Beispiel eines Ergebnisberichts im Anhang unter F). Der Testpersonencode setzte sich aus drei Buchstaben und zwei Ziffern zusammen, wobei die Buchstaben die jeweilige Schulart, den Schulort (bzw. die Abteilung im Falle der HTL) und die Klasse kennzeichneten und die Ziffern eine fortlaufende Nummerierung der Testhefte pro Klasse darstellten.

Nach den allgemeinen Erläuterungen und Anweisungen zur Testung erhielt jede Testperson ein Blatt zur Erhebung der demografischen Daten und je ein Testheft für die sechs Untertests. Die Testhefte lagen in zwei Testformen (A und B) vor, wobei jeder Schüler bzw. jede Schülerin alle Untertests derselben Testform, also entweder alle der Form A oder alle der Form B, bearbeitete. Schüler/innen, die nebeneinander saßen, erhielten jeweils unterschiedliche Testformen, sodass abwechselnd eine Person Version A und die nächste Version B usw. ausgeteilt bekam, und die Testformen ungefähr gleich häufig eingesetzt wurden. Die beiden Testversionen stellten bei allen Untertests – außer im Falle des *FMT*, bei dem beide Testformen exakt gleich waren – Quasi-Parallelformen dar, d. h. allen Testpersonen wurden dieselben Items, jedoch in unterschiedlicher Reihenfolge, vorgegeben. Dies diente vor allem dazu, die Möglichkeit des Abschreibens der Testpersonen

untereinander zu verringern. Das genaue Testdesign der einzelnen Untertests der Testbatterie ist dem Anhang unter A zu entnehmen.

Nach dem Ausfüllen der demografischen Daten wurde mit der Testbearbeitung begonnen. Die Testpersonen wurden vor jedem Untertest mündlich instruiert, wobei die Instruktionen auch zum Mitlesen auf den Testheften aufgedruckt waren³. Für die Bearbeitung der einzelnen Verfahren gab es jeweils eine Zeitbegrenzung, die den Schüler/innen stets im Vorhinein mitgeteilt wurde. Die vorgesehenen Bearbeitungszeiten sowie auch die Anzahl der vorgegebenen Items pro Test sind in Tabelle 2 dargestellt.

Tabelle 2: Itemanzahl und Bearbeitungszeit pro Verfahren

Verfahren	Itemanzahl	Bearbeitungszeit
<i>Formale Folgerichtigkeit</i>	22	7 Minuten
<i>Sprachliche Analogien</i>	21	8 Minuten
<i>Färbiger Matrizentest</i>	18	10 Minuten
<i>N-TOP</i>	15	8 Minuten
<i>Figurale Analogien</i>	24	8 Minuten
<i>Gleichungen</i>	12	8 Minuten

Nach Ablauf der jeweiligen Bearbeitungszeit wurden die Schüler/innen gebeten, einen Strich unter die zuletzt bearbeitete Aufgabe zu ziehen. Dies war notwendig, um bei der Dateneingabe nachvollziehen zu können, ob die jeweilige Testperson Items, bei denen keine Antwort markiert worden war, nicht lösen hatte können oder aufgrund der Zeitvorgabe gar nicht bearbeitet hatte. Sofern alle Schüler/innen einer Klasse bereits vor Ablauf der Zeit mit der Bearbeitung eines Untertests fertig waren, wurde schon etwas früher zum nächsten Test übergegangen, um so größere Unruhen zu vermeiden. Dies kam nur bei den Verfahren *Sprachliche Analogien* und *Figurale Analogien* vor, bei denen in sehr vielen Klassen nicht die gesamten acht Minuten benötigt wurden. Die Testleiterinnen wurden

³ Die Anweisungen für die Testleiterinnen sowie die mündlichen Instruktionen für die Bearbeitung der Tests, die den Schüler/innen gegeben wurden, sind im Anhang unter E zu finden.

vor der Untersuchung angewiesen, die tatsächlich benötigten Bearbeitungszeiten pro Test sowie eventuelle Auffälligkeiten während der Testung auf einem Testleiter/innenprotokoll für jede Klasse zu notieren. Dabei wurde insgesamt relativ selten von auffälligen Verhaltensweisen berichtet. Zu vermehrten Unruhen oder Versuchen der Testpersonen, die Aufgaben gemeinsam zu bearbeiten oder die Lösungen zu vergleichen, kam es hauptsächlich dann, wenn ein Großteil der Schüler/innen vor Ablauf der Zeitbegrenzung mit der Bearbeitung eines Verfahrens fertig war und auf die anderen Teilnehmer/innen warten musste. Insgesamt wurde aber festgehalten, dass sich die Testpersonen in den meisten Klassen sehr ruhig und gewissenhaft verhielten bzw. spätestens nach dem erneuten Hinweis, dass es für die Auswertung unbedingt notwendig sei, dass die Aufgaben von jeder Person selbstständig bearbeitet werden, eventuelle Gespräche oder Ähnliches weitestgehend unterließen.

7. Stichprobe

Die Stichprobe besteht aus insgesamt 265 Schülerinnen und Schülern aus allgemeinbildenden und berufsbildenden höheren Schulen. Davon sind 43 % männlich und 57 % weiblich. 88 Personen stammen aus der achten Klasse eines Gymnasiums (33 %), 93 aus der fünften Klasse einer HAK (35 %) und 84 aus der fünften Klasse einer HTL (32 %).

Tabelle 3: Verteilung der Stichprobe in Bezug auf das Alter in Jahren ($n = 265$)

Alter in Jahren	absolute Häufigkeit	Prozent
16	1	0,4
17	58	21,9
18	139	52,5
19	52	19,6
20	12	4,5
21	3	1,1

Wie aus Tabelle 3 hervorgeht, liegt das Alter der Testpersonen zwischen 16 und 21 Jahren, wobei ungefähr die Hälfte der getesteten Schüler/innen 18 Jahre alt ist. Was die Muttersprache betrifft, so gaben etwa 90 % Deutsch und nur knappe 10 % eine andere Sprache an.

8. Darstellung der Ergebnisse

In den folgenden Abschnitten werden die Ergebnisse der Untersuchung im Detail vorgestellt. Die Kapitel unter 8.1 beschäftigen sich mit den Rasch-Modell-Analysen, wobei zunächst die Vorgehensweise bei der Überprüfung der Verfahren auf Geltung des Rasch-Modells erläutert wird und in den Kapiteln 8.1.1 bis 8.1.6 die diesbezüglichen Ergebnisse bei den Verfahren *Formale Folgerichtigkeit*, *Sprachliche Analogien*, *FMT*, *N-TOP*, *Figureale Analogien* und *Gleichungen* besprochen werden. In Abschnitt 8.2 bzw. 8.2.1 werden die Ergebnisse der Faktorenanalyse und in Kapitel 8.3 die Korrelationen zwischen den Tests präsentiert.

Für die statistischen Analysen wurde die Open-Source-Statistik-Software *R* (Version 3.1.2; R Core Team, 2014) verwendet, wobei bei den Rasch-Modell-Analysen das Programmpaket *eRm* (extended Rasch modeling; Version 0.15-4; Mair, Hatzinger, & Maier, 2014) und für die Faktorenanalyse das Programmpaket *psych* (Version 1.5.1; Revelle, 2015) zur Anwendung kam.

8.1. Analysen zur Überprüfung der Geltung des Rasch-Modells

Zur Überprüfung der Geltung des Rasch-Modells werden bei allen untersuchten Tests die folgenden Teilungskriterien verwendet:

- **Anzahl gelöster Aufgaben (Score):** Rohwert $\leq$ Median vs. Rohwert $>$ Median
- **Geschlecht:** männlich vs. weiblich
- **Schultyp:** AHS (Gymnasium) vs. BHS (HAK und HTL)

Um zu prüfen, ob die Itemparameterschätzungen in den jeweiligen Teilstichproben statistisch gleich sind, wird der Likelihood-Ratio-Test (LRT) nach Andersen (1973) durchgeführt, wobei das akzeptierte Risiko 1. Art auf $\alpha = 0,01$ festgelegt wird, da mehrere Signifikanztests durchgeführt werden (vgl. dazu z. B. Kubinger et al., 2011, S. 270 f.).

Falls es zu einem signifikanten Ergebnis kommt und somit von Modellabweichungen auszugehen ist, erfolgt eine grafische Modellkontrolle, um das bzw. die nicht modellkonformen Items zu identifizieren. Jene Items, deren 95%-Konfidenzellipse die 45°-Gerade nicht schneidet, sind als auffällig zu werten. Die Konfidenzellipse stellt jenen Bereich dar, innerhalb dessen die tatsächliche Itemschwierigkeit in den Teilstichproben mit einer Wahrscheinlichkeit von 95 % liegt. Von einer relevanten Modellabweichung wird in diesen Fällen dann ausgegangen, wenn die Differenz der Itemparameterschätzungen der jeweiligen zwei Teilstichproben mehr als ein Zehntel des Range der gesamten Itemschwierigkeitsparameterschätzungen beträgt (vgl. dazu Goethals, 1994, zitiert nach Kubinger, 2005, S. 389; bzw. Kubinger, 2005). Zusätzlich wird der z-Test nach Fischer und Scheiblechner (1970, zitiert nach Kubinger et al., 2011, S. 556 f.) durchgeführt, der aber aufgrund der Überhöhung des Risikos 1. Art nur deskriptiv zu interpretieren ist.

Die nicht modellkonformen Items werden schrittweise aus dem Itempool entfernt, bis die Ergebnisse des Likelihood-Ratio-Tests für alle Teilungskriterien nicht mehr signifikant ausfallen. Beim Ausscheiden von Items soll vor allem auf mögliche inhaltliche Gründe geachtet werden, welche insbesondere bei *Sprachliche Analogien* und *Figurale Analogien* eine wichtige Rolle spielen.

Für Items, die von keiner oder allen Testpersonen einer Teilstichprobe gelöst wurden, können für das betreffende Teilungskriterium keine Schwierigkeitsparameter geschätzt werden; diese können daher in diesem Fall nicht berücksichtigt werden. Die Anzahl der schätzbaren Items beträgt immer um eins mehr als die Anzahl der Freiheitsgrade (df) und ist jeweils den Tabellen über die Ergebnisse der Likelihood-Ratio-Tests zu entnehmen.

Die Anzahl der Testpersonen, die in die Rasch-Modell-Analysen der einzelnen Verfahren jeweils eingehen, ist in den folgenden Abschnitten stets zu Beginn angeführt. Personen werden dann von den Analysen ausgeschlossen, wenn sie einen Test instruktionswidrig bearbeiteten und daher keine aussagekräftigen Daten lieferten.

8.1.1. Ergebnisse der Rasch-Modell-Analysen bei *Formale Folgerichtigkeit*

Von Untertest 12 des AID-Gruppe, *Formale Folgerichtigkeit* (Kubinger & Hagenmüller, in Vorb.), wurden 22 Items dahingehend überprüft, ob diese mit dem Rasch-Modell konform sind. Die Stichprobe besteht hier aus $n = 264$ Personen, die in die Analysen eingingen.

Item 101 konnte im Teilungskriterium „Schultyp“ nicht geschätzt werden, da es in der Teilstichprobe der AHS-Schüler/innen von keiner Person gelöst wurde, und wurde daher bei den Rasch-Modell-Analysen für dieses Teilungskriterium nicht berücksichtigt.

Der Likelihood-Ratio-Test nach Andersen (1973) für das Teilungskriterium „Schultyp“ ergibt ein signifikantes Ergebnis. Bei den übrigen Teilungskriterien muss von keinen signifikanten Modellabweichungen ausgegangen werden (siehe Tabelle 4).

Tabelle 4: Ergebnisse der Likelihood-Ratio-Tests
(*Formale Folgerichtigkeit*, alle Items)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	25,870	21	38,932	0,211
Geschlecht	26,291	21	38,932	0,196
Schultyp	40,539	20	37,566	0,004

Die Ergebnisse der grafischen Modellkontrollen sind den Abbildungen 8 bis 10 zu entnehmen. Für jene Items, bei denen die Differenz der Parameterschätzungen in den jeweiligen Teilstichproben mehr als ein Zehntel des Range der gesamten Itemschwierigkeitsparameterschätzungen beträgt, sind die 95%-Konfidenzellipsen eingezeichnet. Diese Vorgehensweise findet sich auch bei allen folgenden grafischen Modellkontrollen.

Unter Berücksichtigung der Konfidenzellipsen sind im Teilungskriterium „Schultyp“ die Items 95, 100 und 105 als auffällig zu werten. Die Ergebnisse des z-Tests nach Fischer und Scheiblechner für dieses Teilungskriterium sind in Tabelle 5 dargestellt. Diese lassen auf eine Modellabweichung von Item 95 schließen.

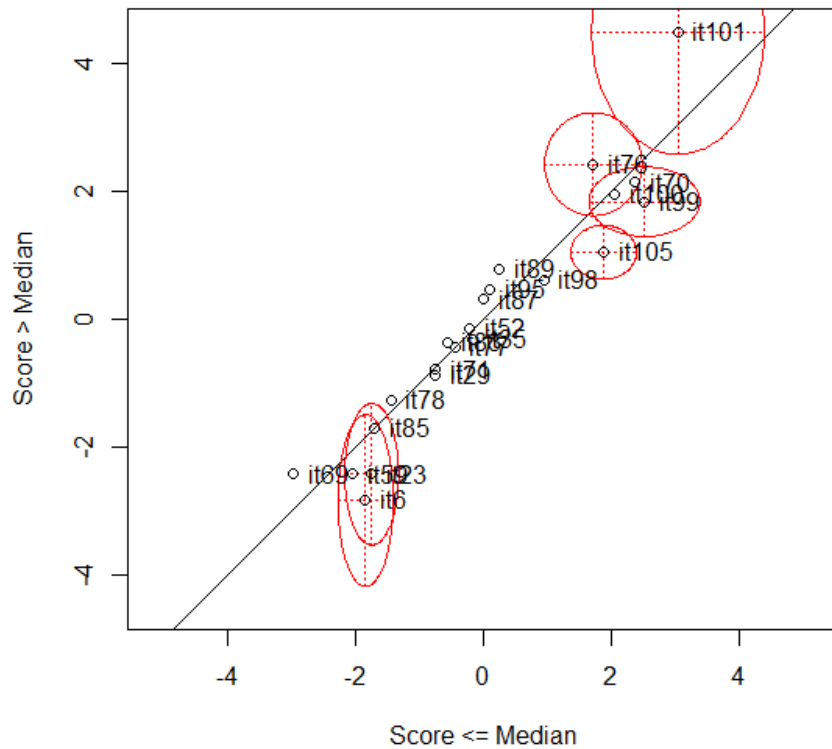


Abbildung 8: Grafische Modellkontrolle für das Teilungskriterium „Score“
(Formale Folgerichtigkeit, alle Items)

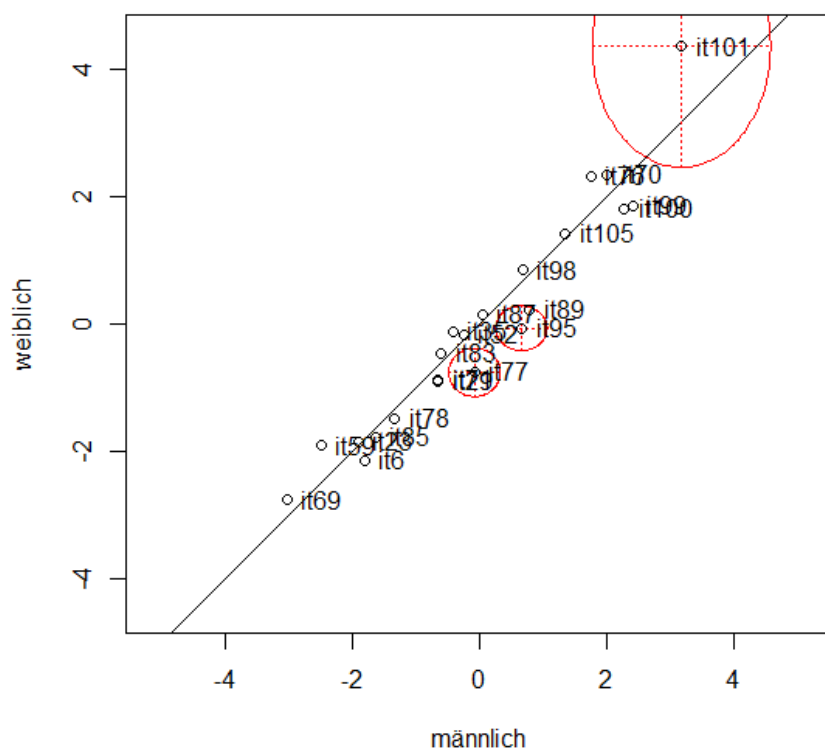


Abbildung 9: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
(Formale Folgerichtigkeit, alle Items)

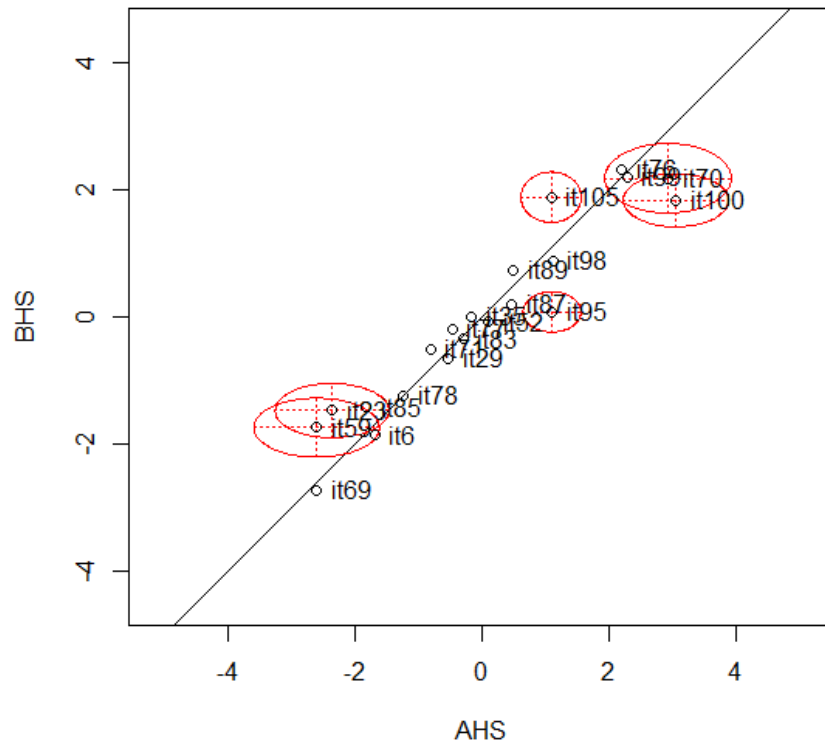


Abbildung 10: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(Formale Folgerichtigkeit, alle Items)

Tabelle 5: Ergebnisse des z-Tests für das Teilungskriterium „Schultyp“
(Formale Folgerichtigkeit, alle Items)

Itemnummer	z-Wert	p
6	-0,385	0,701
23	1,774	0,076
29	-0,373	0,709
35	0,596	0,551
52	-0,633	0,527
59	1,567	0,117
69	-0,188	0,851
70	-1,275	0,202
71	0,990	0,322
76	0,211	0,833

77	0,951	0,342
78	0,013	0,990
83	-0,153	0,878
85	0,957	0,338
87	-0,914	0,361
89	0,864	0,387
95	-3,575	0,000
98	-0,714	0,475
99	-0,175	0,861
100	-2,543	0,011
105	2,545	0,011

Da Item 95 grafisch sowie im z-Test im Teilungskriterium „Schultyp“ und auch grafisch im Teilungskriterium „Geschlecht“ auffällig ist, wird dieses trotz fehlender inhaltlicher Gründe aus dem Itempool entfernt⁴. Nach Ausschluss von Item 95 kommt es nun in den Likelihood-Ratio-Tests bei keinem Teilungskriterium mehr zu einem signifikanten Ergebnis (siehe Tabelle 6).

Tabelle 6: Ergebnisse der Likelihood-Ratio-Tests
(Formale Folgerichtigkeit, nach Ausschluss von Item 95)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	21,726	20	37,566	0,356
Geschlecht	21,245	20	37,566	0,383
Schultyp	26,779	19	36,191	0,110

⁴ Bei Interesse können die auf Basis der Ergebnisse dieser Arbeit ausgeschiedenen Items beim Betreuer, Herrn Univ.-Prof. Mag. Dr. Kubinger, eingesehen werden.

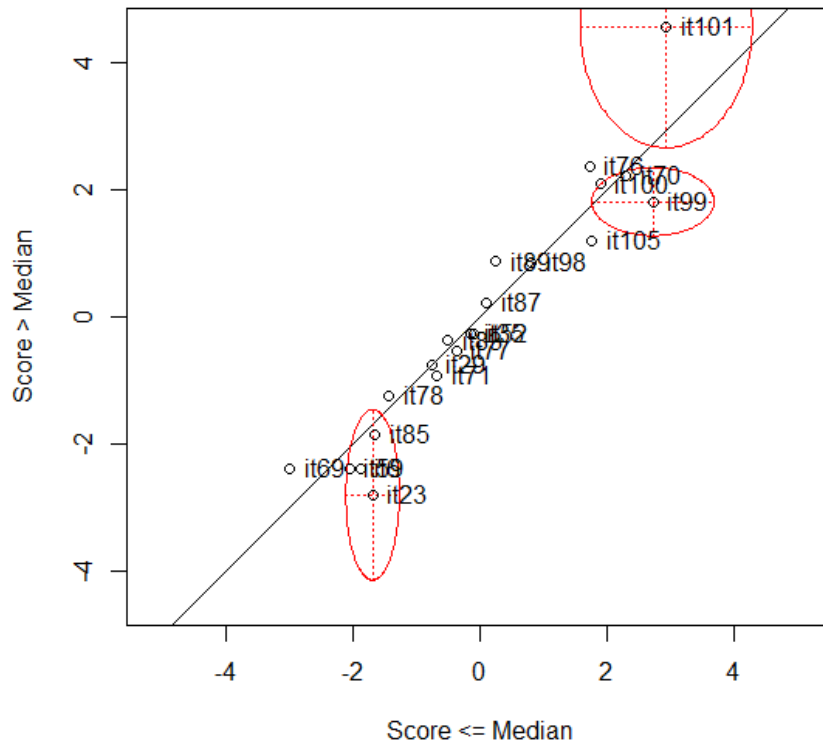


Abbildung 11: Grafische Modellkontrolle für das Teilungskriterium „Score“
 (Formale Folgerichtigkeit, nach Ausschluss von Item 95)

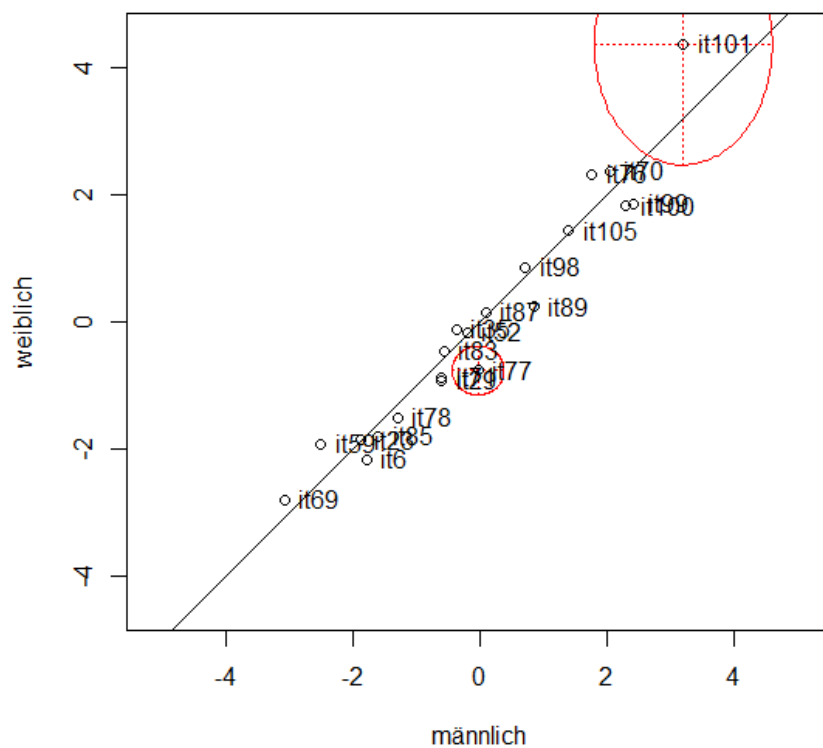


Abbildung 12: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
 (Formale Folgerichtigkeit, nach Ausschluss von Item 95)

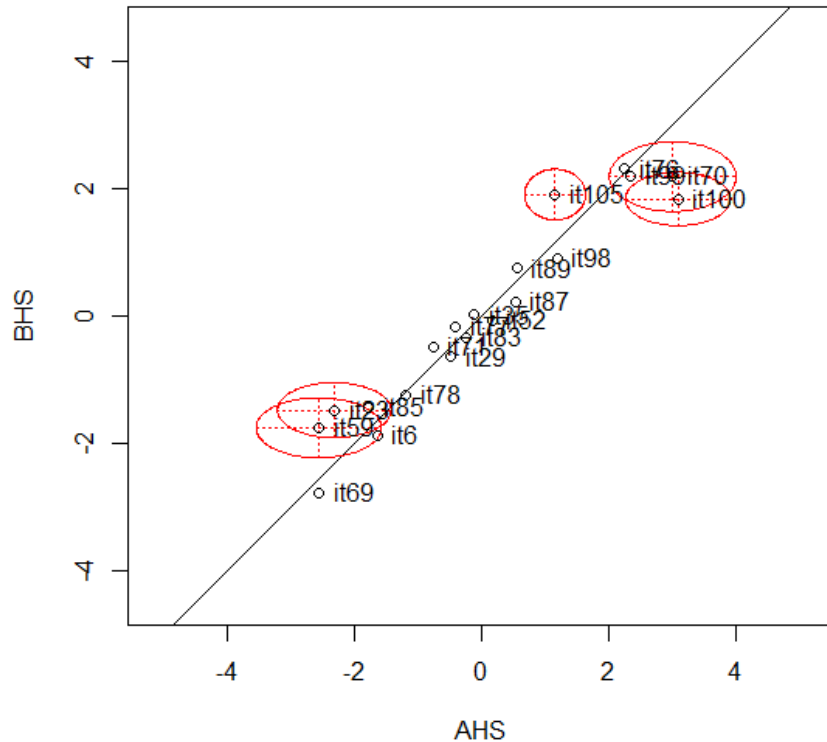


Abbildung 13: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(*Formale Folgerichtigkeit*, nach Ausschluss von Item 95)

Die Abbildungen 11, 12 und 13 zeigen die Ergebnisse der grafischen Modellkontrolle für die drei Teilungskriterien nach Ausschluss von Item 95. In diesen zeigen sich noch einige Items als auffällig (vgl. bspw. die Items 100 und 105 im Teilungskriterium „Schultyp“ in Abbildung 13), da jedoch in den Likelihood-Ratio-Tests keine signifikanten Modellabweichungen mehr beobachtet wurden, werden keine weiteren Items ausgeschlossen.

Die verbleibenden Items des Untertests *Formale Folgerichtigkeit* können somit als dem Rasch-Modell entsprechend angesehen werden. Aufgrund dieser Ergebnisse kann angenommen werden, dass die übrigen Items des Untertests *Formale Folgerichtigkeit* aus AID-Gruppe eindimensional logisch-schlussfolgerndes Denken im figuralen Bereich erfassen.

8.1.2. Ergebnisse der Rasch-Modell-Analysen bei *Sprachliche Analogien*

Vom Test *Sprachliche Analogien* (Kubinger & Ünal, in Vorb.) wurden 21 Items auf Geltung des Rasch-Modells überprüft. Auch in diese Analysen gingen die Daten von $n = 264$ Testpersonen ein.

Wie aus Tabelle 7 ersichtlich ist, fallen die Ergebnisse des Likelihood-Ratio-Tests nach Andersen für alle Teilungskriterien signifikant aus.

Tabelle 7: Ergebnisse der Likelihood-Ratio-Tests
(*Sprachliche Analogien*, alle Items)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	66,671	20	37,566	0,000
Geschlecht	40,063	20	37,566	0,005
Schultyp	276,040	20	37,566	0,000

Im Folgenden sind für alle Teilungskriterien die Ergebnisse der grafischen Modellkontrollen sowie des z-Tests nach Fischer und Scheiblechner für den ersten Berechnungsdurchgang angeführt.

Im Teilungskriterium „Score“ zeigen sich in der grafischen Modellkontrolle die Items 8, 25, 36, 40 und 41 (vgl. Abbildung 14) sowie im z-Test die Items 8, 40 und 41 als auffällig (vgl. Tabelle 8). Im Teilungskriterium „Geschlecht“ schneiden die Konfidenzellipse der Items 8 und 27 die 45°-Gerade nicht und sind somit als auffällig zu werten (vgl. Abbildung 15). Im entsprechenden z-Test wird nur Item 8 auffällig (vgl. Tabelle 9). Eine sehr starke Auffälligkeit ist bei Item 27 im Teilungskriterium „Schultyp“ zu beobachten, wobei in der grafischen Modellkontrolle zudem auch die Items 8, 25, 36, 39 und 41 deutliche Abweichungen zeigen (vgl. Abbildung 16). Im z-Test wird dieses Ergebnis für das Teilungskriterium „Schultyp“ bestätigt (vgl. Tabelle 10).

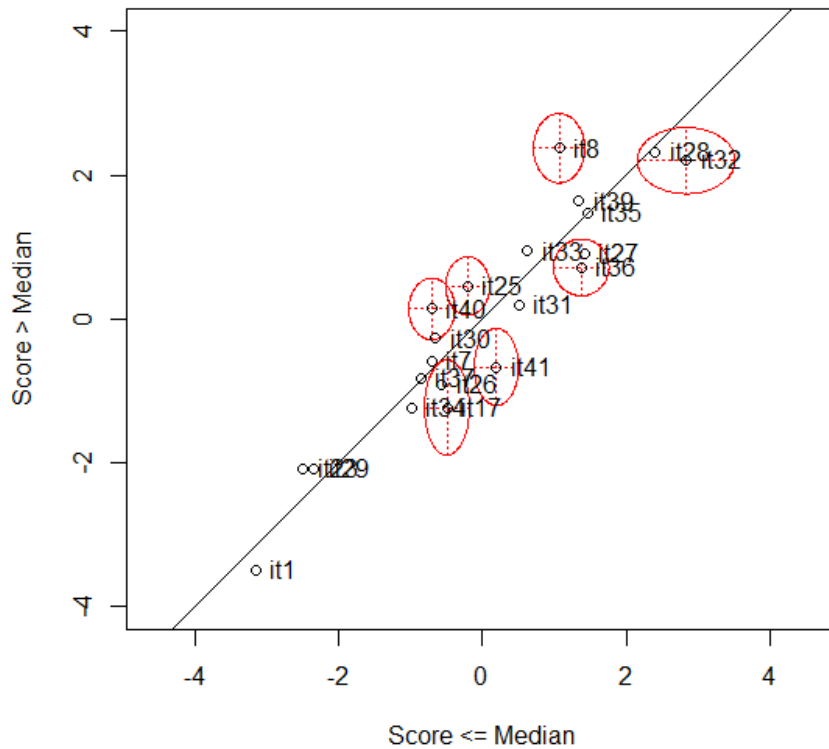


Abbildung 14: Grafische Modellkontrolle für das Teilungskriterium „Score“
(*Sprachliche Analogien*, alle Items)

Tabelle 8: Ergebnisse des z-Tests für das Teilungskriterium „Score“
(*Sprachliche Analogien*, alle Items)

Itemnummer	z-Wert	p
1	-0,321	0,748
7	0,370	0,712
8	4,281	0,000
17	-1,961	0,050
23	0,747	0,455
25	2,563	0,010
26	-1,006	0,314
27	-1,799	0,072
28	-0,235	0,814
29	0,475	0,635

30	1,317	0,188
31	-1,122	0,262
32	-1,502	0,133
33	1,236	0,216
34	-0,629	0,529
35	0,069	0,945
36	-2,327	0,020
37	0,094	0,925
39	1,068	0,286
40	3,104	0,002
41	-2,713	0,007

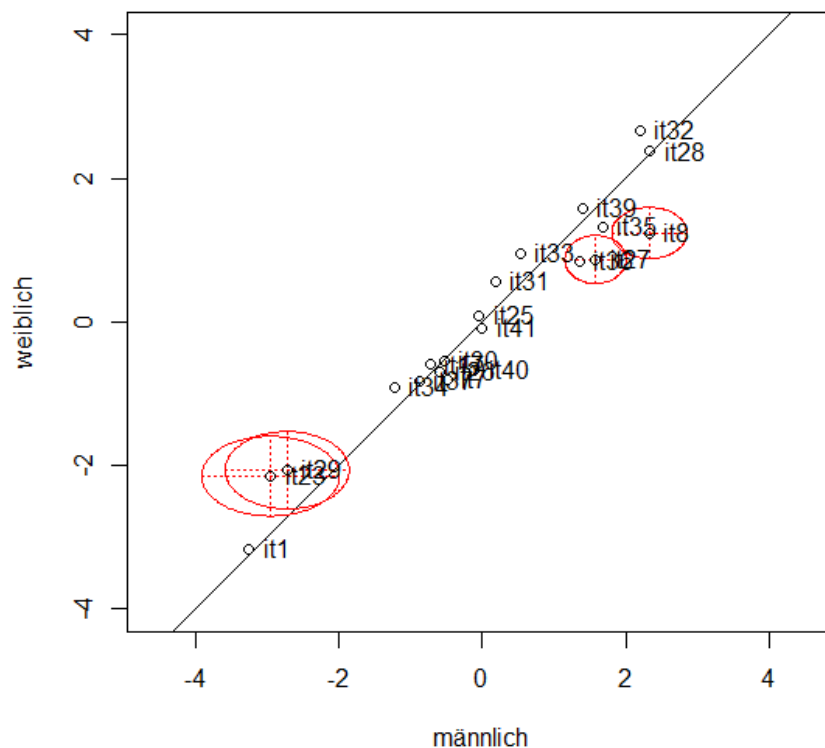


Abbildung 15: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
(*Sprachliche Analogien*, alle Items)

Tabelle 9: Ergebnisse des z-Tests für das Teilungskriterium „Geschlecht“
(*Sprachliche Analogien*, alle Items)

Itemnummer	z-Wert	p
1	0,117	0,907
7	-1,105	0,269
8	-3,364	0,001
17	0,451	0,652
23	1,430	0,153
25	0,558	0,577
26	-0,410	0,682
27	-2,545	0,011
28	0,111	0,912
29	1,268	0,205
30	0,016	0,987
31	1,465	0,143
32	1,265	0,206
33	1,666	0,096
34	1,000	0,317
35	-1,238	0,216
36	-1,880	0,060
37	0,201	0,840
39	0,635	0,525
40	-2,033	0,042
41	-0,326	0,745

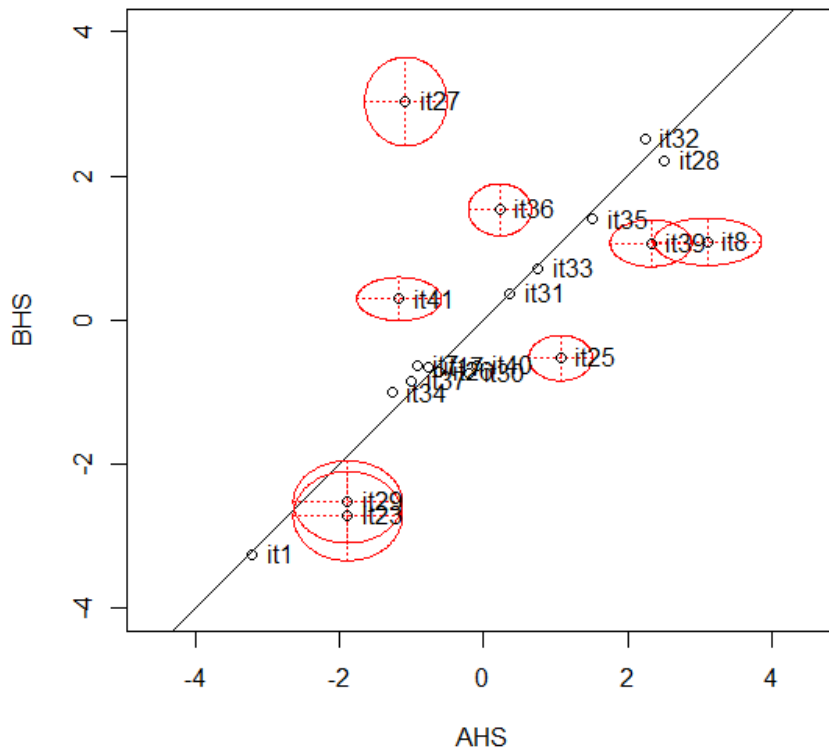


Abbildung 16: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(*Sprachliche Analogien*, alle Items)

Tabelle 10: Ergebnisse des z-Tests für das Teilungskriterium „Schultyp“
(*Sprachliche Analogien*, alle Items)

Itemnummer	z-Wert	p
1	-0,051	0,959
7	0,885	0,376
8	-4,821	0,000
17	0,339	0,735
23	-1,666	0,096
25	-5,754	0,000
26	-0,291	0,771
27	9,594	0,000
28	-0,755	0,450
29	-1,312	0,190

30	-1,886	0,059
31	-0,002	0,998
32	0,746	0,455
33	-0,114	0,909
34	0,736	0,462
35	-0,386	0,700
36	4,511	0,000
37	0,477	0,633
39	-3,759	0,000
40	-1,920	0,055
41	4,359	0,000

Aufgrund der starken Auffälligkeit von Item 27 in der grafischen Modellkontrolle im Teilungskriterium „Schultyp“ wird zunächst dieses Item auf mögliche inhaltliche Gründe für die Modellabweichung überprüft. In Abbildung 16 zeigt sich, dass Item 27 AHS-Schüler/innen im Vergleich zu den anderen Items unverhältnismäßig leichter fällt als dies für BHS-Schüler/innen der Fall ist. Die geschätzte Itemschwierigkeit dieser Aufgabe fällt also für Testpersonen aus Handelsakademien und Höheren Technischen Lehranstalten wesentlich höher aus als für Personen, die ein Gymnasium besuchen. Dies lässt sich vermutlich dadurch erklären, dass Item 27 inhaltlich dem Bereich der Psychologie zugeordnet werden kann und Psychologie nur in allgemeinbildenden höheren Schulen, jedoch nicht in Handelsakademien und Höheren Technischen Lehranstalten unterrichtet wird. Zusätzlich zu dieser starken Auffälligkeit bezüglich des Teilungskriteriums „Schultyp“ zeigen sich auch deutlich unterschiedliche Itemschwierigkeitsparameterschätzungen bei männlichen und weiblichen Testpersonen (vgl. Abbildung 15). Frauen dürften hier aufgrund des oftmals stärker ausgeprägten Interesses für psychologische oder sozialwissenschaftliche Inhalte bei der Lösung des Items einen Vorteil haben.

Item 27 wird daher aufgrund der inhaltlichen Problematik und der starken Auffälligkeit in der grafischen Modellkontrolle als erstes aus dem Itempool ausgeschieden.

Die nach Ausschluss von Item 27 durchgeführten Likelihood-Ratio-Tests fallen in den Teilungskriterien „Score“ und „Schultyp“ erneut signifikant aus. Im Teilungskriterium „Geschlecht“ ergibt sich nun ein nicht signifikantes Ergebnis (siehe Tabelle 11).

Tabelle 11: Ergebnisse der Likelihood-Ratio-Tests
(*Sprachliche Analogien*, nach Ausschluss von Item 27)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	39,722	19	32,191	0,004
Geschlecht	34,383	19	32,191	0,017
Schultyp	131,006	19	32,191	0,000

Basierend auf den Ergebnissen der grafischen Modellkontrollen und der z-Tests, die für die nächsten Berechnungsdurchgänge nicht mehr im Detail angeführt werden, sowie vor allem auch unter Berücksichtigung möglicher inhaltlicher Auffälligkeiten, wurden in den nächsten Schritten sukzessive die Items 8, 41, 25 und 36 aus dem Itempool entfernt.

Item 8, das inhaltlich dem Bereich Mathematik angehört, wurde in allen drei Teilungskriterien grafisch auffällig. Für BHS-Schüler/innen weist es im Vergleich zu AHS-Schüler/innen eine deutlich niedrigere Itemschwierigkeit auf (vgl. Abbildung 16), was sich möglicherweise dadurch erklären lässt, dass mathematische Kenntnisse vor allem bei HTL-Schüler/innen im Vergleich zu manch anderen Wissensgebieten stärker ausgeprägt sein dürften als dies beispielsweise bei AHS-Schüler/innen der Fall ist. Im Teilungskriterium „Geschlecht“ zeigte sich überraschenderweise eine höhere Itemparameterschätzung für Männer als für Frauen (vgl. Abbildung 15). Des Weiteren fiel es leistungstärkeren Personen im Verhältnis zu den übrigen Items schwerer als leistungsschwächeren (vgl. Abbildung 14).

Die nach Entfernung von Item 8 durchgeführten Likelihood-Ratio-Tests ergeben erneut ein signifikantes Ergebnis für die Teilungskriterien „Score“ und „Schultyp“ (siehe Tabelle 12).

Tabelle 12: Ergebnisse der Likelihood-Ratio-Tests
(*Sprachliche Analogien*, nach Ausschluss der Items 27 und 8)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	38,458	18	34,805	0,003
Geschlecht	22,813	18	34,805	0,198
Schultyp	105,955	18	34,805	0,000

Das als nächstes ausgeschiedene Item 41 stammt aus dem Themenbereich Medizin bzw. Biologie und fällt Gymnasiast/innen deutlich leichter als HAK- und HTL-Schüler/innen (vgl. Abbildung 16), was sich inhaltlich wiederum auf den stärkeren Fokus auf den biologischen Bereich in Gymnasien im Gegensatz zu den anderen Schultypen zurückführen lässt. Zudem ist die geschätzte Itemschwierigkeit hier bei leistungsschwächeren Personen wesentlich höher als bei leistungsstarken (vgl. Abbildung 14).

In den nach Ausschluss von Item 41 erneut durchgeführten Likelihood-Ratio-Tests zeigt sich neben dem Teilungskriterium „Geschlecht“ nun auch im Teilungskriterium „Score“ kein signifikantes Ergebnis mehr, während bezüglich des „Schultyps“ weiterhin Modellabweichungen zu beobachten sind (siehe Tabelle 13).

Tabelle 13: Ergebnisse der Likelihood-Ratio-Tests
(*Sprachliche Analogien*, nach Ausschluss der Items 27, 8 und 41)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	30,392	16*	32,000	0,016
Geschlecht	22,449	17	33,409	0,168
Schultyp	79,402	17	33,409	0,000

*Für Item 23 konnten in diesem Berechnungsdurchgang für dieses Teilungskriterium keine Itemparameter geschätzt werden, da es von allen Personen gelöst wurde, die einen Rohwert größer dem Median erreicht haben.

Für die Eliminierung des Items 25 im nächsten Berechnungsdurchgang wurde wiederum aufgrund der hohen Differenz der Itemparameterschätzungen der beiden Teilstichproben des Teilungskriteriums „Schultyp“ entschieden. Die Aufgabe, mit der betriebswirtschaftliches Wissen überprüft wird, fiel BHS-Schüler/innen im Vergleich zu den übrigen Items unverhältnismäßig leichter als dies für AHS-Schüler/innen der Fall war; wahrscheinlich stellt der wirtschaftliche Bezug hier einen Vorteil für HAK-Schüler/innen dar.

Nach Ausschluss von Item 25 ergibt sich wieder ein signifikantes Ergebnis im LRT für das Teilungskriterium „Schultyp“, wie aus Tabelle 14 ersichtlich ist.

Tabelle 14: Ergebnisse der Likelihood-Ratio-Tests
(*Sprachliche Analogien*, nach Ausschluss der Items 27, 8, 41 und 25)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	22,778	16	32,000	0,120
Geschlecht	22,239	16	32,000	0,136
Schultyp	50,419	16	32,000	0,000

Nachdem im nächsten Schritt noch Item 36 ausgeschlossen wurde, das für Gymnasiast/innen vermutlich aufgrund des sozialwissenschaftlichen Inhalts sowie der Verwendung von Fremdwörtern leichter geschätzt wurde als für Schüler/innen aus Handelsakademien und Höheren Technischen Lehranstalten, fielen die Likelihood-Ratio-Tests nun in allen drei Teilungskriterien nicht mehr signifikant aus (siehe Tabelle 15).

Die grafischen Modellkontrollen für den letzten Berechnungsdurchgang können den Abbildungen 17 bis 19 entnommen werden. Obwohl in den Teilungskriterien „Score“ und „Geschlecht“ Item 40 und im Teilungskriterium „Schultyp“ Item 39 grafisch noch immer auffällig sind, werden keine weiteren Items ausgeschieden, da aufgrund der Ergebnisse der Likelihood-Ratio-Tests von keinen signifikanten Modellabweichungen mehr auszugehen ist.

Tabelle 15: Ergebnisse der Likelihood-Ratio-Tests
(*Sprachliche Analogien*, nach Ausschluss der Items 27, 8, 41, 25 und 36)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	23,860	15	30,578	0,068
Geschlecht	17,813	15	30,578	0,273
Schultyp	24,367	15	30,578	0,059

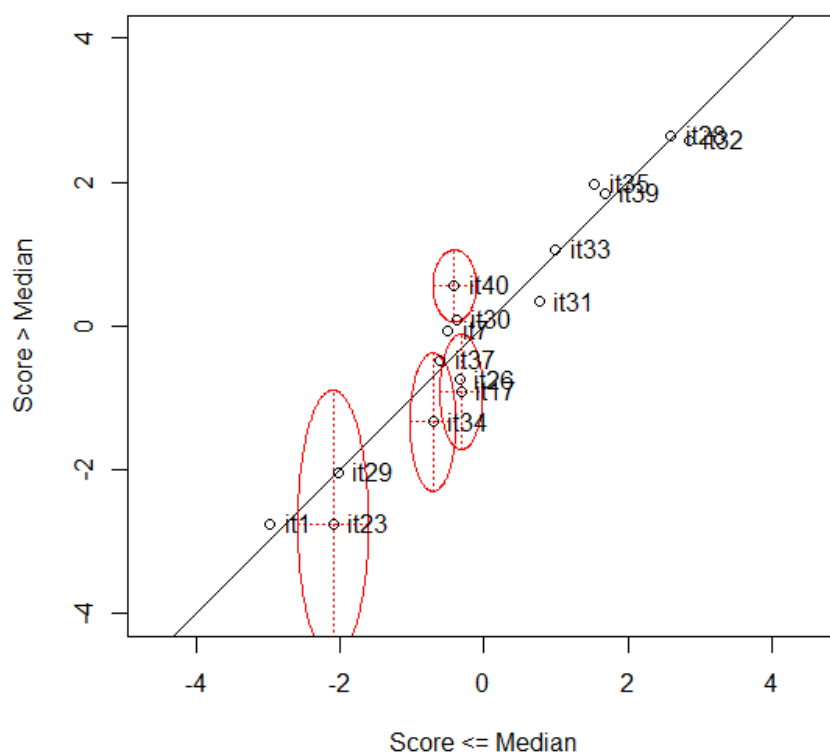


Abbildung 17: Grafische Modellkontrolle für das Teilungskriterium „Score“
(*Sprachliche Analogien*, nach Ausschluss der Items 27, 8, 41, 25 und 36)

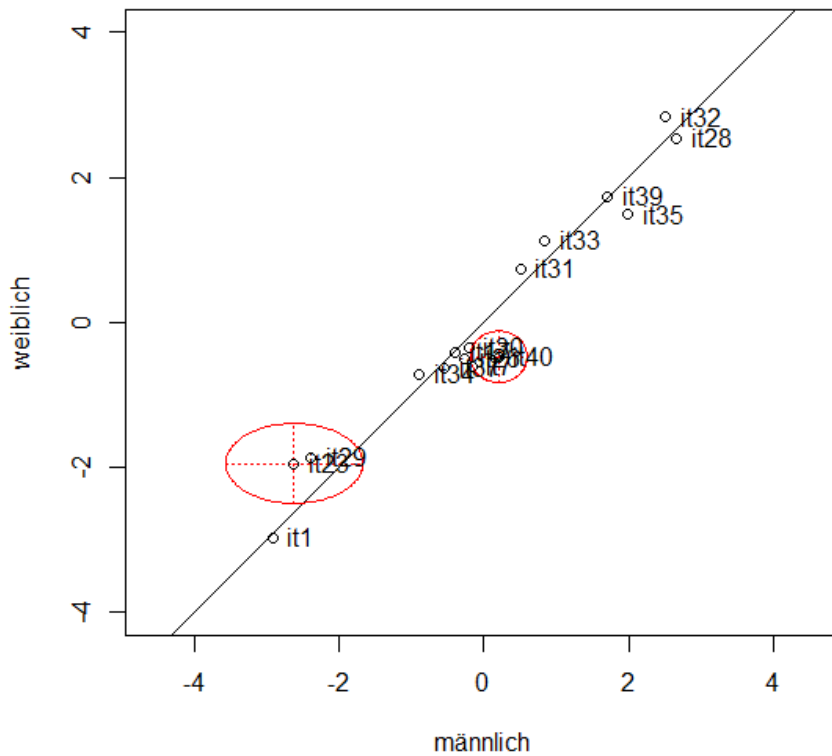


Abbildung 18: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
(*Sprachliche Analogien*, nach Ausschluss der Items 27, 8, 41, 25 und 36)

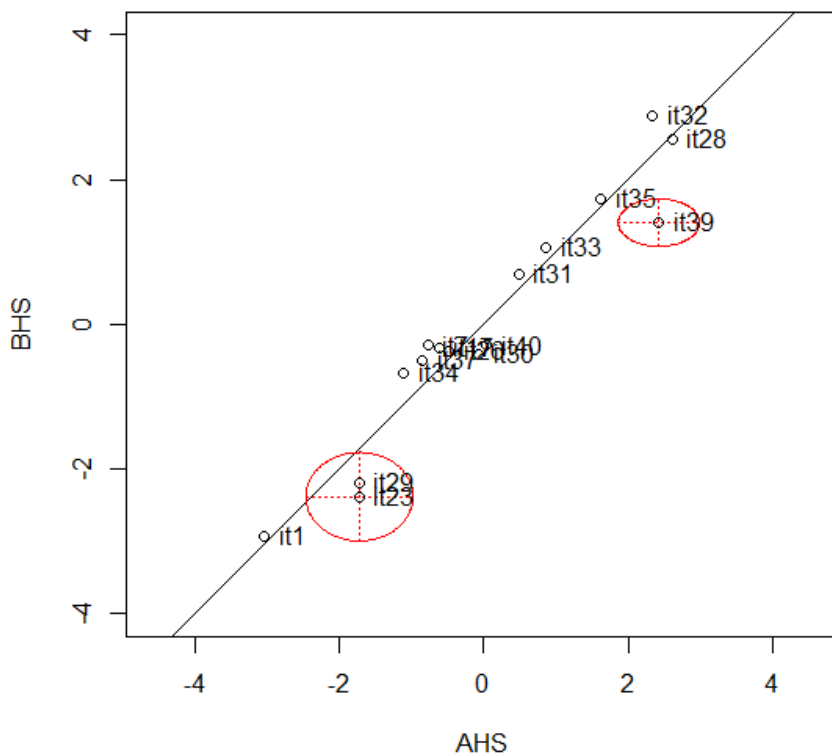


Abbildung 19: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(*Sprachliche Analogien*, nach Ausschluss der Items 27, 8, 41, 25 und 36)

Somit können die verbleibenden 16 Items des Tests *Sprachliche Analogien* als mit dem dichotomen logistischen Modell von Rasch konform angesehen werden und für das Verfahren kann a-posteriori Modellgeltung angenommen werden. Es kann nun also davon ausgegangen werden, dass der Test *Sprachliche Analogien* eindimensional die Fähigkeit Reasoning im kristallisierten Intelligenzbereich misst.

8.1.3. Ergebnisse der Rasch-Modell-Analysen beim *Färbigen Matrizentest*

Vom *FMT* (Undeutsch, 2010a) wurden 18 Items auf Rasch-Modell-Geltung überprüft ($n = 262$).

Wie in Tabelle 16 dargestellt ist, fallen die Likelihood-Ratio-Tests nach Andersen in allen Teilungskriterien nicht signifikant aus. Die Ergebnisse der grafischen Modellkontrollen sind den Abbildungen 20, 21 und 22 zu entnehmen. Es werden keine Items ausgeschlossen, da aufgrund der Ergebnisse der LRTs von keiner signifikanten Modellabweichung auszugehen ist.

Tabelle 16: Ergebnisse der Likelihood-Ratio-Tests
(*FMT*, alle Items)

Teilungskriterium	χ^2 (<i>LRT</i>)	<i>df</i>	χ^2_{krit}	<i>p</i>
<i>Score</i>	26,954	17	33,409	0,059
<i>Geschlecht</i>	24,268	17	33,409	0,112
<i>Schultyp</i>	31,723	17	33,409	0,016

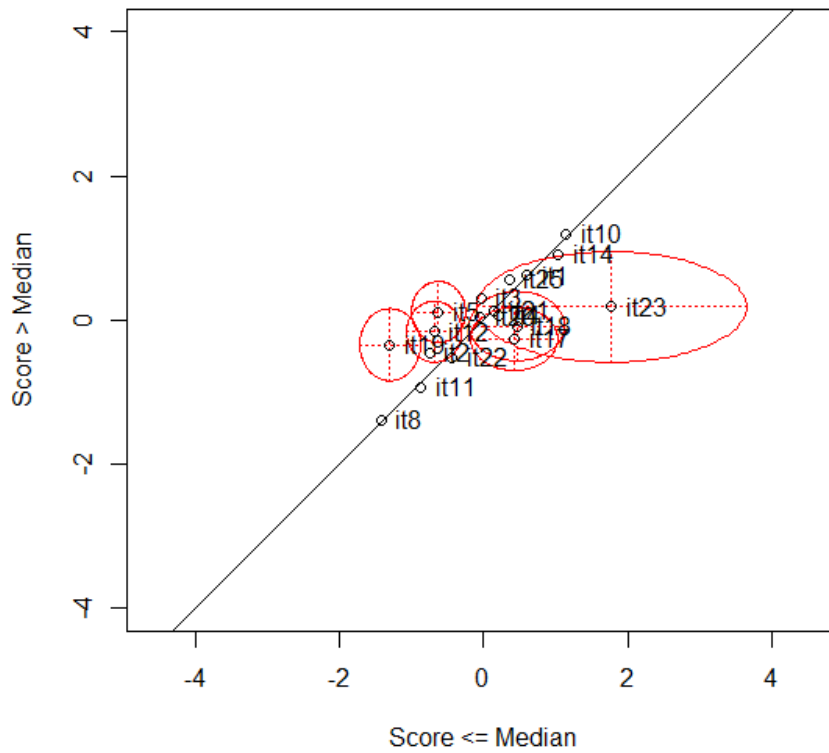


Abbildung 20: Grafische Modellkontrolle für das Teilungskriterium „Score“
(FMT, alle Items)

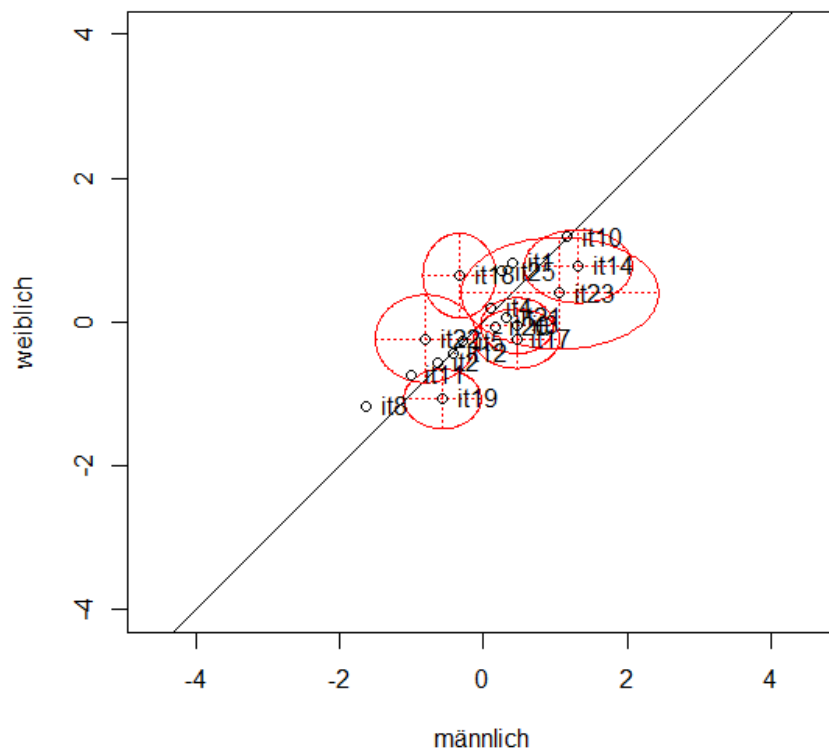


Abbildung 21: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
(FMT, alle Items)

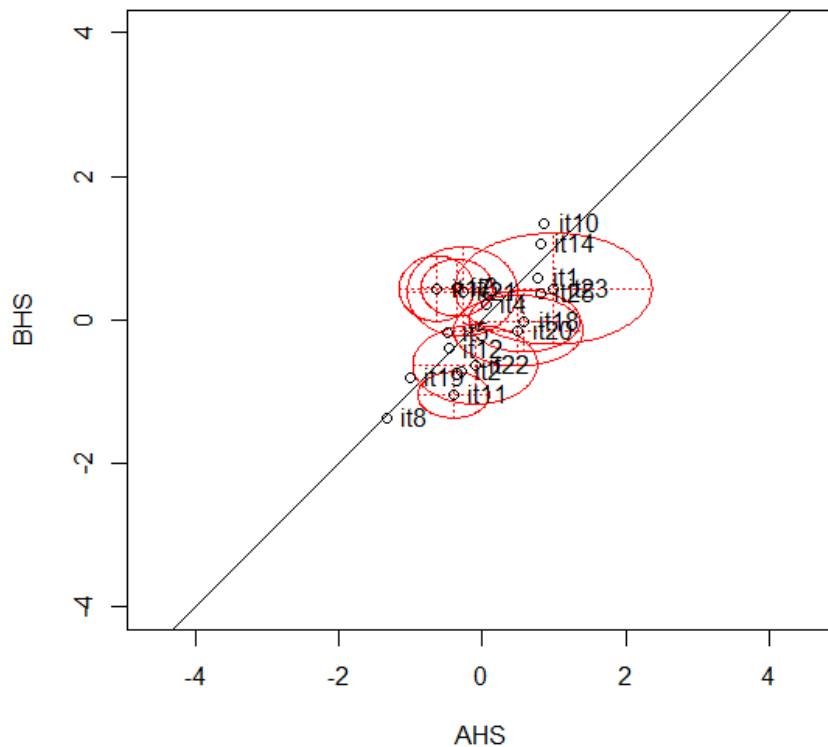


Abbildung 22: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(*FMT*, alle Items)

Somit kann den 18 überprüften Items des *Färbigen Matrizentests* Rasch-Modell-Geltung zugesprochen werden. Das Verfahren erfasst also eindimensional logisch-schlussfolgerndes Denken.

8.1.4. Ergebnisse der Rasch-Modell-Analysen bei *N-TOP*

15 Items des Tests *Numerische Topologien* (Kubinger & Heuberger, in Vorb.) wurden dahingehend überprüft, ob sie mit dem dichotomen logistischen Modell von Rasch konform sind ($n = 264$). Die Itemnummerierung entspricht jener aus Ünal (2014) und Gamsjäger (2012).

Bei den Berechnungen für das Teilungskriterium „Score“ konnten die Items 1b und 19a nicht geschätzt werden, da diese von allen Testpersonen der Teilstichprobe „Rohwert > Median“ gelöst wurden, und wurden daher in den entsprechenden Analysen nicht berücksichtigt.

Wie Tabelle 17 zu entnehmen ist, fallen die Likelihood-Ratio-Tests nach Andersen für alle Teilungskriterien nicht signifikant aus.

Tabelle 17: Ergebnisse der Likelihood-Ratio-Tests
(*N-TOP*, alle Items)

Teilungskriterium	χ^2 (<i>LRT</i>)	<i>df</i>	χ^2_{krit}	<i>p</i>
<i>Score</i>	24,102	12	26,217	0,020
<i>Geschlecht</i>	19,472	14	29,141	0,148
<i>Schultyp</i>	27,981	14	29,141	0,014

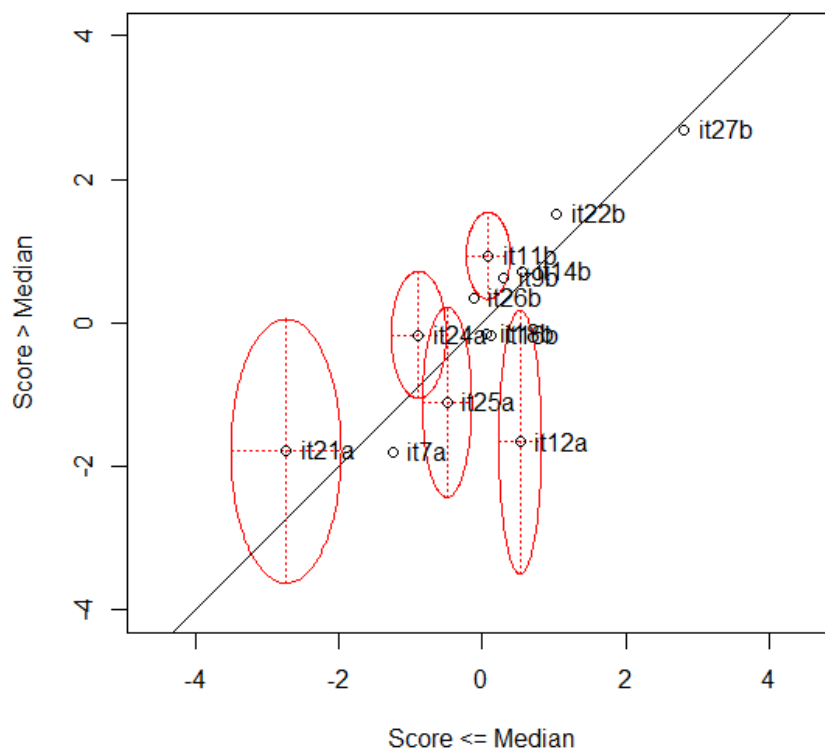


Abbildung 23: Grafische Modellkontrolle für das Teilungskriterium „Score“
(*N-TOP*, alle Items)

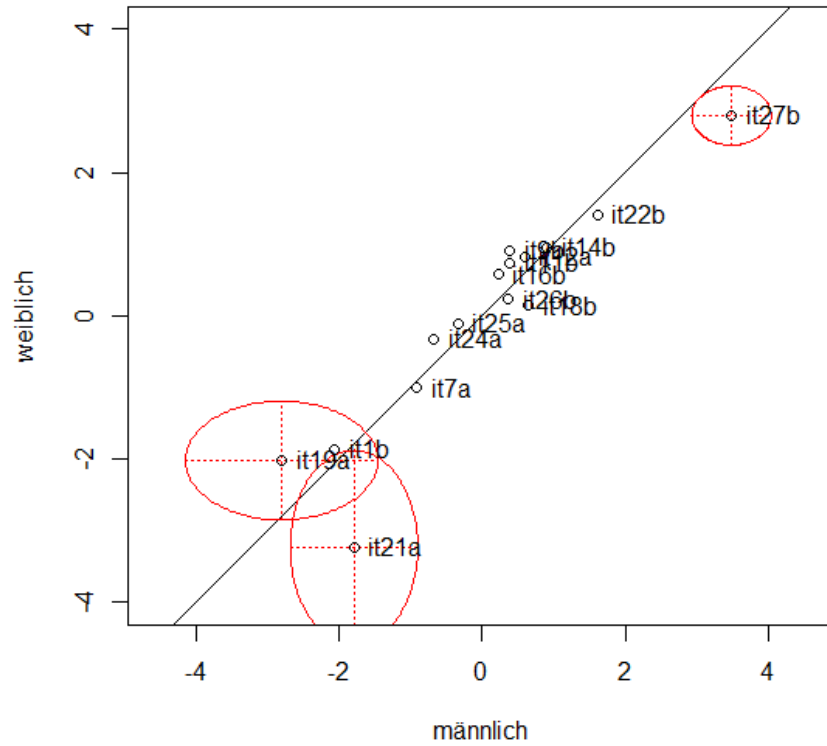


Abbildung 24: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
(*N-TOP*, alle Items)

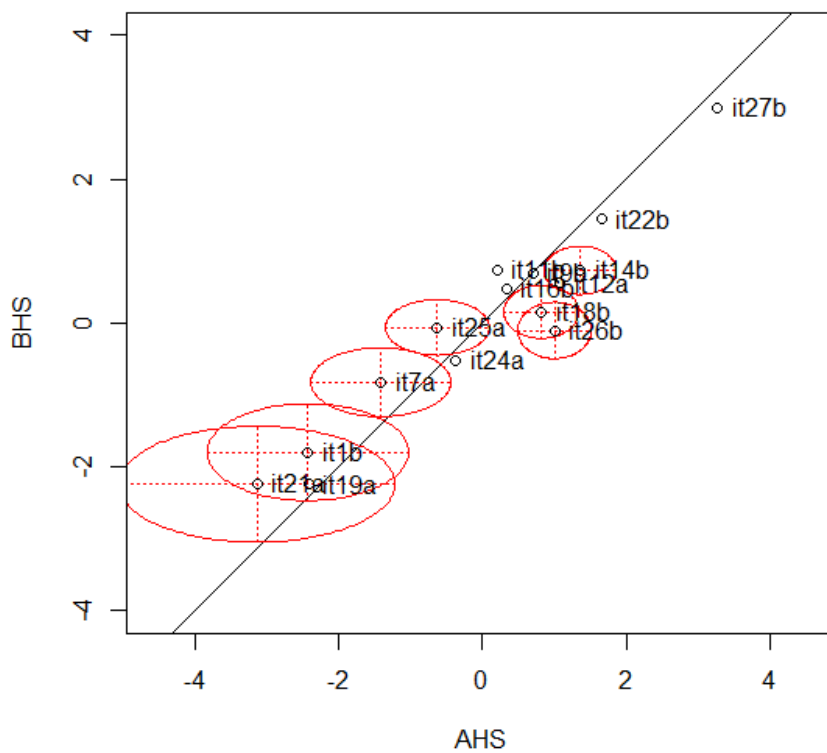


Abbildung 25: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(*N-TOP*, alle Items)

Die Ergebnisse der grafischen Modellkontrollen für die drei Teilungskriterien sind in den Abbildungen 23, 24 und 25 dargestellt. Obwohl die grafische Modellkontrolle nur für das Teilungskriterium „Geschlecht“ vollständig zufriedenstellend ist, werden keine Items ausgeschlossen, da die Likelihood-Ratio-Tests Rasch-Modell-Geltung für die überprüften Items von *N-TOP* ergeben. Somit kann den 15 untersuchten Items des Verfahrens *N-TOP* zugesprochen werden, dass sie eindimensional das Konstrukt Reasoning erfassen.

8.1.5. Ergebnisse der Rasch-Modell-Analysen bei *Figurale Analogien*

Der Itempool der auf das Rasch-Modell zu überprüfenden Items von *Figurale Analogien* (Ünal & Kubinger, in Vorb.) besteht aus 24 Items ($n = 261$).

Item 9 konnte in keinem der drei Teilungskriterien geschätzt werden, da dieses von nur einer Person der gesamten Stichprobe nicht gelöst wurde. Für das Teilungskriterium „Score“ konnten für einige weitere Items keine Itemparameter geschätzt werden, weil diese von keiner oder allen Personen von zumindest einer Teilstichprobe gelöst wurden. Die nicht berücksichtigten Items sowie die jeweiligen Berechnungsdurchgänge, in denen diese nicht geschätzt werden konnten, sind in Tabelle 18 angeführt.

Tabelle 18: Bei den Rasch-Modell-Analysen nicht berücksichtigte Items von *Figurale Analogien*

Teilungskriterium	Itemnummer	Berechnungsdurchgang (von insgesamt 6)
<i>Score</i>	9	1-6
	26	2-6
	27	2-4
	20	3-6
	18	6
<i>Geschlecht</i>	9	1-6
<i>Schultyp</i>	9	1-6

Der Likelihood-Ratio-Test nach Andersen ergibt, wie aus Tabelle 19 hervorgeht, für alle Teilungskriterien ein signifikantes Ergebnis.

Tabelle 19: Ergebnisse der Likelihood-Ratio-Tests
(*Figurale Analogien*, alle Items)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	81,059	22	40,289	0,000
Geschlecht	82,275	22	40,289	0,000
Schultyp	45,529	22	40,289	0,002

Die grafischen Modellkontrollen für die drei Teilungskriterien finden sich in den Abbildungen 26, 27 und 28. Dabei wurden im Teilungskriterium „Score“ die Items 2, 6, 7 und 10 auffällig, im Teilungskriterium „Geschlecht“ die Items 2, 7, 23 und 24 sowie im Teilungskriterium „Schultyp“ die Items 5 und 12. Die zugehörigen Ergebnisse des z-Tests nach Fischer und Scheiblechner sind in den Tabellen 20, 21 bzw. 22 aufgeführt. Diese bestätigen die Auffälligkeit der Items 2, 7 und 10 für das Teilungskriterium „Score“, der Items 2 und 7 für das Teilungskriterium „Geschlecht“ sowie der Items 5 und 12 für das Teilungskriterium „Schultyp“.

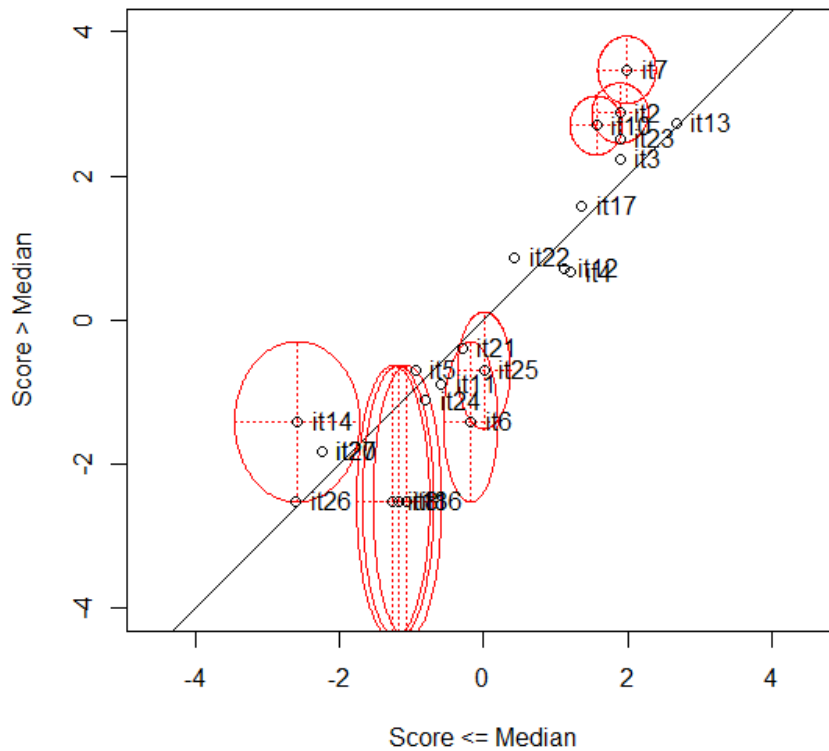


Abbildung 26: Grafische Modellkontrolle für das Teilungskriterium „Score“
(*Figurale Analogien*, alle Items)

Tabelle 20: Ergebnisse des z-Tests für das Teilungskriterium „Score“
(*Figurale Analogien*, alle Items)

Itemnummer	z-Wert	p
2	3,362	0,001
3	1,192	0,233
4	-1,764	0,078
5	0,518	0,604
6	-2,049	0,040
7	4,717	0,000
8	-1,336	0,181
10	4,078	0,000
11	-0,596	0,551
12	-1,307	0,191

13	0,194	0,847
14	1,632	0,103
16	-1,467	0,142
17	0,761	0,446
18	-1,266	0,205
20	0,545	0,586
21	-0,245	0,807
22	1,496	0,135
23	2,156	0,031
24	-0,560	0,575
25	-1,546	0,122
26	0,082	0,935
27	0,542	0,588

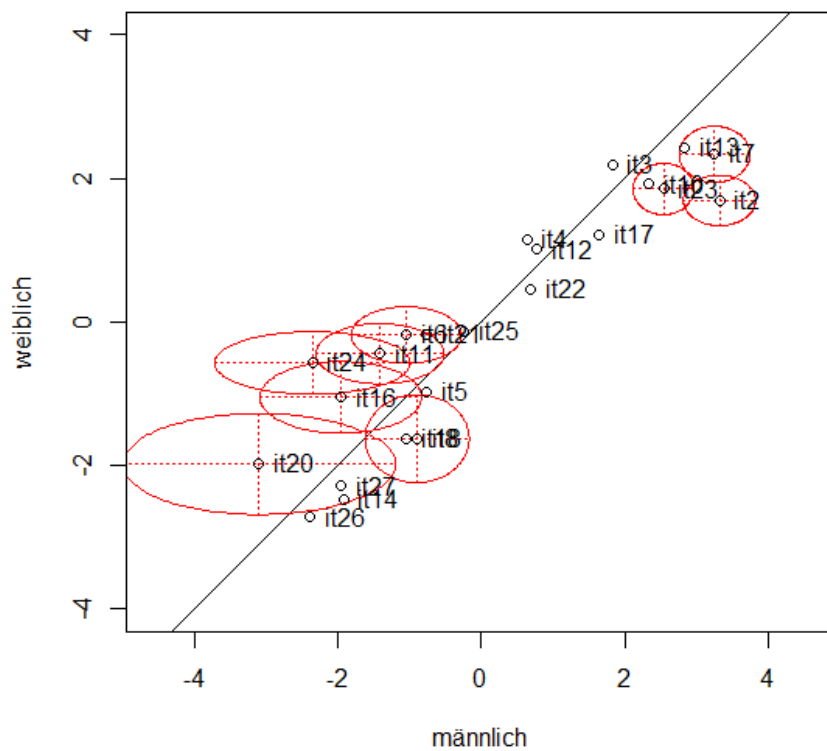


Abbildung 27: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
(*Figurale Analogien*, alle Items)

Tabelle 21: Ergebnisse des z-Tests für das Teilungskriterium „Geschlecht“
(Figurale Analogien, alle Items)

Itemnummer	z-Wert	p
2	-5,197	0,000
3	1,306	0,192
4	1,791	0,073
5	-0,511	0,609
6	1,959	0,050
7	-2,846	0,004
8	-1,503	0,133
10	-1,510	0,131
11	1,935	0,053
12	0,808	0,419
13	-1,344	0,179
14	-0,763	0,446
16	1,450	0,147
17	-1,611	0,107
18	-1,154	0,249
20	1,082	0,279
21	1,432	0,152
22	-0,764	0,445
23	-2,423	0,015
24	2,446	0,014
25	0,311	0,756
26	-0,390	0,696
27	-0,461	0,645

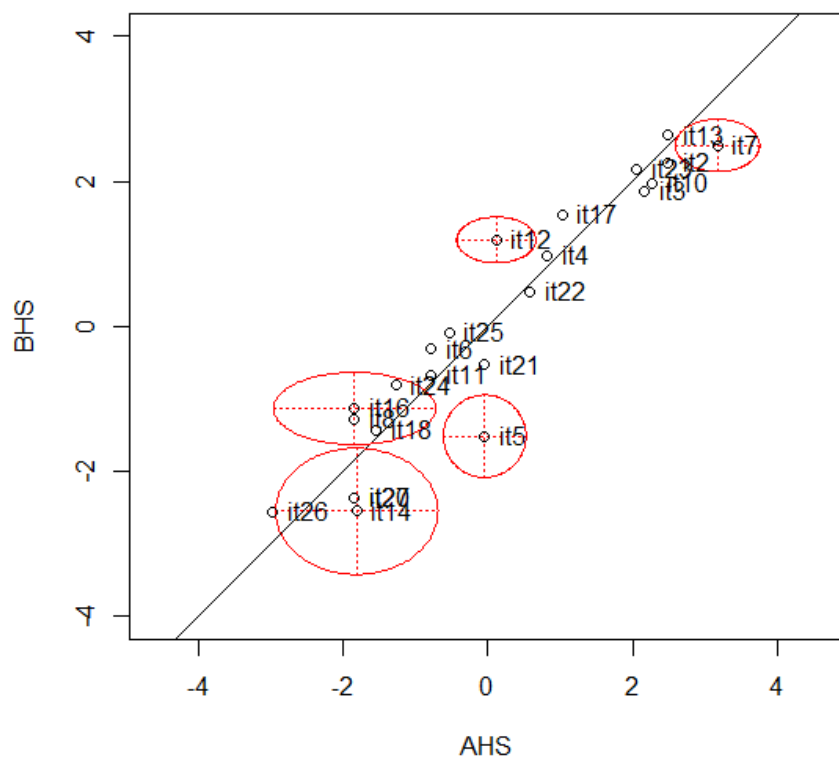


Abbildung 28: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(*Figurale Analogien*, alle Items)

Tabelle 22: Ergebnisse des z-Tests für das Teilungskriterium „Schultyp“
(*Figurale Analogien*, alle Items)

Itemnummer	z-Wert	p
2	-0,757	0,449
3	-0,995	0,320
4	0,528	0,597
5	-3,573	0,000
6	1,122	0,262
7	-1,927	0,054
8	0,910	0,363
10	-0,981	0,327
11	0,267	0,790
12	3,351	0,001

13	0,471	0,637
14	-1,017	0,309
16	1,142	0,253
17	1,795	0,073
18	0,199	0,842
20	-0,745	0,456
21	-1,349	0,177
22	-0,323	0,747
23	0,389	0,697
24	0,897	0,370
25	1,159	0,246
26	0,386	0,700
27	-0,742	0,458

Inhaltlich erklärbar ist am ehesten die Auffälligkeit der Items 2 und 24 im Teilungskriterium „Geschlecht“. Item 2 fiel weiblichen Personen im Vergleich zu den übrigen Items unverhältnismäßig leichter als dies bei männlichen Personen der Fall war, was sich vermutlich dadurch erklären lässt, dass in dieser Aufgabe haushaltstechnisches Wissen eine Rolle spielt. Bei Item 24 verhält es sich umgekehrt; hier wird Wissen aus dem Themenbereich Chemie abgefragt.

Aufgrund der möglichen inhaltlichen Gründe und der Tatsache, dass Item 2 auch im Teilungskriterium „Score“ grafisch auffällig ist, wird diese Aufgabe als erste aus dem Itempool entfernt.

Nach Ausschluss von Item 2 ergibt sich im Likelihood-Ratio-Test weiterhin für alle drei Teilungskriterien ein signifikantes Ergebnis (siehe Tabelle 23).

Tabelle 23: Ergebnisse der Likelihood-Ratio-Tests
(*Figurale Analogien*, nach Ausschluss von Item 2)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	56,609	19	36,191	0,000
Geschlecht	55,120	21	38,932	0,000
Schultyp	44,937	21	38,932	0,002

Basierend auf den nach Ausschluss von Item 2 durchgeführten grafischen Modellkontrollen, den Ergebnissen der z-Tests, welche ab diesem Berechnungsdurchgang jeweils nicht mehr genau dargestellt werden, sowie aufgrund von inhaltlichen Analysen wurden in den nächsten Schritten sukzessive die Items 7, 24, 5 und 23 aus dem Itempool entfernt.

Für Item 7, in dem physikalische Kenntnisse abgefragt werden, fielen die Itemschwierigkeitsparameterschätzungen für männliche Personen höher aus als für weibliche, wofür sich inhaltlich nur schwer eine Erklärung finden lässt. Dennoch wurde diese Aufgabe ausgeschieden, da sich auch eine grafische Auffälligkeit in den Teilungskriterien „Score“ und „Schultyp“ zeigt.

Die nachfolgenden Likelihood-Ratio-Tests ergeben für alle drei Teilungskriterien wiederum ein signifikantes Ergebnis (siehe Tabelle 24).

Tabelle 24: Ergebnisse der Likelihood-Ratio-Tests
(*Figurale Analogien*, nach Ausschluss der Items 2 und 7)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	35,826	17	33,409	0,005
Geschlecht	46,780	20	37,566	0,001
Schultyp	40,673	20	37,566	0,004

Für das im nächsten Durchgang ausgeschlossene Item 24 zeigte sich, wie weiter oben bereits erwähnt, eine niedrigere Itemschwierigkeitsparameterschätzung für männliche Testpersonen, die sich inhaltlich möglicherweise durch stärker ausgeprägte Kenntnisse in Chemie bei diesen Personen erklären lässt, womit sie im Vergleich zu Frauen einen Vorteil bei der Lösung dieser Aufgabe haben könnten.

Auch nach Ausschluss von Item 24 sind die Ergebnisse der Likelihood-Ratio-Tests weiterhin signifikant (siehe Tabelle 25).

Tabelle 25: Ergebnisse der Likelihood-Ratio-Tests
(*Figurale Analogien*, nach Ausschluss der Items 2, 7 und 24)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	37,047	16	32,000	0,002
Geschlecht	38,286	19	36,191	0,005
Schultyp	39,945	19	36,191	0,003

Im nächsten Schritt wurde Item 5 aus dem Itempool entfernt. Für die Tatsache, dass diese Aufgabe BHS-Schüler/innen im Vergleich zu den anderen Items leichter fiel als dies bei AHS-Schüler/innen der Fall war, lässt sich kein inhaltlicher Grund finden. Dennoch musste das Item aufgrund der starken Auffälligkeit ausgeschlossen werden.

Tabelle 26: Ergebnisse der Likelihood-Ratio-Tests
(*Figurale Analogien*, nach Ausschluss der Items 2, 7, 24 und 5)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	38,009	16	32,000	0,002
Geschlecht	38,120	18	34,805	0,004
Schultyp	25,397	18	34,805	0,114

Nachdem Item 5 ausgeschlossen wurde, zeigt sich nun beim LRT kein signifikantes Ergebnis mehr im Teilungskriterium „Schultyp“, jedoch sind die Ergebnisse in den Teilungskriterien „Score“ und „Geschlecht“ noch immer signifikant (siehe Tabelle 26).

Die letzte Aufgabe, die aus dem Itempool eliminiert werden musste, ist Item 23; diese stammt aus dem Themenbereich Religion. Obwohl die unterschiedlichen Itemparameterschätzungen für männliche und weibliche Personen inhaltlich schwer zu erklären sind, musste das Item ausgeschieden werden.

Nach Ausschluss von Item 23 fallen die Likelihood-Ratio-Tests in keinem Teilungskriterium mehr signifikant aus (siehe Tabelle 27), d. h. es ist nun von keinen signifikanten Modellabweichungen mehr auszugehen und es werden daher keine weiteren Items ausgeschieden. Auch die grafischen Modellkontrollen für den letzten Berechnungsdurchgang, welche den Abbildungen 29 bis 31 zu entnehmen sind, sind weitgehend zufriedenstellend.

Tabelle 27: Ergebnisse der Likelihood-Ratio-Tests
(*Figurale Analogien*, nach Ausschluss der Items 2, 7, 24, 5 und 23)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	17,060	14	29,141	0,253
Geschlecht	32,220	17	33,409	0,014
Schultyp	26,068	17	33,409	0,073

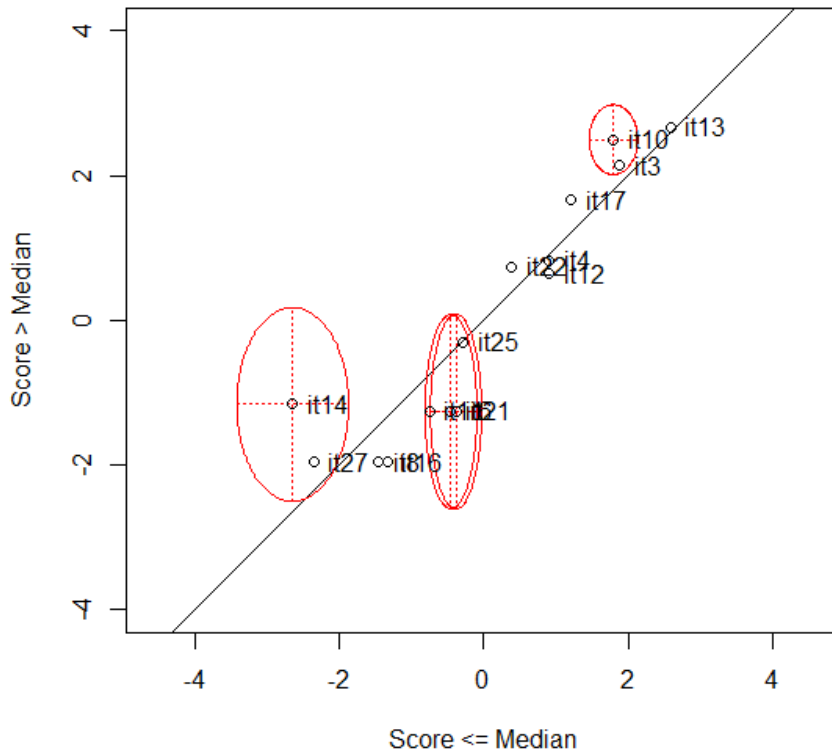


Abbildung 29: Grafische Modellkontrolle für das Teilungskriterium „Score“
(*Figurele Analogien*, nach Ausschluss der Items 2, 7, 24, 5 und 23)

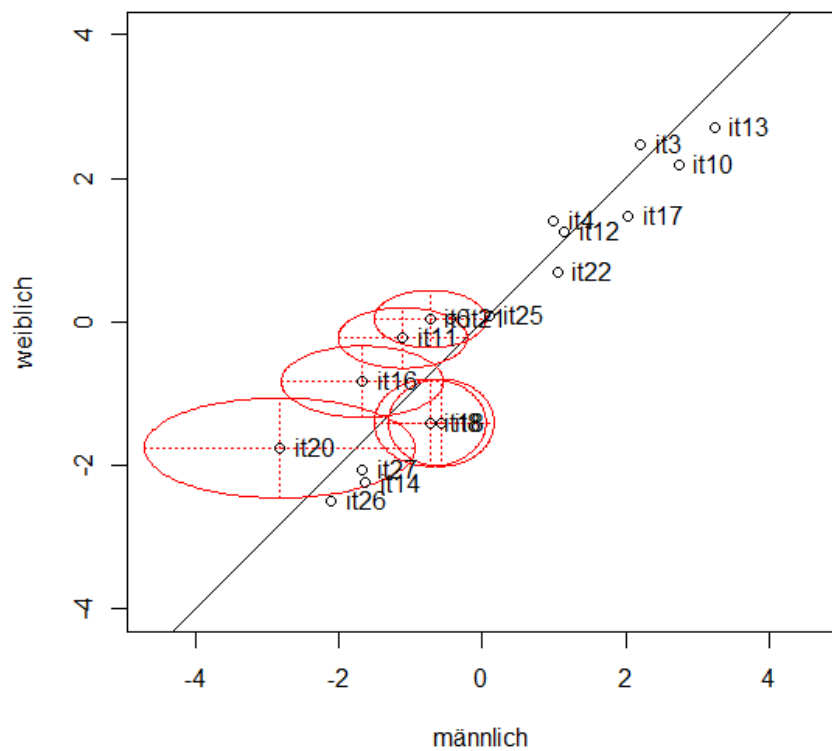


Abbildung 30: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
(*Figurele Analogien*, nach Ausschluss der Items 2, 7, 24, 5 und 23)

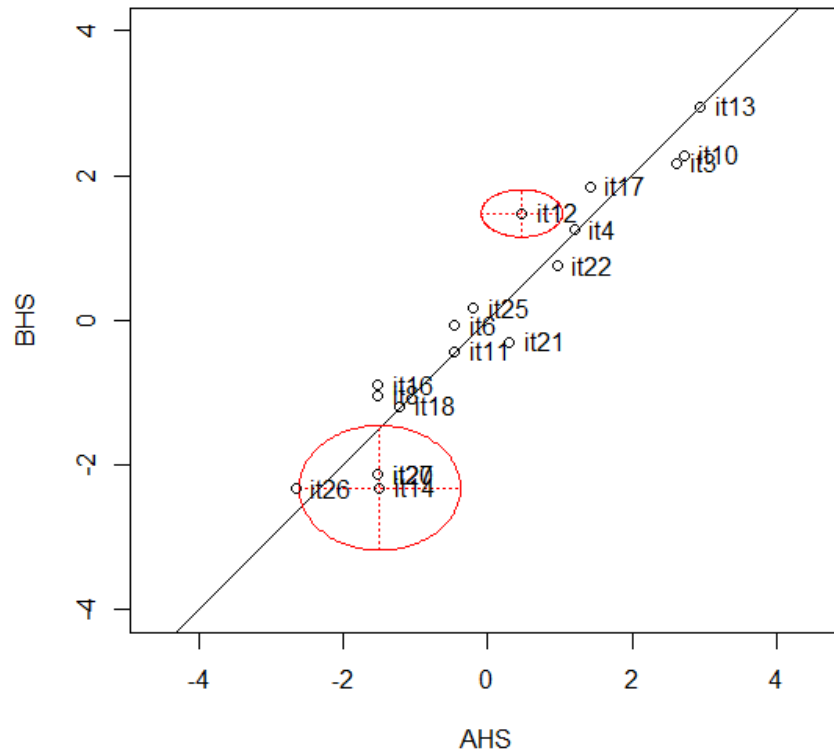


Abbildung 31: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(*Figurale Analogien*, nach Ausschluss der Items 2, 7, 24, 5 und 23)

Den verbleibenden Items von *Figurale Analogien* kann somit zugesprochen werden, dass sie Rasch-Modell-Geltung aufweisen und eindimensional kristallisiertes schlussfolgerndes Denken erfassen.

8.1.6. Ergebnisse der Rasch-Modell-Analysen bei *Gleichungen*

Vom Verfahren *Gleichungen* (Kubinger & Gamsjäger, in Vorb.) wurden 11 Items an einer Stichprobe von $n = 258$ Personen auf Geltung des dichotomen logistischen Modells von Rasch überprüft. Eines der ursprünglich den Testpersonen vorgegebenen 12 Items (Item 24) musste aufgrund eines festgestellten Konstruktionsfehlers aus den Analysen ausgeschlossen werden. Dieses muss aus dem Itempool von *Gleichungen* ausgeschieden werden.

Die Likelihood-Ratio-Tests nach Andersen ergeben für alle drei Teilungskriterien ein nicht signifikantes Ergebnis (siehe Tabelle 28).

Tabelle 28: Ergebnisse der Likelihood-Ratio-Tests
(Gleichungen, alle Items)

Teilungskriterium	χ^2 (LRT)	df	χ^2_{krit}	p
Score	21,221	10	23,209	0,020
Geschlecht	6,725	10	23,209	0,751
Schultyp	10,299	10	23,209	0,415

Für die Teilungskriterien „Geschlecht“ und „Schultyp“ sind die Ergebnisse der grafischen Modellkontrollen ebenfalls zufriedenstellend (vgl. Abbildung 33 bzw. 34). Im Teilungskriterium „Score“ sind die Items 3 und 39 grafisch auffällig (vgl. Abbildung 32); sie werden aber aufgrund fehlender inhaltlicher Gründe und – noch wichtiger – aufgrund des nicht signifikanten Ergebnisses des LRT nicht aus dem Itempool ausgeschieden.

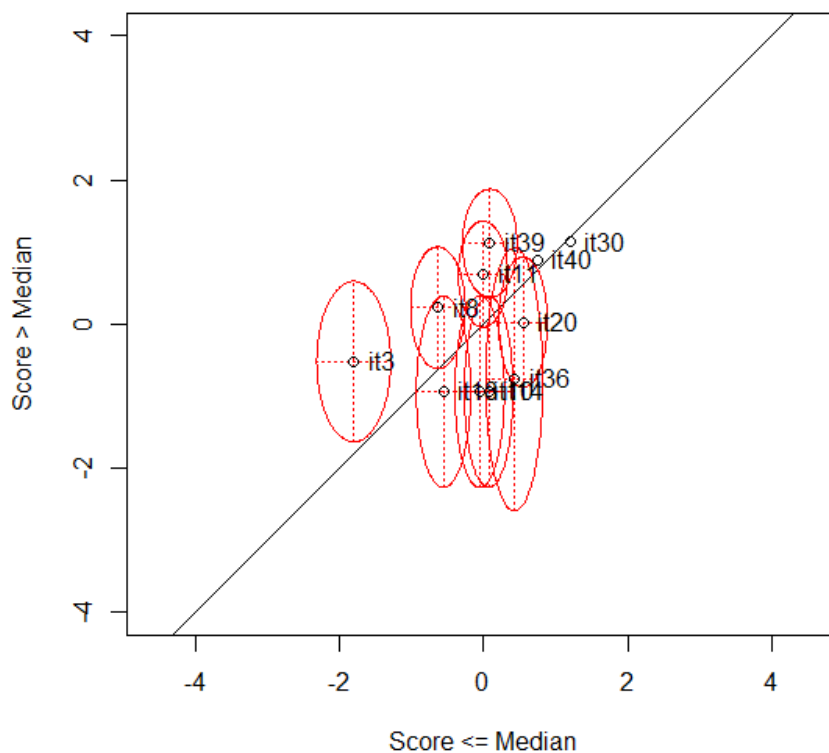


Abbildung 32: Grafische Modellkontrolle für das Teilungskriterium „Score“
(Gleichungen, alle Items)

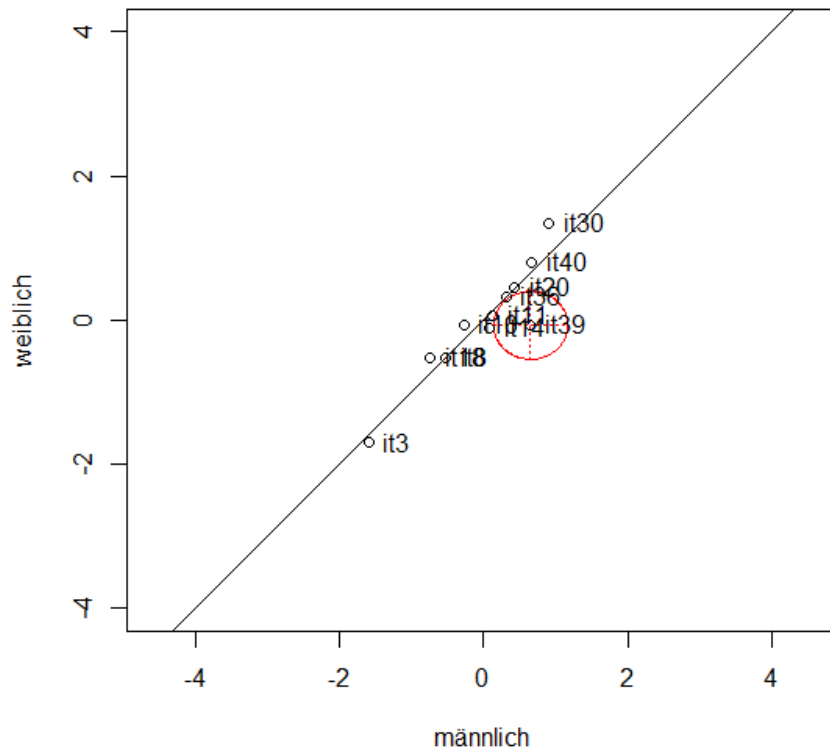


Abbildung 33: Grafische Modellkontrolle für das Teilungskriterium „Geschlecht“
(*Gleichungen*, alle Items)

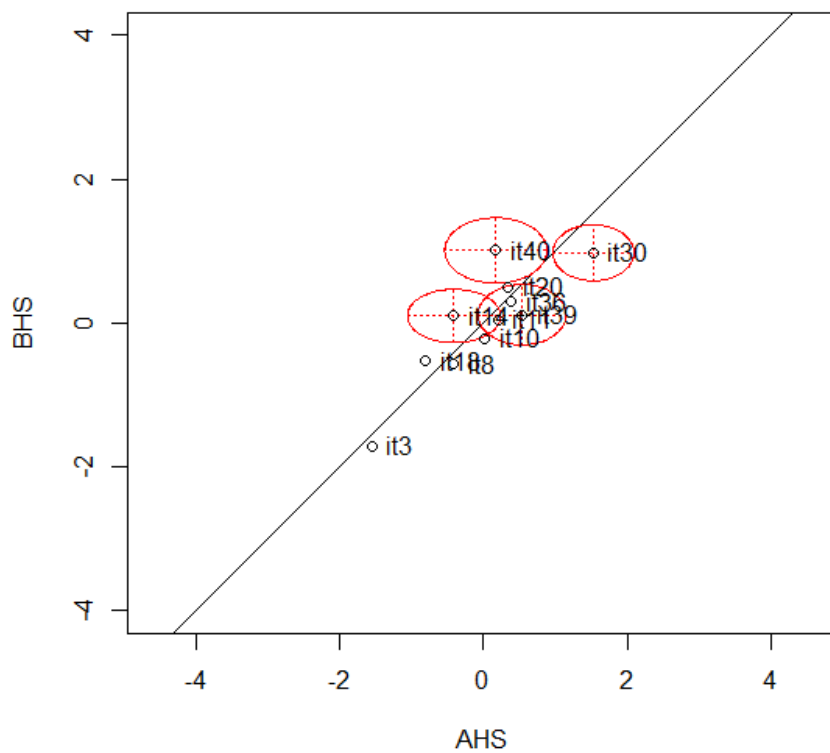


Abbildung 34: Grafische Modellkontrolle für das Teilungskriterium „Schultyp“
(*Gleichungen*, alle Items)

Die untersuchten Items des Verfahrens *Gleichungen* können somit als dem Rasch-Modell entsprechend beurteilt werden und messen eindimensional die Fähigkeit zum schlussfolgernden Denken im kristallisierten Intelligenzbereich.

8.2. Durchführung der exploratorischen Faktorenanalyse

Zur Bestimmung der Dimensionalität der Reasoning-Test-Batterie wurde eine Hauptkomponentenanalyse durchgeführt (vgl. dazu Abschnitt 3.3). Dafür wurden zunächst die Personenparameter pro Untertest für alle Testpersonen ermittelt, wobei diese für Personen, die keine oder alle Items eines Untertests gelöst hatten, extrapoliert wurden. Bei der Durchführung der Hauptkomponentenanalyse wurden nur Testpersonen berücksichtigt, die in keiner der Rasch-Modell-Analysen ausgeschlossen worden waren. Somit besteht die Stichprobe bei der Faktorenanalyse aus insgesamt $n = 255$ Schülerinnen und Schülern.

Zur Bestimmung der Faktorenzahl wurde einerseits das Kaiser-Kriterium herangezogen und andererseits der Scree-Test durchgeführt. Anschließend war, sofern für die Extraktion von mehr als einem Faktor entschieden würde, zur besseren Interpretierbarkeit der Faktoren eine Varimax-Rotation vorgesehen.

8.2.1. Ergebnisse der Hauptkomponentenanalyse

Die Hauptkomponentenanalyse wurde auf Basis der Korrelationsmatrix der Personenparameter der untersuchten Reasoning-Tests unter Ausschluss jener Items, die sich als nicht Rasch-Modell-konform erwiesen haben, durchgeführt. Die resultierenden Korrelationen zwischen den Verfahren sind in Kapitel 8.3 genauer dargestellt.

Bei der Hauptkomponentenanalyse entspricht die Anzahl der extrahierten Faktoren anfänglich der Anzahl der Variablen oder Tests, über die die Faktorenanalyse durchgeführt wird. Dementsprechend ergeben sich im vorliegenden Fall zu Beginn sechs Faktoren, deren Eigenwerte ebenso wie der prozentuelle Varianzanteil der Testbatterie, der durch sie jeweils erklärt wird, in Tabelle 29 dargestellt sind. Wie sich daraus ablesen lässt, weist nur ein Faktor einen deutlich höheren Eigenwert als 1 auf, während der Ei-

genwert des zweiten nur noch ganz knapp über 1 liegt und somit durch diesen kaum mehr Varianz erklärt wird als durch einen einzelnen Untertest der Testbatterie.

Auch der Scree-Plot in Abbildung 35 zeigt einen deutlichen Abfall der Eigenwerte nach dem ersten Faktor und einen Knick beim zweitgrößten Eigenwert. Auf Basis dieses Diagramms wäre also klar für eine Generalfaktorlösung zu entscheiden.

Tabelle 29: Ergebnis der Hauptkomponentenanalyse: Eigenwerte und erklärter Varianzanteil pro Faktor

<i>Faktor</i>	<i>Anfängliche Eigenwerte</i>		
	<i>Gesamt</i>	<i>Prozent der Varianz</i>	<i>Kumulierte Prozent</i>
1	2,307	38,45	38,45
2	1,037	17,28	55,73
3	0,815	13,58	69,31
4	0,678	11,30	80,61
5	0,631	10,52	91,13
6	0,532	8,87	100,00

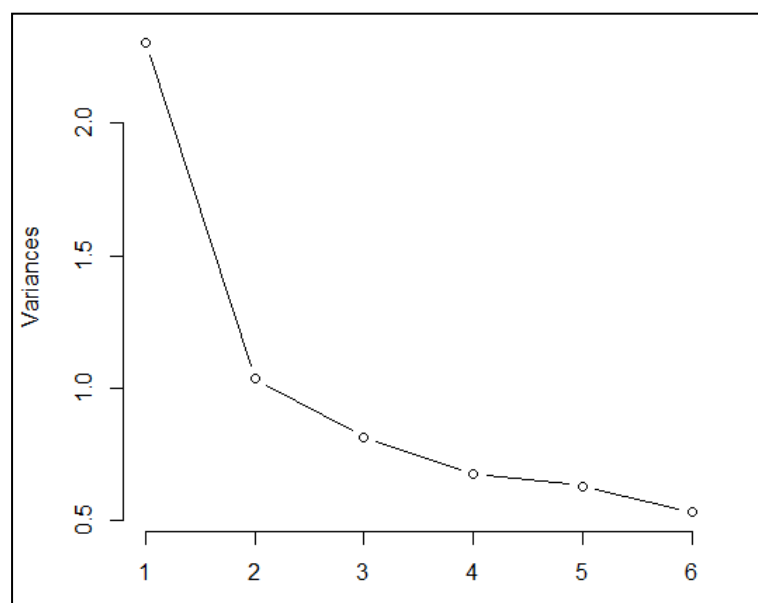


Abbildung 35: Scree-Plot der Eigenwerte (Hauptkomponentenanalyse)

Wie aus Tabelle 30 hervorgeht, wird der gemeinsame Faktor im Falle einer Generalfaktorlösung am stärksten durch den Test *Gleichungen* beschrieben, der eine Ladung von 0,72 aufweist. Die nächsthöchsten Ladungen zeigen sich bei den Verfahren *Formale Folgerichtigkeit* und *N-TOP* mit jeweils 0,65. Am niedrigsten laden der *Färbige Matrizentest* sowie *Sprachliche Analogien* auf dem extrahierten Faktor. Etwa die Hälfte der Varianz des Tests *Gleichungen* wird durch den Generalfaktor erklärt, während die Kommunalität beim *Färbigen Matrizentest* und bei *Sprachliche Analogien* nur jeweils 0,29 beträgt, was bedeutet, dass von diesen Verfahren nur etwas über ein Viertel der Varianz erklärt wird. Insgesamt ist der Faktor, der inhaltlich als gemeinsame Reasoning-Komponente verstanden werden könnte, die stark durch numerisches bzw. abstraktes Denken bestimmt ist, für 38 % der Varianz der Testleistungen in den verschiedenen Tests verantwortlich.

Tabelle 30: Ladungen und Kommunalitäten der Generalfaktorlösung

	Faktor 1	Kommunalitäten
<i>Formale Folgerichtigkeit</i>	0,65	0,43
<i>Sprachliche Analogien</i>	0,54	0,29
<i>Färbiger Matrizentest</i>	0,54	0,29
<i>N-TOP</i>	0,65	0,42
<i>Figurale Analogien</i>	0,60	0,36
<i>Gleichungen</i>	0,72	0,51
Eigenwert des Faktors	2,31	
Erklärte Varianz (%)	38	

Da der zweite Faktor einen Eigenwert von knapp über 1 aufweist, wird zusätzlich zur Generalfaktorlösung auch eine Zwei-Faktoren-Lösung betrachtet. Im Falle von zwei extrahierten Faktoren wird eine Faktorenrotation durchgeführt, wobei die Varimax-Rotation gewählt wird, welche sich dadurch auszeichnet, dass die Faktorenlösung orthogonal so rotiert wird, dass die Variabilität der Ladungen innerhalb jedes Faktors mög-

lichst groß wird und auch die einzelnen Variablen jeweils auf nur einem (oder wenigen) Faktoren hoch und auf dem bzw. den anderen eher niedrig laden.

Tabelle 31: Ladungen und Kommunalitäten der Varimax-rotierten
Zwei-Faktoren-Lösung

	Faktor 1	Faktor 2	Kommunalitäten
<i>Formale Folgerichtigkeit</i>	0,79	0,08	0,63
<i>Sprachliche Analogien</i>	0,03	0,80	0,64
<i>Färbiger Matrizentest</i>	0,73	-0,03	0,53
<i>N-TOP</i>	0,55	0,35	0,43
<i>Figurale Analogien</i>	0,15	0,77	0,61
<i>Gleichungen</i>	0,55	0,46	0,51
Eigenwerte der Faktoren	1,78	1,56	
Erklärte Varianz (%)	30	26	

Die Varimax-rotierte Zwei-Faktoren-Lösung der Reasoning-Test-Batterie ist in Tabelle 31 dargestellt. Dem ersten Faktor, der 30 % der Varianz der Testleistungen in den verschiedenen Reasoning-Tests erklärt, können die Verfahren *Formale Folgerichtigkeit*, *Färbiger Matrizentest*, *N-TOP* sowie *Gleichungen* zugeordnet werden, wobei ersteres mit 0,79 die höchste Ladung aufweist. Faktor 2, der für 26 % der Varianz verantwortlich ist, wird ähnlich stark durch die Tests *Sprachliche Analogien* und *Figurale Analogien* definiert; die Ladungen betragen hier 0,80 bzw. 0,77. Inhaltlich wird Faktor 1 also hauptsächlich durch Verfahren beschrieben, die Reasoning im fluiden Intelligenzbereich erfassen (mit Ausnahme des Tests *Gleichungen*, der aber auch relativ stark auf Faktor 2 lädt) und zudem nur figural-abstraktes oder numerisches Material verwenden. Im Gegensatz dazu laden auf Faktor 2 jene beiden Verfahren sehr stark, die kristallisiertes schlussfolgerndes Denken erfassen sollen und sich durch eine deutliche Wissenskomponente auszeichnen, gleichzeitig aber auch einen hohen sprachlichen Anteil aufweisen. Auch auf den Test *Gleichungen*, der auf diesem Faktor immerhin mit 0,46 lädt, treffen diese Eigenschaften größtenteils zu, da auch dieser dem kristallisierten Intelligenzbereich zuzuordnen ist. Ins-

gesamt kann durch die beiden latenten Faktoren gemeinsam etwas mehr als die Hälfte der Varianz der Testleistungen in den sechs Verfahren erklärt werden.

Zusammenfassend lässt sich die Reasoning-Test-Batterie also entweder durch einen Generalfaktor, nämlich die gemeinsame Reasoning-Komponente (*abstraktes Denken*), erklären, oder durch die folgenden zwei Faktoren:

- **Faktor 1: logisch-schlussfolgerndes Denken (fluid, figural-abstrakt/numerisch):** *Formale Folgerichtigkeit, Färbiger Matrizentest, N-TOP, Gleichungen*
- **Faktor 2: Allgemeinwissen (kristallisiertes schlussfolgerndes Denken, sprachlich):** *Sprachliche Analogien, Figurale Analogien*

8.3. Korrelative Zusammenhänge zwischen den Reasoning-Tests

Zusätzlich zu den soeben beschriebenen Ergebnissen der Faktorenanalyse sollen hier noch die linearen Zusammenhänge zwischen den untersuchten Reasoning-Tests beschrieben werden. Die Korrelationen wurden auf Basis der Personenparameter, die aufgrund der Geltung des Rasch-Modells als normalverteilt und intervallskaliert angesehen werden können, mittels Produkt-Moment-Korrelation nach Pearson berechnet.

Aus der Interkorrelationsmatrix in Tabelle 32 ist ersichtlich, dass sich die Korrelationskoeffizienten zwischen $r = 0,14$ und $r = 0,38$ bewegen. Es ergeben sich also durchwegs relativ niedrige positive Zusammenhänge zwischen den verschiedenen Tests. Das Verfahren *Formale Folgerichtigkeit* (abgekürzt in Tabelle: FF) weist die höchste Korrelation mit dem *Färbigen Matrizentest* (FMT) und dem Test *N-TOP* zu jeweils $r = 0,35$ sowie in ähnlicher Höhe zu *Gleichungen* (GL) auf. Beim *FMT* sind ansonsten nur noch niedrigere Korrelationen mit den übrigen Verfahren festzustellen. Die Personenparameter des Tests *Gleichungen* korrelieren am höchsten mit jenen von *N-TOP* zu $r = 0,38$, gefolgt von *Figurale Analogien* (FA), zu welchem der Zusammenhang ähnlich hoch ausfällt wie zu *Formale Folgerichtigkeit*. *Figurale Analogien* wiederum zeigt den größten Zusammenhang mit *Sprachliche Analogien* (SA) zu $r = 0,35$ und eine nur unwesentlich niedri-

gere Korrelation mit *Gleichungen*. Für *N-TOP* und *Sprachliche Analogien* wurden die höchsten Korrelationen bereits genannt.

Tabelle 32: Interkorrelationsmatrix der Reasoning-Tests

	FF	SA	FMT	N-TOP	FA	GL
FF	1	0,14	0,35	0,35	0,21	0,32
SA	0,14	1	0,14	0,22	0,35	0,25
FMT	0,35	0,14	1	0,17	0,15	0,27
N-TOP	0,35	0,22	0,17	1	0,22	0,38
FA	0,21	0,35	0,15	0,22	1	0,33
GL	0,32	0,25	0,27	0,38	0,33	1

Der stärkste Zusammenhang innerhalb der gesamten Korrelationsmatrix ist zwischen *Gleichungen* und *N-TOP* zu beobachten. Das Bestimmtheitsmaß beträgt in diesem Fall $B = 0,14$, was bedeutet, dass nur etwa 14 % der Varianz der Personenparameter des Tests *N-TOP* durch *Gleichungen* erklärt wird und umgekehrt. Damit ist auch der stärkste Zusammenhang innerhalb der untersuchten Reasoning-Test-Batterie mit Kubinger et al. (2011, S. 359) praktisch als kaum relevant zu betrachten. Dieses Ergebnis ist so zu werten, dass die verschiedenen Verfahren nur in sehr geringem Maße ähnliche Fähigkeiten erfassen und somit voneinander relativ unterschiedliche und nicht austauschbare Reasoning-Tests darstellen.

IV. DISKUSSION UND AUSBLICK

Ziel der vorliegenden Untersuchung war es, einen Beitrag zur Validierung einer Testbatterie aus sechs Reasoning-Tests, von denen sich der Großteil noch in Entwicklung befindet, zu leisten.

Einerseits wurden die einzelnen Verfahren dafür dahingehend überprüft, ob sie mit dem dichotomen logistischen Modell von Rasch konform sind und damit jeweils eindimensional die Fähigkeit zum logisch-schlussfolgernden Denken erfassen. Bei den diesbezüglichen Analysen zeigten sich alle untersuchten Items des *Färbigen Matrizentests* sowie der Tests *N-TOP* und *Gleichungen* als dem Rasch-Modell entsprechend. Hier mussten also keine Items aus dem Itempool entfernt werden. Im Gegensatz dazu wurde ein Item aus *Formale Folgerichtigkeit* (entspricht ca. 5 % aller Items) und jeweils fünf Items aus *Sprachliche Analogien* (ca. 24 %) und *Figurale Analogien* (ca. 21 %) ausgeschieden, um zumindest a-posteriori Rasch-Modell-Geltung der Verfahren zu erreichen. Vor allem bei den beiden Analogie-Tests war der Inhalt der Aufgaben dafür verantwortlich, dass sich die entfernten Items als auffällig erwiesen.

Bei *Sprachliche Analogien* wurden in vielen Fällen unterschiedliche Itemschwierigkeitsparameterschätzungen für Schülerinnen und Schüler aus allgemein bildenden höheren Schulen und jene aus berufsbildenden höheren Schulen festgestellt, was bedeutet, dass der Besuch eines bestimmten Schultyps durch das Erlernen spezifischer Themenbereiche einen Vorteil bei der Lösung mancher Wissensaufgaben brachte. Diese Items sind aufgrund dessen also nicht geeignet, eine valide Aussage über die Fähigkeit einer Testperson zu machen, da neben der kristallisierten Reasoning-Fähigkeit sozusagen auch noch der besuchte Schultyp durch das jeweilige Item miterfasst wird. Ähnliches gilt für die Fälle, in denen sich unterschiedliche Itemparameterschätzungen für weibliche und männliche Testpersonen ergaben, was vor allem bei einigen Items aus *Figurale Analogien* beobachtet wurde. Im Falle des Tests *Formale Folgerichtigkeit* konnte kein inhaltlicher Grund für die Auffälligkeit des ausgeschiedenen Items gefunden werden; es musste aber dennoch aus dem Itempool entfernt werden. Zu den Verfahren *FMT* und *Gleichungen* ist noch anzumerken, dass die hier untersuchten Items einen geringen Range in den Schwierigkeiten aufweisen und daher insgesamt weniger gut zwischen Personen mit unterschied-

lichem Fähigkeitsniveau differenzieren können. Diese Tatsache mag auch dazu beitragen, dass bei den Items keine Abweichungen vom Rasch-Modell festgestellt werden konnten.

Neben den Rasch-Modell-Analysen wurde weiters eine exploratorische Faktorenanalyse, genauer gesagt eine Hauptkomponentenanalyse, zur Überprüfung der Dimensionalität der Reasoning-Test-Batterie durchgeführt. Hierfür wurden zunächst die Personenfähigkeitsparameter auf Basis der Rasch-Modell-konformen Items berechnet. Das Eigenwertdiagramm der extrahierten Faktoren und der Scree-Plot weisen in erster Linie auf eine Generalfaktorlösung hin, was bedeutet, dass die Leistungen in allen sechs Verfahren zumindest teilweise auf einer gemeinsamen latenten Fähigkeit beruhen. Entsprechend der Ladungsmatrix könnte dieser gemeinsame Faktor als „abstrakte Reasoning-Fähigkeit“ oder kurz „abstraktes Denken“ bezeichnet werden, da er am stärksten durch Verfahren, die numerisches oder abstrakt-bildhaftes Material verwenden, beschrieben wird, und zwar in erster Linie durch *Gleichungen*, *Formale Folgerichtigkeit* und *N-TOP*. Auf Basis des Eigenwert-Kriteriums wurde neben der Generalfaktorlösung auch noch eine Zweifaktoren-Lösung betrachtet. Diese entspricht inhaltlich sehr gut der theoretischen Konzeption der einzelnen Tests, da hier ein Faktor gefunden wurde, der Reasoning im fluiden Intelligenzbereich abdeckt, und ein zweiter Faktor, der vor allem Allgemeinwissen oder kristallisiertes schlussfolgerndes Denken repräsentiert. Die Zuteilung der einzelnen Verfahren zu den zwei Faktoren anhand der Ladungsmatrix entspricht weitgehend der theoretischen Zuordnung der Tests zu den Intelligenzbereichen nach Cattell. Nur beim Verfahren *Gleichungen* lässt sich laut Faktorenanalyse keine eindeutige Zugehörigkeit feststellen.

Die Hypothese bezüglich der korrelativen Zusammenhänge, in der postuliert wurde, dass Verfahren, die demselben Intelligenzbereich sensu Cattell zugeordnet werden, stärker miteinander korrelieren sollten als solche, die unterschiedlichen Bereichen angehören, konnte teilweise bestätigt werden. Bei *Sprachliche Analogien* und *Figurale Analogien* zeigen sich die höchsten Korrelationen relativ eindeutig mit jenen Verfahren, die ebenfalls dem kristallisierten Intelligenzbereich zuzurechnen sind. Somit kann in diesem Fall die Hypothese beibehalten werden, wobei die Korrelationskoeffizienten durchwegs als relativ niedrig bezeichnet werden müssen. Ähnliches gilt für die Verfahren *Formale Folgerichtigkeit* und *Färbiger Matrizentest*. Erstgenannter Test korreliert am stärksten mit dem *FMT* sowie mit *N-TOP*, die allesamt fluides Reasoning erfassen sollen, wobei beim *Färbigen Matrizentest* zudem auch noch die gleiche Materialart (figural) zum Ein-

satz kommt. Eine ähnlich hohe Korrelation zeigt sich jedoch auch mit dem Verfahren *Gleichungen*, das weder demselben Intelligenzbereich noch derselben Materialform zuzuordnen ist. Der *FMT* korreliert hypothesenkonform am stärksten mit *Formale Folgerichtigkeit*, während der Zusammenhang mit *N-TOP* entgegen den Erwartungen weit niedriger ausfällt. Für die Tests *Gleichungen* und *N-TOP* sind die größten Zusammenhänge mit dem jeweils anderen Test festzustellen; diese beiden weisen innerhalb der gesamten Interkorrelationsmatrix die höchste Korrelation auf. Hier dürften materialabhängige Fähigkeiten, d. h. numerische Fähigkeiten, für den Zusammenhang verantwortlich sein. *N-TOP* zeigt die zweithöchste Korrelation mit *Formale Folgerichtigkeit*, jedoch, wie soeben erwähnt, eine viel niedrigere mit dem *FMT*, was durch die gemeinsame Zuordnung zum Bereich der fluiden Intelligenz im ersten Fall als erwartungskonform und im zweiten als erwartungswidrig zu bezeichnen ist. Bei *Gleichungen* wiederum ist der zweitgrößte Zusammenhang mit *Figurale Analogien* zu beobachten, was sich ebenfalls durch die Erfassung des kristallisierten Schlussfolgerns in beiden Verfahren erklären lässt. Zu *Sprachliche Analogien* weist der Test *Gleichungen* jedoch eine deutlich niedrigere Korrelation auf; hier wäre theoretisch ein ähnliches Ausmaß wie bei *Figurale Analogien* zu erwarten gewesen. Dies könnte dahingehend interpretiert werden, dass kristallisierte numerische und figurale Reasoning-Aufgaben ähnlichere Fähigkeiten erfordern als numerische und sprachliche Aufgabenstellungen.

Über die gesamte Testbatterie hinweg wäre auf Basis des theoretischen Hintergrunds anzunehmen gewesen, dass der größte Zusammenhang zwischen *Formale Folgerichtigkeit* und *FMT* besteht, da beide Verfahren Reasoning im fluiden Intelligenzbereich mithilfe von figuralem Material erfassen sollen. Diese Korrelation ist allerdings nur am zweithöchsten innerhalb der gesamten Korrelationsmatrix ausgefallen.

Zusammenfassend lässt sich bezüglich der beobachteten Korrelationen also feststellen, dass in den meisten Fällen die Leistungen einer Testperson in den untersuchten Verfahren stärker zusammenhängen, wenn die verglichenen Tests demselben Intelligenzbereich (kristallisiert/fluid) angehören; in anderen Fällen führt dieselbe verwendete Materialart (insbesondere „numerisch“) zu einer größeren Ähnlichkeit zwischen den Tests. Die Ergebnisse lassen sich daher weitgehend mit den anfänglich formulierten Erwartungen in Einklang bringen. Insgesamt lässt sich aber aufgrund der durchgehend nicht sehr hohen Korrelationen schlussfolgern, dass die einzelnen Reasoning-Tests im Wesentlichen doch relativ spezifische Fähigkeiten messen und somit möglicherweise durch ihren gemeinsa-

men Einsatz die Fähigkeit zum logisch-schlussfolgernden Denken in ihren vielen Facetten gut abdecken können.

Kritisch anmerken lässt sich in Bezug auf die untersuchte Testbatterie, dass in Hinblick auf die Konstruktvalidierung im Sinne einer MTMM-Matrix nicht alle Kombinationen aus zu erfassendem Fähigkeitsbereich (fluides vs. kristallisiertes Reasoning) und Materialart (lexikalisch, numerisch, figural) realisiert wurden und kein Verfahren vorgegeben wurde, das Reasoning im fluiden Intelligenzbereich mithilfe von lexikalischem Material erfasst. Des Weiteren wäre es vermutlich sinnvoller gewesen, auch beim *Färbigen Matrizentest* – wie bei den anderen Verfahren – Quasi-Parallelförmigkeiten einzusetzen, da dieser den Testpersonen insgesamt recht schwer gefallen ist bzw. die Zeit für viele nicht für die Bearbeitung aller Aufgaben ausgereicht hat und somit die am Ende vorkommenden Items von vielen Personen nicht bearbeitet werden konnten, wodurch sich für diese nur wenige Daten ergaben. Bei zukünftigen Untersuchungen des *FMT* sollte daher darauf geachtet werden, vor allem für die betreffenden Items genügend Daten zu sammeln und diese erneut einer Rasch-Modell-Analyse zu unterziehen. Das hohe Schwierigkeitsniveau des *FMT* bei dieser Stichprobe führte auch dazu, dass der Range der Rohscores bzw. damit auch der Personenparameter nicht sehr groß ausfiel, womit in diesem Fall nicht gut zwischen Personen mit höherem Fähigkeitsniveau diskriminiert werden kann. Diese Tatsache könnte möglicherweise auch zu der unerwartet niedrigen Korrelation des *FMT* mit *Formale Folgerichtigkeit* beigetragen haben.

Wünschenswert wäre es zudem, die untersuchten Reasoning-Tests im Sinne einer „Art Kreuzvalidierung“ (Kubinger, 2005; Kubinger et al., 2011, S. 509, 561) an einer anderen Stichprobe nochmals zu überprüfen. Hierfür sollten für die Items, die sich als Rasch-Modell-konform erwiesen haben, an einem neuen Datensatz Analysen durchgeführt werden, beispielsweise mit Personen unterschiedlicher Altersgruppen, anderer Schultypen, Bildungsschichten, o. Ä., wobei auch die Verwendung anderer als der hier verwendeten Teilungskriterien (z. B. Alter) empfohlen werden kann. Dies gilt vor allem für das Verfahren *Sprachliche Analogien*, da für dieses noch keine anderen Daten als die hier erhobenen vorliegen. Außerdem sollten auch die im Rahmen dieser Arbeit nicht überprüften Items der Reasoning-Tests Rasch-Modell-Analysen unterzogen werden.

Zusammenfassend kann festgehalten werden, dass insgesamt mit der vorliegenden Untersuchung aufschlussreiche Erkenntnisse bezüglich der Validität und Dimensionalität

der Reasoning-Test-Batterie gewonnen werden konnten. Die Ergebnisse der Rasch-Modell-Analysen entsprachen weitgehend den Erwartungen und können im Großen und Ganzen als zufriedenstellend bezeichnet werden. Mit den hier untersuchten Verfahren stehen – zumindest nach Ausschluss einiger nicht-modellkonformer Items – jeweils ein-dimensional messende Reasoning-Tests zur Verfügung, denen insgesamt als Testbatterie auch faktorielle Validität zugesprochen werden kann. Trotz der gemeinsamen Erfassung des Konstrukts Reasoning erwiesen sich die Verfahren dennoch als relativ verschieden, wodurch sie in der Kombination die Fähigkeit zum schlussfolgernden Denken in ihren verschiedenen Facetten gut abzudecken vermögen und je nach psychologisch-diagnostischer Fragestellung unterschiedlich relevant sein könnten. Unter der Voraussetzung weiterer Validitätsprüfungen und ergänzender Rasch-Modell-Kontrollen eignen sich die Verfahren also durchaus für den Einsatz bei Fragestellungen, in denen die Erfassung des logisch-schlussfolgernden Denkens von Bedeutung ist.

V. ZUSAMMENFASSUNG

In der vorliegenden Arbeit wurden die Dimensionalität und Validität einer Testbatterie bestehend aus sechs in Entwicklung befindlichen psychologisch-diagnostischen Verfahren zur Erfassung der Fähigkeit zum logisch-schlussfolgernden Denken überprüft. Drei dieser Reasoning-Tests, nämlich der Untertest *Formale Folgerichtigkeit* aus der Intelligenz-Testbatterie AID-Gruppe (Kubinger & Hagenmüller, in Vorb.), der *Färbige Matrizentest* (Undeutsch, 2010a) sowie das Verfahren *N-TOP* (Kubinger & Heuberger, in Vorb.), messen laut theoretischer Konzeption schlussfolgerndes Denken im Bereich der fluiden Intelligenz sensu Cattell (z. B. 1943, 1963) und drei, d. s. die Verfahren *Sprachliche Analogien* (Kubinger & Ünal, in Vorb.), *Figurale Analogien* (Ünal & Kubinger, in Vorb.) und *Gleichungen* (Kubinger & Gamsjäger, in Vorb.), dienen der Erfassung von Reasoning im kristallisierten Intelligenzbereich.

Zur Überprüfung der Items auf Geltung des dichotomen logistischen Modells von Rasch sowie zur Analyse der Dimensionalität und Konstruktvalidität der Verfahren wurden Daten an insgesamt 265 Schülerinnen und Schülern aus Maturaklassen allgemeinbildender (AHS) und berufsbildender höherer Schulen (BHS) in Niederösterreich erhoben. Diese bearbeiteten die sechs Reasoning-Tests jeweils schriftlich in Form einer Gruppentestung, wobei die Bearbeitungszeit pro Test beschränkt war.

Die Analysen zur Prüfung der Verfahren auf Konformität mit dem Rasch-Modell erfolgten einerseits mit Hilfe des Likelihood-Ratio-Tests nach Andersen (1973), andererseits wurden grafische Modellkontrollen und der z -Test nach Fischer und Scheiblechner (1970, zitiert nach Kubinger et al., 2011, S. 556 f.) zur Überprüfung der Verfahren auf Itemebene durchgeführt. Als Teilungskriterien wurden die Anzahl gelöster Aufgaben (Score), das Geschlecht sowie der besuchte Schultyp (AHS vs. BHS) herangezogen. Für alle untersuchten Items des *Färbigen Matrizentests* wie auch der Verfahren *N-TOP* und *Gleichungen* ergab sich eine a-priori Geltung des Rasch-Modells. Eine Aufgabe des Tests *Formale Folgerichtigkeit* (entspricht ca. 5 % aller Items), etwa 24 % der Items von *Sprachliche Analogien* und ca. 21 % aus *Figurale Analogien* mussten sukzessive aus dem Itempool entfernt werden, um a-posteriori Modellgeltung zu erreichen. Nach Ausschluss der nicht-modellkonformen Aufgaben kann angenommen werden, dass die überprüften

Items aller sechs Tests jeweils eindimensional die Fähigkeit zum logisch-schlussfolgernden Denken messen.

Zur Bestimmung der Konstruktvalidität bzw. der faktoriellen Validität wurde zunächst eine Hauptkomponentenanalyse über alle Verfahren durchgeführt. Diese zeigte, dass die Leistungen der Testpersonen in den sechs verschiedenen Tests entweder auf einen Generalfaktor, der als „abstraktes Denken“ interpretiert wird, oder auf zwei zugrunde liegende Faktoren, nämlich „logisch-schlussfolgerndes Denken im fluiden Intelligenzbereich“ einerseits und „kristallisiertes schlussfolgerndes Denken (Allgemeinwissen)“ andererseits, zurückgeführt werden können. Beide Lösungen stehen im Einklang mit dem theoretischen Hintergrund der Reasoning-Tests.

Des Weiteren wurde die konvergente bzw. diskriminante Validität anhand der Produkt-Moment-Korrelation nach Pearson bestimmt. Die Interkorrelationen zwischen den Personenparametern der sechs Verfahren bewegten sich zwischen $r = 0,14$ und $r = 0,38$, wobei der größte Zusammenhang zwischen *N-TOP* und *Gleichungen* zu beobachten war. Die durchwegs relativ niedrigen positiven Korrelationen sprechen dafür, dass die Verfahren jeweils verschiedene Reasoning-Fähigkeiten erfassen.

In zukünftigen Untersuchungen sollten die im Rahmen dieser Arbeit überprüften sowie auch die hier nicht untersuchten Items der Reasoning-Tests anhand einer neuen Stichprobe weiteren Rasch-Modell-Analysen unterzogen sowie weitere Validierungsstudien durchgeführt werden.

VI. LITERATURVERZEICHNIS

- 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.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40(3), 153-193.
- Cattell, R. B. (1961). 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.
- Conrad, W. (1983). Intelligenzdiagnostik. In K.-J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik* (Enzyklopädie der Psychologie, Themenbereich B, Serie II, Band 2, S. 104-201). Göttingen: Hogrefe.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Huber.
- Formann, A. K., Waldherr, K., & Piswanger, K. (2011). *Wiener Matrizen-Test 2. Manual*. Göttingen: Beltz Test GmbH.
- Gamsjäger, C. (2012). *Entwicklung des Tests „Gleichungen“ zur Erfassung von Reasoning*. Unveröffentlichte Diplomarbeit, Universität Wien, Wien.
- Guilford, J. P., & Hoepfner, R. (1976). *Analyse der Intelligenz*. Weinheim: Beltz.
- Guthke, J. (2003). Intelligenztest. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 209-216). Weinheim: Beltz.

- Hagenmüller, B. (2011). *Entwicklung des Untertests „Formale Folgerichtigkeit“ zur Erfassung von Reasoning in der Intelligenz-Testbatterie AID 3*. Unveröffentlichte Diplomarbeit, Universität Wien, Wien.
- 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.
- Horn, J. L., & Cattell, R. B. (1982). Whimsy and misunderstandings of Gf-Gc Theory: A comment on Guilford. *Psychological Bulletin*, 91(3), 623-633.
- Jäger, A. O. (1973). *Dimensionen der Intelligenz* (3. Auflage). Göttingen: Hogrefe.
- Klauer, K. J. (Hrsg.) (2001a). *Handbuch Kognitives Training* (2., überarbeitete und erweiterte Auflage). Göttingen: Hogrefe.
- Klauer, K. J. (2001b). Training des induktiven Denkens. In K. J. Klauer (Hrsg.), *Handbuch Kognitives Training* (2., überarbeitete und erweiterte Auflage) (S. 165-209). Göttingen: Hogrefe.
- Kubinger, K. D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie: Ein Abriß samt neuesten Beiträgen* (2., verbesserte Auflage) (S. 19-83). 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. (2009a). *Adaptives Intelligenz Diagnostikum - Version 2.2 (AID 2) samt AID 2-Türkisch*. Göttingen: Beltz.

- Kubinger, K. D. (2009b). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2., überarbeitete und erweiterte Auflage). Göttingen: Hogrefe.
- Kubinger, K. D., & Gamsjäger, C. (in Vorb.). *Gleichungen*.
- Kubinger, K. D., & Hagenmüller, B. (in Vorb.). *AID-Gruppe*.
- Kubinger, K. D., & Heuberger, N. (in Vorb.). *Numerische Topologien (N-TOP)*.
- Kubinger, K. D., & Holocher-Ertl, S. (2014). *Adaptives Intelligenz Diagnostikum 3 – Version 3.1 (AID 3)*. Göttingen: Hogrefe.
- Kubinger, K. D., Rasch, D., & Yanagida, T. (2011). *Statistik in der Psychologie. Vom Einführungskurs bis zur Dissertation*. Göttingen: Hogrefe.
- Kubinger, K. D., & Ünal, S. (in Vorb.). *Sprachliche Analogien*.
- Kubinger, K. D., & Wurst, E. (1985). *Adaptives Intelligenz Diagnostikum (AID)*. Weinheim: Beltz.
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Beltz.
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). Manual* (2., erweiterte und überarbeitete Auflage). Göttingen: Hogrefe.
- Mair, P., Hatzinger, R., & Maier, M. J. (2014). eRm: Extended Rasch Modeling (R package version 0.15-4). Retrieved from <http://erm.r-forge.r-project.org/>
- Meiser, T., & Klauer, K. C. (2001). Training des deduktiven Denkens. In K. J. Klauer (Hrsg.), *Handbuch Kognitives Training* (2., überarbeitete und erweiterte Auflage) (S. 211-234). Göttingen: Hogrefe.
- Moosbrugger, H. (2012). Item-Response-Theorie (IRT). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage) (S. 227-274). Heidelberg: Springer.

- Moosbrugger, H., & Hartig, J. (2003). Faktorenanalyse. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 137-145). Weinheim: Beltz.
- Moosbrugger, H., & Schermelleh-Engel, K. (2012). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage) (S. 325-343). Heidelberg: Springer.
- R Core Team (2014). R: A language and environment for statistical computing (Version 3.1.2) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Revelle, W. (2015). psych: Procedures for Personality and Psychological Research (R package version 1.5.1). Evanston, Illinois, USA: Northwestern University. Retrieved from <http://CRAN.R-project.org/package=psych>
- Schermelleh-Engel, K., & Schweizer, K. (2003). Diskriminante Validität. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 103-110). Weinheim: Beltz.
- Schermelleh-Engel, K., & Schweizer, K. (2012). Multitrait-Multimethod-Analysen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2., aktualisierte und überarbeitete Auflage) (S. 345-362). Heidelberg: Springer.
- 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: Their nature and measurement*. London: Macmillan & Co.
- Srp, G. (1994). *Syllogismen. Manual*. Frankfurt/Main: Swets Test Services.
- Stemmler, G., Hagemann, D., Amelang, M., & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung* (7., vollständig überarbeitete Auflage). Stuttgart: Kohlhammer.

- 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.
- Ünal, S., & Kubinger, K. D. (in Vorb.). *Figurale Analogien*.
- Undeutsch, N. C. (2010a). *Färbiger Matrizentest (FMT)*. Großhöflein: P&T Prieler Tometich Verlag.
- Undeutsch, N. C. (2010b). *Zwei neue Subtests für das „Self-Assessment“ Psychologie durch regelgeleitete Konstruktion*. Unveröffentlichte Diplomarbeit, Universität Wien, Wien.
- Waldmann, M., & Weinert, F. E. (1990). *Intelligenz und Denken: Perspektiven der Hochbegabungsforschung*. Göttingen: Hogrefe.

VII. ANHANG

A. Testdesign

In der ersten Zeile der Tabelle sind jeweils die Nummern der Items angeführt, wie sie in den Verfahren verwendet sind, und in den darunterliegenden die entsprechende Nummerierung in den im Rahmen dieser Diplomarbeit vorgegebenen Testformen A und B.

Testdesign von *Formale Folgerichtigkeit* (Kubinger & Hagenmüller, in Vorb.)

Itemnummer											
	6	23	29	35	52	59	69	70	71	76	77
Testheft A	4	2	3	11	10	6	1	20	8	21	9
Testheft B	6	1	2	10	12	5	3	19	7	22	8
Itemnummer											
	78	83	85	87	89	95	98	99	100	101	105
Testheft A	5	13	7	12	14	15	16	19	18	22	17
Testheft B	4	15	9	11	13	14	18	20	17	21	16

Testdesign von *Sprachliche Analogien* (Kubinger & Ünal, in Vorb.)

Itemnummer											
	1	7	8	17	23	25	26	27	28	29	30
Testheft A	12	18	21	11	1	8	3	4	5	6	7
Testheft B	3	9	12	2	13	20	15	16	17	18	19
Itemnummer											
	31	32	33	34	35	36	37	39	40	41	
Testheft A	9	10	19	13	2	14	20	15	16	17	
Testheft B	21	1	10	4	14	5	11	6	7	8	

Testdesign des *Färbigen Matrizentests* (Undeutsch, 2010a) (TH = Testheft)

Itemnummer									
	1	2	3	4	5	8	10	11	12
TH A und B	2	1	3	4	5	6	7	8	9
Itemnummer									
	14	17	18	19	20	21	22	23	25
TH A und B	10	11	12	13	14	15	16	17	18

Testdesign von *N-TOP* (Kubinger & Heuberger, in Vorb.) (TH = Testheft)

Itemnummer															
	1b	7a	9b	11b	12a	14b	16b	18b	19a	21a	22b	24a	25a	26b	27b
TH A	1	10	15	2	8	7	11	13	4	6	9	3	5	14	12
TH B	8	2	7	9	15	14	3	5	11	13	1	10	12	6	4

Testdesign von *Figurale Analogien* (Ünal & Kubinger, in Vorb.)

Itemnummer												
	2	3	4	5	6	7	8	9	10	11	12	13
Testheft A	2	4	5	7	8	6	3	13	10	11	12	19
Testheft B	13	10	16	3	2	24	11	20	4	5	14	22
Itemnummer												
	14	16	17	18	20	21	22	23	24	25	26	27
Testheft A	18	9	14	15	17	16	20	21	22	23	1	24
Testheft B	23	1	18	9	6	15	17	12	21	8	7	19

Testdesign von *Gleichungen* (Kubinger & Gamsjäger, in Vorb.)

Itemnummer												
	3	8	10	11	14	18	20	24	30	36	39	40
Testheft A	1	2	3	4	5	6	7	8	9	10	11	12
Testheft B	2	5	4	6	1	3	8	7	10	12	9	11

B. Genehmigung der Untersuchung durch den Landesschulrat für Niederösterreich

Landesschulrat für Niederösterreich



Rennbahnstraße 29
3109 St. Pölten

Frau
Bernadette SCHNAIT
Bernadette.Schnait@gmx.at

Sachbearbeiterin:
Maria Meister
t: +43 2742 280 4311
f: +43 2742 280 1111
e: maria.meister@lsr-noe.gv.at

Beilage(n): ---
Bezug: ----

Präs.-420/1448-2014

Datum: 10.09.2014

Betrifft:

Genehmigung der Durchführung einer empirischen Untersuchung in Form einer
Fragebogenerhebung

Der Landesschulrat für Niederösterreich genehmigt die Durchführung der vorgelegten empirischen Studie zum Thema „*Analyse und Vergleich psychologischer Verfahren zur Erfassung der Fähigkeit zum schlussfolgernden Denken*“ durch Frau Bernadette Schnait an folgenden Schulen:

Berufsbildende höhere Schulen	Allgemein bildende höhere Schulen
HTBLuVA Mödling	BG Wr. Neustadt, Zehnergasse
HTL der Stadtgemeinde Ybbs/Donau	BORG Deutsch-Wagram
HAK/HAS der Stadtgemeinde Ybbs/Donau	BG/BRG Schwechat
BHAK/BHAS Horn	BORG Mistelbach
BHAK/BHAS Waidhofen/Ybbs	
HAK/HAS der Stadtgemeinde Tulln	
BHAK/BHAS Amstetten	
BHAK/BHAS Hollabrun	

Es ist bei dieser Untersuchung ein Elterninformationsblatt vorzulegen, mit welchem vor Durchführung der Untersuchung die Zustimmung der Erziehungsberechtigten einzuholen ist. SchülerInnen, welche bereits das 18. Lebensjahr vollendet haben, können selbst über die Teilnahme entscheiden.

Die an dieser Untersuchung teilnehmenden SchülerInnen sind vor Beginn der Erhebung ausdrücklich auf die Freiwilligkeit der Teilnahme hinzuweisen, außerdem ist deren Anonymität jedenfalls zu wahren.

Auf die Einhaltung der Datenschutzbestimmungen wird hingewiesen.

Parteienverkehr
Dienstag 8-12 Uhr

<http://www.lsr-noe.gv.at>
office@lsr-noe.gv.at
DVR:0064394

Amtsstunden
Mo.-Fr. 8-16 Uhr

Vor Beginn der Erhebung ist die Zustimmung der jeweiligen Direktion einzuholen. Bei Privatschulen ist neben der Zustimmung der Schulleitung auch das Einverständnis des Privatschulenerhalters einzuholen.

Es ist darauf zu achten, dass die Durchführung der Fragebogenerhebung nur die unbedingt erforderliche Zeit, längstens jedoch eine Unterrichtsstunde in Anspruch nimmt.

Einer Übermittlung der Ergebnisse wird entgegengesehen.

Für den Amtsführenden Präsidenten:

Mag. Loibl

Hofrat

Elektronisch gefertigt

Signaturwert	cd400830fc6547a2a41080cc1aa27129	
	Unterzeichner	Landesschulrat fuer Niederoesterreich
	Datum/Zeit-UTC	12.09.2014 16:08:37
	Aussteller-Zertifikat	CN=a-sign-corporate-light-02, OU=a-sign-corporate-light-02, O=A-Trust Ges. f. Sicherheitssysteme im elektr. Datenverkehr GmbH, C=AT
	Serien-Nr.	710400110544
Prüfinformation	Informationen zur Prüfung der elektronischen Signatur finden Sie unter: https://www.a-trust.at/pdfverify Informationen zur Prüfung des Ausdrucks finden Sie unter: http://www.lsr-noe.gv.at/amtssignatur-bildmarke.html	
Hinweis	Dieses Dokument wurde amtssigniert. Auch ein Ausdruck dieses Dokuments hat gemäß § 20 E-Government-Gesetz die Beweiskraft einer öffentlichen Urkunde.	

Parteienverkehr
Dienstag 8-12 Uhr

<http://www.lsr-noe.gv.at>
office@lsr-noe.gv.at
DVR:0064394

Amtsstunden
Mo.-Fr. 8-16 Uhr

C. Beispiel eines Informationsblattes für die Schüler/innen



universität
wien

Bernadette Schnait, B.A.

Diplomandin an der Fakultät für Psychologie

Institut für Angewandte Psychologie:

Gesundheit, Entwicklung und Förderung

Universität Wien

E-Mail: bernadette.schnait@gmx.at

Tel.: +43 (0) 660 / 482 99 42

Liebe Schülerin, lieber Schüler!

In den nächsten Wochen werden mit Einverständnis deines Direktors Testungen in den Maturaklassen deiner Schule stattfinden, zu denen ich dich herzlich einladen möchte.

Worum geht's?

Im Rahmen meiner Diplomarbeit an der Fakultät für Psychologie der Universität Wien beschäftige ich mich mit der **Analyse und dem Vergleich – großteils relativ neu entwickelter – psychologischer Tests, mit Hilfe derer die Fähigkeit zum logisch-schlussfolgernden Denken erfasst werden kann**. Tests dieser Art werden beispielsweise in der Berufs- und Bildungsberatung oder bei Eignungsbeurteilungen im Rahmen von Auswahl-/Aufnahmeverfahren eingesetzt. Es ist also nicht unwahrscheinlich, dass du mit derartigen Verfahren zukünftig einmal konfrontiert wirst. Dabei ist es wichtig, dass die verwendeten Tests gut und ausreichend erprobt sind, um zuverlässige Aussagen über die Stärken und Schwächen der jeweiligen Person machen zu können.

Zur Überprüfung dieser psychologischen Verfahren sind die Daten von ca. 200 SchülerInnen aus Abschlussklassen höherer Schulen notwendig. Daher möchte ich dich hiermit bitten, mich bei meiner Arbeit zu unterstützen und an der Testung teilzunehmen. Du wirst dabei verschiedene Tests vorgegeben bekommen, die die Fähigkeit zum logischen Denken erheben. Die Untersuchung findet **während der Schulzeit** statt und dauert **ca. 60-75 Minuten**. Die Testung wird **in der Klasse** durchgeführt, wobei jede/r SchülerIn die Tests **schriftlich** bearbeiten wird. Es kommt bei der Untersuchung nicht auf das Abschneiden der einzelnen SchülerInnen an, sondern die Ergebnisse werden zusammengefasst und in ihrer Gesamtheit ausgewertet. Die Teilnahme ist natürlich **freiwillig** und die Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke verwendet und **anonymisiert**.

Was hast du davon?

- Du hast die Möglichkeit, einen **schriftlichen Ergebnisbericht** über deine Stärken und Schwächen zu erhalten.
- Du bekommst einen Einblick in (neue) psychologische Testverfahren.
- Du unterstützt damit die Wissenschaft und eine engagierte Studentin.

Ich wäre dir sehr dankbar, wenn du an der Testung teilnehmen und dich mit deinem Namen und deiner Unterschrift in die Liste eintragen würdest.

Für Fragen kannst du mich gerne telefonisch oder per E-Mail kontaktieren oder dich am Testungstag an die TestleiterInnen und mich wenden.

Mit lieben Grüßen und herzlichem Dank im Voraus!

Bernadette Schnait

D. Beispiel eines Informationsbriefes für die Eltern der Schüler/innen



universität
wien

Bernadette Schnait, B.A.

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

E-Mail: bernadette.schnait@gmx.at

Tel.: +43 (0) 660 / 482 99 42

Liebe Eltern und Erziehungsberechtigte!

Viele Kinder und Jugendliche werden im Laufe ihrer schulischen Karriere aus verschiedenen Gründen mit psychologischen Tests untersucht, zum Beispiel zur Abklärung einer Hochbegabung, aufgrund von Lern- und Leistungsproblemen oder auch im Rahmen einer Berufs- oder Bildungsberatung. Meist kommen in diesem Zusammenhang auch Tests zur Erfassung der Fähigkeit zum schlussfolgernden Denken zum Einsatz, um die intellektuellen Fähigkeiten des Kindes oder Jugendlichen abschätzen zu können. Um adäquate Aussagen über die Stärken und Schwächen des Kindes/des Jugendlichen machen zu können, sind gute und ausreichend erprobte psychologische Verfahren nötig, unter anderem solche, die logisches Denken überprüfen.

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 logisch-schlussfolgernden Denken. Neben bereits etablierten psychologischen Tests werden in meine 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.

Zu diesem Zweck werden Schülerinnen und Schülern aus Abschlussklassen höherer Schulen verschiedene psychologische Tests zur Bearbeitung vorgegeben, da gerade zu diesem Zeitpunkt im Laufe ihrer Ausbildung oftmals derartige Verfahren eingesetzt werden, um die Jugendlichen in Bezug auf eine mögliche Studien- oder Berufswahl angemessen beraten zu können.

Beim Aufgabenmaterial, das im Rahmen meiner Studie zur Anwendung kommt, handelt es sich um sechs verschiedene psychologische Tests, die die Fähigkeit zum logischen Schlussfolgern auf unterschiedliche Weise erheben. Die Untersuchungen finden während der Schulzeit statt und dauern ca. 60-75 Minuten. Die Testungen werden im Klassenverband durchgeführt, wobei die SchülerInnen die Tests jede/r für sich schriftlich bearbeiten werden.

Die Direktion der Schule Ihres Sohnes/Ihrer Tochter hat sich zur Teilnahme an meiner Studie einverstanden erklärt. Daher führe ich in den nächsten Wochen in der Klasse Ihres Kindes eine Testung mit den erwähnten psychologischen Verfahren durch.

Ich würde es sehr begrüßen, auch Ihren Sohn/Ihre Tochter in meine Studie einbeziehen zu können. Ich wende mich daher mit der Bitte an Sie, Ihren Sohn/Ihre Tochter an der Untersuchung teilnehmen zu lassen, vorausgesetzt natürlich, Ihr Kind ist damit einverstanden. Es kommt bei der Untersuchung nicht auf das Abschneiden der einzelnen SchülerInnen an, sondern die Daten werden zusammengefasst und in ihrer Gesamtheit ausgewertet. Jedoch besteht die Möglichkeit – falls erwünscht –, Ihrer Tochter/Ihrem Sohn einen kurzen schriftlichen Ergebnisbericht über seine/ihre Stärken und Schwächen zukommen zu lassen.

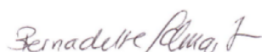
Für Rückfragen stehe ich gerne telefonisch oder per E-Mail zur Verfügung:

Bernadette Schnait
E-Mail: bernadette.schnait@gmx.at
Tel.: +43 (0) 660 482 99 42

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

Ich bitte Sie, mit Ihrer Unterschrift auf dem unten angefügten Abschnitt Ihr Einverständnis zur Teilnahme Ihres Sohnes/Ihrer Tochter an meiner Untersuchung zu erteilen.

Mit freundlichen Grüßen und herzlichem Dank im Voraus!



Bernadette Schnait
Diplomandin an der Fakultät für Psychologie
Universität Wien

----- bitte hier abtrennen -----

Ich erkläre mich mit der Teilnahme meiner Tochter/meines Sohnes _____

_____, geboren am _____,

an der Testung für die Diplomarbeit von Bernadette Schnait einverstanden.

.....

Datum

.....

Unterschrift

E. Testleiter/innenanweisung für die Durchführung der Testungen

Testleiteranweisung

Testungen Bernadette Schnait

Informationen zum Ablauf der Testungen

1) Einleitungstext (siehe S. 3)

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

Die Testbatterie besteht aus 6 Untertests und einem Blatt für die demografischen Daten in folgender **Reihenfolge**:

0. Demografische Daten (weißes Papier)
1. Formale Folgerichtigkeit (weißes Papier)
2. Sprachliche Analogien (farbiges Papier)
3. Färbiger Matrizentest (weißes Papier)
4. N-TOP (farbiges Papier)
5. Figurale Analogien (weißes Papier)
6. Gleichungen (farbiges Papier)

Die Testhefte liegen in **Form A und B** vor. Ein/e SchülerIn bekommt alle Untertests in derselben Testform (zB immer A). SchülerInnen, die nebeneinander sitzen, sollen jeweils eine unterschiedliche Testform erhalten. Das heißt, die Testhefte werden abwechselnd in Form A und B an die SchülerInnen ausgeteilt.

Die Tests sind **vor** der Testung in der richtigen Reihenfolge sowie auch abwechselnd nach Testform A und B zu sortieren, um ein einfaches Austeilen zu gewährleisten.

SchülerIn	Reihenfolge der Testhefte
1	0 – 1A – 2A – 3A – 4A – 5A – 6A
2	0 – 1B – 2B – 3B – 4B – 5B – 6B
3	0 – 1A – 2A – 3A – 4A – 5A – 6A
4	0 – 1B – 2B – 3B – 4B – 5B – 6B
usw.	...

- 3)** Zeitpunkt des **Beginns der Testung** (ab Einleitungstext) und des **Endes der Testung** (sobald die Bearbeitungszeit des letzten Tests vorbei ist) sollen bitte am **Testleiterprotokoll** notiert werden, wie auch **Datum, Schule und Klasse sowie Namen der Testleiter**.
- 4)** Um den SchülerInnen Rückmeldung über ihre Testergebnisse geben zu können, sollen sie sich den Zettel vorne mit meiner E-Mail-Adresse und dem Probandencode mitnehmen.
- 5)** Die SchülerInnen 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**, zB durch Einkreisen.
- 6)** Wenn eine Instruktion nicht gleich verstanden wird (zB Beispielaufgabe) oder Fragen dazu sind, so wird diese bitte nochmals langsam 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 **Strich unter die zuletzt bearbeitete Aufgabe** zu ziehen.
- 8) Sollten **vor Ablauf** der Bearbeitungszeit alle Testpersonen bereits fertig sein, so kann schon früher zum nächsten Untertest über gegangen werden. Die Testpersonen sollen aber **bitte trotzdem einen Strich unter die zuletzt bearbeitete (in diesem Fall letzte) Aufgabe als Kontrolle ziehen**. Die tatsächlich benötigte Zeit wird dann auf dem Testleiterprotokoll notiert.
- 9) Sollten SchülerInnen nachfragen, ob sie bei Unsicherheit raten sollen oder die Aufgabe auslassen sollen, ist ihnen zu antworten, dass sie lieber raten sollen. Dies aber bitte nur auf Nachfrage sagen!
- 10) Bitte die Schüler **animieren, die Aufgaben der Reihe nach zu bearbeiten**, das ist für die anschließende Auswertung sehr wichtig. Wenn eine Aufgabe aber viel zu schwierig ist, dann kann man schon sagen, dass sie dann die nächste Aufgabe versuchen sollen, weil auch wieder einfachere kommen. Aber es soll nicht zum im-Testheft-Herumspringen animieren.
- 11) Bitte darauf achten, dass die SchülerInnen auch **umblättern**, wenn „**BITTE UMBLÄTTERN**“ am Ende der Seite steht.
- 12) Am Ende eines Tests müssen die SchülerInnen warten, bis sie zum Umblättern aufgefordert werden. Bitte darauf achten, dass sie **nicht** zum nächsten Untertest oder während der Testinstruktion zu den Testaufgaben **weiterblättern, ohne dass sie dazu aufgefordert wurden**. Anhand der **Farbe des Papiers** kann auch leicht überblickt werden, ob alle Testpersonen beim selben Test sind.
- 13) Es ist **darauf zu achten**, dass die SchülerInnen **nicht vom Nachbarn abschreiben** oder miteinander **reden**. Liegt der Verdacht nahe, dass eine SchülerIn abgeschrieben hat, so ist das am besten 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** werden, auch keine Hinweise zur Lösung gegeben werden. Bitte auch keine inhaltlichen Fragen beantworten (zB bei Figurale Analogien: was ein bestimmtes Bild darstellen soll) – bei derartigen Fragen bitte sagen, dass du diese leider nicht beantworten darfst.
- 16) Es ist darauf zu achten, dass alle SchülerInnen die Testhefte wieder **abgeben** und nicht mit nach Hause nehmen. Es muss auch **verhindert** werden, dass **Fotos** von den Tests gemacht werden.

Zur Information:

- Kursiv geschriebener Text ist der vom/von der TestleiterIn gesprochene Text, der an die SchülerInnen gerichtet ist.
- „Normal“ geschriebener Text ist Erklärung für die TestleiterInnen.

Testleiter-Instruktionen

1. Einleitung und Austeilen der Testhefte

Einen schönen Vormittag! Mein Name ist und das ist/sind meine KollegIn(nen) Ist es in Ordnung, wenn ich euch dutze?

*Wir möchten euch zur heutigen Testung **begrüßen** und uns erst einmal ganz herzlich dafür **bedanken**, dass ihr euch zur Teilnahme an der (meiner) Untersuchung von Bernadette Schnait bereit erklärt habt. Wie ihr wisst, werdet ihr heute ein paar **psychologische Tests** bearbeiten, in denen es um logisches Denken geht.*

***Ich möchte euch kurz erklären, wie das Ganze ablaufen wird.** Ihr werdet von uns gleich ein Testheft ausgeteilt bekommen, das aus insgesamt **6 Untertests** und einem **Blatt für eure persönlichen Daten** besteht. Auf diesem Blatt sollt ihr bitte zu Beginn einige Angaben zu eurer Person machen (zB Geschlecht, Alter usw.). Danach beginnt die eigentliche Testbearbeitung.*

*Wenn ihr das **Testheft** gleich bekommt, **lasst es bitte noch genauso liegen**, wie es vor euch hingelegt wird. Ich möchte euch erst alles zu Ende erklären und dann beginnen wir gemeinsam.*

[Der/Die andere TestleiterIn beginnt jetzt mit dem Austeilen.]

*Ich werde euch vor jedem einzelnen Untertest erklären, was bei den jeweiligen Aufgaben zu tun ist und wir werden das anhand von Beispielen gemeinsam üben. Ihr habt dann für jeden Untertest eine **bestimmte Zeit** zur Verfügung, die ich euch davor immer nennen werde. Mit der Bearbeitung der Aufgaben sollt ihr bitte **immer erst dann beginnen, wenn ich es euch sage**. Bearbeitet die Aufgaben bitte **der Reihe nach**, das ist für die Auswertung wichtig. Solltet ihr einmal eine Antwort ändern wollen, dann streicht bitte die erste Antwort durch und markiert die neue Antwort deutlich durch Ankreuzen und Einkreisen.*

***Nach Ablauf der Zeit** werde ich euch immer Bescheid geben und ihr **hört dann bitte zu arbeiten auf**. Ich werde euch dann immer bitten, einen **Strich unter die letzte bearbeitete Aufgabe** zu ziehen. Es spielt dabei keine Rolle, ob ihr die zuletzt gemachte Aufgabe lösen konntet oder nicht, bitte einfach einen Strich unter der Aufgabe ziehen, bei der ihr gerade wart.*

*Solltet ihr **früher** mit den Aufgaben eines Untertests **fertig** sein, **wartet bitte auf mein Zeichen** und blättert nicht vorher um. Wir kommen dann gemeinsam zum nächsten Untertest. Auf den Seiten, auf denen unten „**Bitte umblättern**“ steht, könnt ihr selbstständig umblättern und die nächsten Aufgaben bearbeiten.*

*Die einzelnen Tests werden euch wahrscheinlich unterschiedlich schwer fallen. Es kann auch sein, dass ihr nicht bei allen Untertests oder vielleicht sogar bei keinem alle Aufgaben lösen könnt. Das macht überhaupt nichts, jeder findet die Aufgaben unterschiedlich schwierig und hat andere Stärken und Schwächen. Wichtig ist, dass ihr die Aufgaben **gewissenhaft** bearbeitet und euch Mühe gebt. **Arbeitet immer zügig, aber trotzdem genau, und versucht so viele Aufgaben in der Zeit zu lösen wie möglich.***

*Bitte konzentriert euch auf euer eigenes Testheft und **schaut nicht vom Nachbarn ab**. Euer Sitznachbar hat auch nicht genau dasselbe Testheft wie ihr. Und bitte während der Testung **nicht miteinander reden** oder Sonstiges, also jeder bearbeitet die Aufgaben bitte eigenständig. Bitte gebt jetzt auch **alles vom Tisch weg** – auch **Handys bitte ausschalten** und weggeben – und bereitet einen **Stift** vor.*

Habt ihr bisher alles verstanden? Gibt es noch Fragen?

Erklärung Code-Zettel

Ganz vorne auf eurem Testheft findet ihr einen kleinen Zettel angeheftet, auf dem meine/eine E-Mail-Adresse und ein Code steht. Das ist derselbe Code, wie der, der auf eurem Testheft steht. Ihr könnt euch den Zettel jetzt runternehmen und wenn ihr gerne eine **Rückmeldung** haben wollt, wie gut ihr bei den einzelnen Tests abgeschnitten habt, könnt ihr mir (Frau Schnait) eine E-Mail schreiben, in der ihr bitte diesen Code angebt. Ich/Sie schicke/t euch dann, sobald ich (sie) mit der Auswertung fertig bin/ist, einen schriftlichen Ergebnisbericht zu. Das kann aber ein paar Monate dauern, bis ihr eine Rückmeldung bekommt, also bitte um etwas Geduld.

Die Testung ist natürlich **anonym**, das heißt, bei der Auswertung wird nicht mehr nachzuvollziehen sein, wer welchen Test ausgefüllt hat. Deswegen wird auch für die Rückmeldung ein Code verwendet.

Wenn es keine Fragen mehr gibt, beginnen wir jetzt mit der Testbearbeitung.

2. Demografische Daten (weißes Papier)

Bitte füllt jetzt zu Beginn die Daten zu eurer Person auf der ersten Seite aus.

3. UT Formale Folgerichtigkeit (weißes Papier)

Wenn die Schüler mit dem Ausfüllen der demografischen Daten fertig sind, wird wie folgt instruiert:

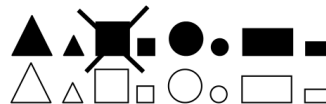
Wir beginnen jetzt mit der Bearbeitung der einzelnen Tests. Bitte blättert jetzt zum ersten Test „**Formale Folgerichtigkeit**“ um.

Der/die Testleiter(in) liest Folgendes vor, die Schüler(innen) können im Testheft mitlesen:

Auf den nächsten Seiten sind Folgen von geometrischen Figuren gegeben, die du immer um eine Figur verlängern sollst. Es geht also darum, welche Figur die Reihe logisch richtig fortsetzt.

Dir stehen dafür mehrere Antwortmöglichkeiten zur Auswahl. Kreuze die Lösung an. Es gibt immer nur eine richtige Antwortmöglichkeit.

Ihr seht dazu ein Beispiel:

Beispiel 1:

Die Lösung des **Beispiels 1** ist ein **großes schwarzes Quadrat**. Dieses ist bereits angekreuzt. Die Folge besteht aus lauter großen schwarzen Quadraten. Damit die nächste Figur die Folge richtig fortsetzt, muss diese wieder ein **großes schwarzes Quadrat** sein.

Versucht nun das **Beispiel 2** zu lösen und kreuzt entsprechend an.

Beispiel 2:



Wenn alle SchülerInnen angekreuzt haben, ist zu fragen:

Was ist die Lösung? [jemanden drannehmen, der aufgezeigt hat] *Ja genau, das **kleine weiße Quadrat**.*

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

*Auf den nächsten Seiten findest du dazu **22 Aufgaben**.*

*Für die Bearbeitung stehen dir **7 Minuten** zur Verfügung. Falls du vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig bist, warte bitte auf ein Zeichen des Testleiters.*

Gibt es noch Fragen?

Der/die Testleiter(in) gibt ein Startzeichen zum Umblättern, indem er/sie instruiert:

Los geht's. Bitte umblättern.

Für die Bearbeitung des Untertests stehen den Schülern **7 Minuten** zur Verfügung. Nach **3 Minuten 30 Sekunden** ist den Schülern mitzuteilen, dass sie nun noch die Hälfte der Zeit übrig haben.

Nach den gesamten **7 Minuten** ist mit dem Wort „*Stopp!*“ abzubrechen.

Die Schüler werden anschließend darum gebeten, im **Testheft unter die zuletzt bearbeitete Aufgabe einen Strich zu ziehen**.

Bitte macht nun einen Strich unter die Aufgabe, die ihr zuletzt bearbeitet habt.

Alternativ: Alle die fertig geworden sind, machen nun bitte einen Strich unter die letzte Aufgabe (Aufgabe 22). Die anderen machen den Strich unter die Aufgabe, bei der sie gerade sind.

4. UT Sprachliche Analogien (farbiges Papier)

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

*Blättert jetzt bitte auf die nächste Seite zum Test „**Sprachliche Analogien**“ auf dem hellorangen Papier.*

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

In diesem Test geht es darum, eine Beziehung zwischen zwei Objekten oder Begriffen zu erkennen und entsprechend anzuwenden: Zu einem zweiten Paar von Objekten oder Begriffen ist dazu jeweils nur eines/r vorgegeben, das/der entsprechend passende zweite gesucht. Drei Beispielaufgaben illustrieren das.

Beispielaufgabe a)

Fußballspielen : grün = Schifahren : ...

- ☐ rot
- ☐ schwarz
- ☐ gelb
- ☒ weiß
- ☐ blau

Die Lösung ist „weiß“. Fußballspielen findet auf einem grünen Untergrund statt, Schifahren auf einem weißen; die Beziehung zwischen Fußballspielen und grün betrifft also den Untergrund.

Beispielaufgabe b)

Feuer : Wasser = Säure : ...

- ☐ Salz
- ☐ Metall
- ☐ Kalk
- ☐ Hitze
- ☒ Base

Die Lösung ist „Base“. Feuer und Wasser neutralisieren sich gewissermaßen, Säure und Base verhalten sich ähnlich; die Beziehung zwischen Feuer und Wasser betrifft also „neutralisieren“.

Beispielaufgabe c)

49 : 7 = ... : 12

- ☐ 44
- ☐ 94
- ☒ 144
- ☐ 169
- ☐ 196

Die Lösung ist „144“. 7 und 49 beschreiben eine quadratische Funktion, 12 und 144 ebenfalls; die Beziehung zwischen 7 und 49 betrifft also die quadratische Funktion.

Auch bei allen folgenden Aufgaben gibt es immer eine bestimmte Gesetzmäßigkeit, die das gesuchte Wort (Objekt/Begriff) eindeutig festlegt.

*Für die Bearbeitung stehen dir **8 Minuten** zur Verfügung. Falls du vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig bist, warte bitte auf ein Zeichen des Testleiters.*

Gibt es noch Fragen?

Der/die Testleiter(in) gibt ein Startzeichen zum Umblättern, indem er/sie instruiert:

Los geht's. Bitte umblättern.

Für die Bearbeitung des Untertests stehen den Schülern **8 Minuten** zur Verfügung. Nach **4 Minuten** ist den Schülern mitzuteilen, dass sie nun noch die Hälfte der Zeit übrig haben.

Nach den gesamten **8 Minuten** ist mit dem Wort „*Stopp!*“ abzubrechen.

Die Schüler werden anschließend darum gebeten, im **Testheft unter die zuletzt bearbeitete Aufgabe einen Strich zu ziehen**.

Bitte macht nun einen Strich unter die Aufgabe, die ihr zuletzt bearbeitet habt.

Alternativ: Alle die fertig geworden sind, machen nun bitte einen Strich unter die letzte Aufgabe (Aufgabe 22). Die anderen machen den Strich unter die Aufgabe, bei der sie gerade sind.

5. **UT Färbiger Matrizentest** (weißes Papier)

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

*Blättert jetzt bitte auf die nächste Seite zum Test „**Färbiger Matrizentest**“.*

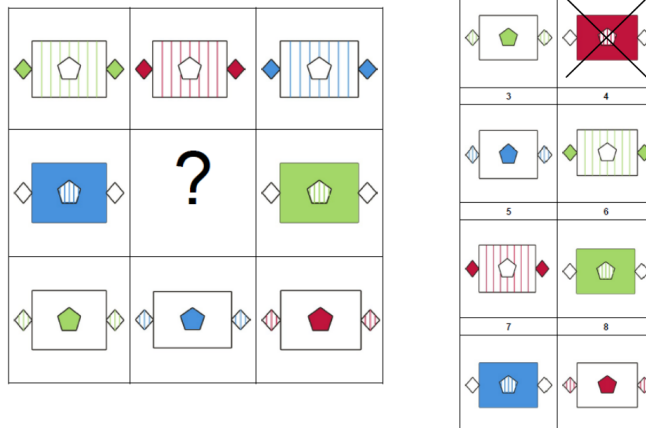
Der/Die TestleiterIn liest vor, die SchülerInnen können im Testheft mitlesen.

Die folgenden 18 Aufgaben bestehen jeweils aus einer Matrix mit einem Fragezeichen in der Mitte. Deine Aufgabe ist es, diejenige Figur aus den acht möglichen Figuren auf der rechten Seite auszuwählen, die logisch sinnvoll an die Stelle des Fragezeichens gesetzt werden kann.

Sieh dir bitte zuerst die zwei folgenden Beispiele an:

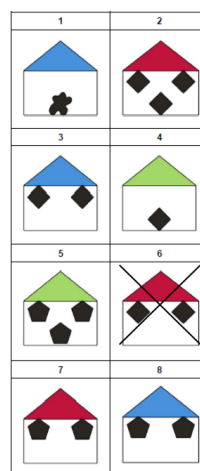
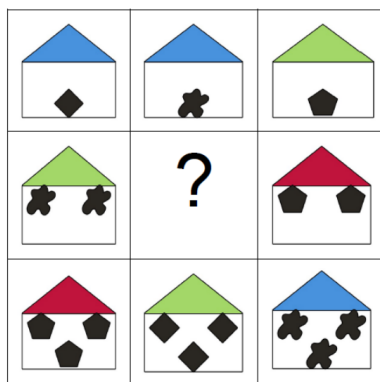
Der/Die Testleiterin zeigt mit einem Stift mit, wenn er/sie die folgenden Beispiele erklärt:

Beispiel 1:



Hier siehst du acht Figuren mit unterschiedlichen Farben und Mustern. In der ersten Reihe ist das Rechteck gestreift, in der zweiten Reihe ausgemalt und in der dritten Reihe weiß. Außerdem gibt es in jeder Reihe eine grüne, eine rote und eine blaue Figur. An die Stelle des Fragezeichens gehört also eine ausgemalte Figur, die rot ist.

Beispiel 2:



Hier gibt es in jeder Spalte ein blaues, ein grünes und ein rotes Dach. In jeder Zeile sind Fenster, die aussehen wie ein Klecks, ein Karo oder ein Fünfeck. Außerdem befinden sich in der ersten Reihe nur Häuser mit einem Fenster, in der zweiten Reihe mit zwei Fenstern und in der dritten Reihe mit drei Fenstern. Folglich fehlt in der Mitte ein Haus mit rotem Dach, das zwei Fenster mit der Form eines Karos hat.

Wie in den Beispielen markiere deine Antwort bitte, indem du die betreffende Figur ankreuzt. Es gibt immer nur eine richtige Lösung.

Für die Testbearbeitung stehen dir 10 Minuten zur Verfügung.

Falls du vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig bist, warte bitte auf ein Zeichen des Testleiters.

Gibt es von deiner Seite noch Fragen?

Der/die Testleiter(in) gibt ein Startzeichen zum Umblättern, indem er/sie instruiert:

Los geht's. Bitte umblättern.

Für die Bearbeitung des Untertests stehen den Schülern **10 Minuten** zur Verfügung. Nach **5 Minuten** ist den Schülern mitzuteilen, dass sie nun noch die Hälfte der Zeit übrig haben.

Nach den gesamten **10 Minuten** ist mit dem Wort „*Stopp!*“ abzuberechnen.

Die Schüler werden anschließend darum gebeten, im **Testheft unter die zuletzt bearbeitete Aufgabe einen Strich zu ziehen.**

Bitte macht nun wieder einen Strich unter die Aufgabe, die ihr zuletzt bearbeitet habt.

6. UT N-TOP (farbiges Papier)

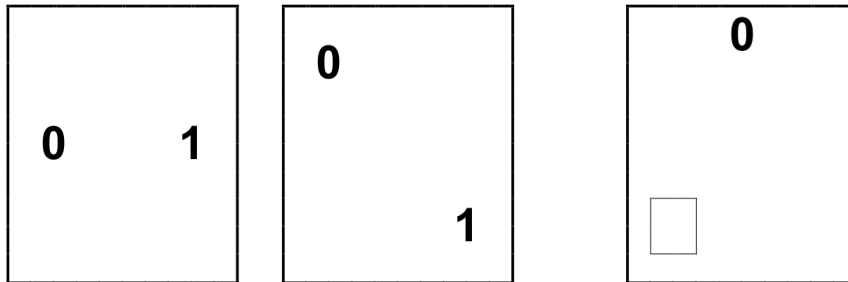
Wenn alle einen Strich unter die letzte Aufgabe gemacht haben, wird wie folgt instruiert:

Blättert jetzt bitte auf die nächste Seite zum Test „N-TOP“ auf dem orangefarbenen Papier.

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

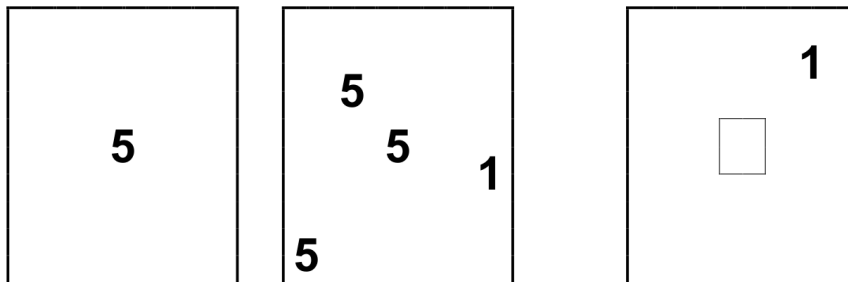
In diesem Test geht es darum, Gesetzmäßigkeiten zu erkennen und entsprechend anzuwenden. Die dabei jeweils gesuchte Zahl ist in das vorgegebene Kästchen einzutragen. Fünf Beispielaufgaben illustrieren das.

Beispielaufgabe a)



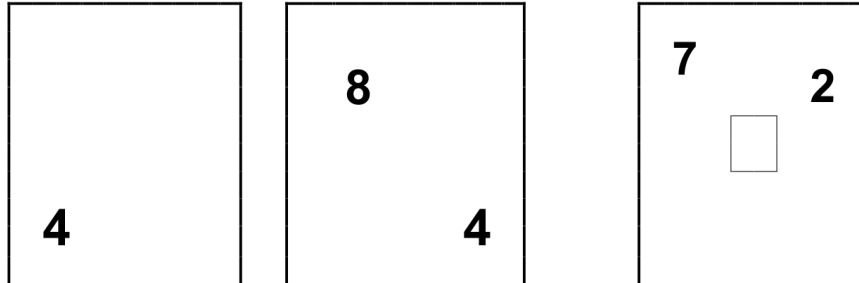
Die Lösung ist „1“. In allen drei Rahmen kommt die Zahl „0“ vor, in den beiden ersten auch die Zahl „1“, also muss folgerichtig auch im dritten Rahmen „1“ vorkommen.

Beispielaufgabe b)



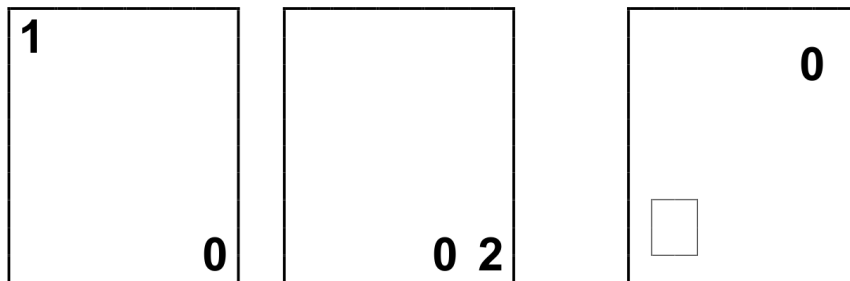
Die Lösung ist „5“. Gemeinsam ist den beiden ersten Rahmen, dass genau in der Mitte die Zahl „5“ steht, also muss folgerichtig auch im dritten Rahmen an dieser Stelle „5“ kommen.

Beispielaufgabe c)



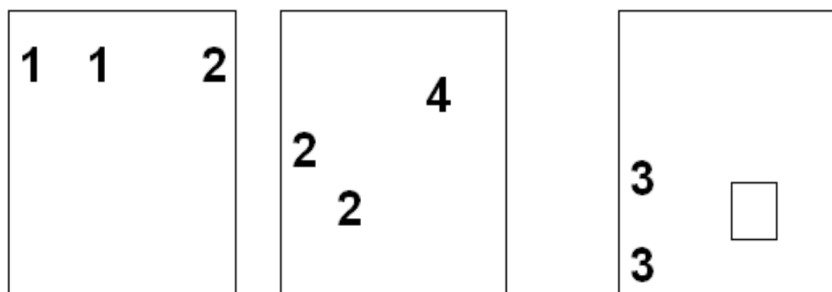
Die Lösung ist „4“. Nichts anderes ist gemeinsam in den beiden ersten Rahmen; folgerichtig muss auch im dritten Rahmen „4“ vorkommen. Die Zahlen „8“, „7“ und „2“ spielen keine Rolle, weil zwischen ihnen keine Beziehung besteht; es könnten auch ganz andere statt diesen verwendet worden sein.

Beispielaufgabe d)



Die Lösung ist „3“. In allen drei Rahmen kommt die Zahl „0“ vor, in den beiden ersten auch die Zahl „1“ bzw. die Zahl „2“, also muss folgerichtig im dritten Rahmen die Zahl „3“ enthalten sein.

Beispielaufgabe e)



Die Lösung ist „6“. Im ersten Rahmen steht eher links zweimal die Zahl „1“ und rechts die Summe davon „2“; im zweiten Rahmen steht eher links zweimal die Zahl „2“ und rechts die Summe davon „4“, so dass im dritten Rahmen mit eher links zweimal der Zahl „3“ folgerichtig rechts die Summe „6“ stehen muss.

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

Für die Testbearbeitung stehen dir 8 Minuten zur Verfügung.

Falls du vor Ablauf der Zeit mit der Bearbeitung der 15 Aufgaben fertig bist, warte bitte auf ein Zeichen des Testleiters.

Gibt es noch Fragen?

Der/die Testleiter(in) gibt ein Startzeichen zum Umblättern, indem er/sie instruiert:

Los geht's. Bitte umblättern.

Für die Bearbeitung des Untertests stehen den Schülern **8 Minuten** zur Verfügung. Nach **4 Minuten** ist den Schülern mitzuteilen, dass sie nun noch die Hälfte der Zeit übrig haben.

Nach den gesamten **8 Minuten** ist mit dem Wort „*Stopp!*“ abzubrechen.

Die Schüler werden anschließend darum gebeten, im **Testheft unter die zuletzt bearbeitete Aufgabe einen Strich zu ziehen**.

Bitte macht nun wieder einen Strich unter die Aufgabe, die ihr zuletzt bearbeitet habt.

7. UT Figurale Analogien (weißes Papier)

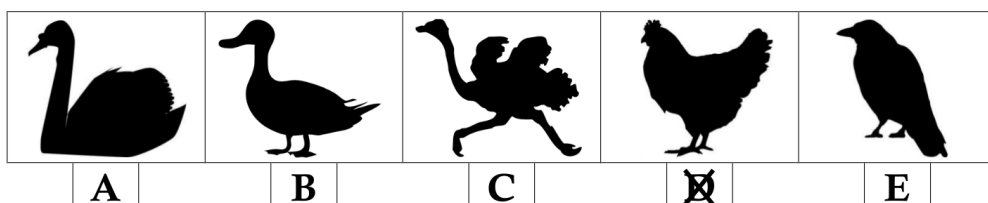
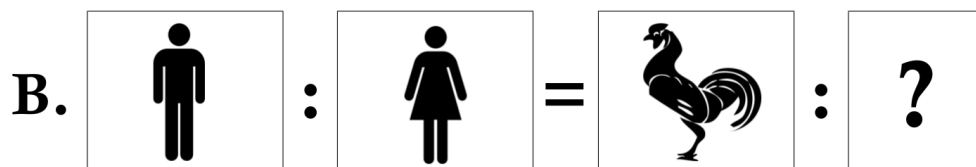
Wenn alle einen Strich unter die letzte Aufgabe gemacht haben, wird wie folgt instruiert:

*Blättert jetzt bitte auf die nächste Seite zum Test „**Figurale Analogien**“ auf dem weißen Papier.*

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

Auf den nächsten Seiten findest du 24 Aufgaben. Bei jeder Aufgabe werden dir drei Bilder vorgegeben. Zwischen dem ersten und zweiten Bild besteht eine Beziehung. Eine ähnliche Beziehung besteht zwischen dem dritten und einem der fünf Wahlbilder. Dieses Bild sollst du herausfinden.

Sieh dir bitte zuerst das folgende Beispiel an.

Beispielaufgabe:

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

Wie bei diesem Beispiel, markiere deine Antwort durch Ankreuzen des Buchstabens unter dem Bild. Beachte bitte, dass immer nur eine Lösung richtig ist.

Für die Bearbeitung stehen dir 8 Minuten zur Verfügung. Falls du vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig bist, warte bitte auf ein Zeichen des Testleiters.

Gibt es noch Fragen?

Der/die Testleiter(in) gibt ein Startzeichen zum Umblättern, indem er/sie instruiert:

Los geht's. Bitte umblättern.

Für die Bearbeitung des Untertests stehen den Schülern **8 Minuten** zur Verfügung. Nach **4 Minuten** ist den Schülern mitzuteilen, dass sie nun noch die Hälfte der Zeit übrig haben.

Nach den gesamten **8 Minuten** ist mit dem Wort „*Stopp!*“ abzubrechen.

Bitte macht nun wieder einen Strich unter die Aufgabe, die ihr zuletzt bearbeitet habt.

8. UT Gleichungen (farbiges Papier)

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

*Blättert jetzt bitte auf die nächste Seite zum letzten Test „**Gleichungen**“ (auf dem orangefarbenen Papier).*

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

Auf der nächsten Seite findest du insgesamt 12 Gleichungen. Aus jeweils vier Vorschlägen sind diejenigen Werte für x und y zu finden, welche die Gleichung lösen.

Beispielaufgabe:

$$x + y = 10$$

x	2	8	5	10
y	9	0	1	3

Die Lösung ist $x = 10$ und $y = 0$. Nur diese beiden Werte ergeben in Summe 10.

Wie bei diesem Beispiel gibt es auch bei allen weiteren Aufgaben immer nur eine Lösung. Bitte markiere diese eindeutig durch Ankreuzen der zutreffenden Zahl. Solltest du eine Auswahl korrigieren wollen, streiche bitte die angekreuzte Zahl durch und markiere deine Lösung, indem du einen Kreis um die Zahl zeichnest. Wenn du bei einer Aufgabe nicht weiterkommst, gehe zur Nächsten über. Zur Lösung der Aufgaben darfst du keine Hilfsmittel (Taschenrechner, Handy usw.) verwenden und dir auch keine Notizen machen.

Für die Testbearbeitung stehen dir 8 Minuten zur Verfügung.

Falls du vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig bist, warte bitte auf ein Zeichen des Testleiters.

Gibt es noch Fragen?

Der/die Testleiter(in) gibt ein Startzeichen zum Umblättern, indem er/sie instruiert:

Los geht's. Bitte umblättern.

Für die Bearbeitung des Untertests stehen den Schülern **8 Minuten** zur Verfügung. Nach **4 Minuten** ist den Schülern mitzuteilen, dass sie nun noch die Hälfte der Zeit übrig haben.

Nach den gesamten **8 Minuten** ist mit dem Wort „Stopp!“ abzubrechen.

Bitte macht nun wieder einen Strich unter die Aufgabe, die ihr zuletzt bearbeitet habt.

Nach diesem letzten Untertest ist wie folgt zu instruieren:

Bitte schließt jetzt alle das Aufgabenheft. Wir sind nun am Ende dieser Testung! Vielen Dank, dass ihr alle so toll mitgemacht habt.

Bitte kontrolliert beim Absammeln der Testhefte noch einmal, ob alle persönlichen Daten am Deckblatt ausgefüllt sind! Bitte darauf achten, dass alle Testpersonen ihr Testheft auch tatsächlich abgeben und niemand eines mit nach Hause nimmt.

Bitte bedankt euch anschließend bei dem/der Lehrer(in), dass wir diese Testung in seiner/ihrer Schulstunde durchführen konnten.

F. Beispiel eines schriftlichen Ergebnisberichts für die Schüler/innen

Probandencode: GSE01

Datum der Testung: 30.10.2014

ERGEBNISBERICHT

Liebe Schülerin, lieber Schüler!

Im Folgenden sind deine Ergebnisse der psychologischen Tests beschrieben, die im Rahmen der Untersuchung von Bernadette Schnait durchgeführt wurden.

Bei den Ergebnissen ist zu beachten, dass diese lediglich eine **Momentaufnahme** darstellen, d.h. sie lassen nur eine Aussage über die zum Zeitpunkt der Testung erbrachte Leistung zu und sind nicht als allgemeingültig zu betrachten. In die Testwerte fließt neben der tatsächlichen Fähigkeit auch noch ein, wie gewissenhaft und ernsthaft die Tests bearbeitet wurden. Je nachdem sind die Ergebnisse also als mehr oder weniger zuverlässige Darstellung deiner Fähigkeiten anzusehen.

Des Weiteren 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 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 bzw. niedrige Fähigkeitsausprägung dar, Prozentränge **über 75** eine **überdurchschnittliche** Leistung bzw. hohe Fähigkeitsausprägung. 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 Maturaklassen allgemeinbildender und berufsbildender höherer Schulen.

NAME DES TESTS UND AUFGABENBESCHREIBUNG	TEST- ERGEBNIS	ERGEBNISBESCHREIBUNG
FORMALE FOLGERICHTIGKEIT (<i>Kubinger & Hagenmüller, in Vorb.</i>)		
Der Test besteht aus Folgen von geometrischen Formen, die logisch korrekt fortgesetzt werden sollen.	PR = 13	Dieser Wert drückt die Fähigkeit aus, aus vorgegebenen Reihen Regeln abzuleiten und daraus korrekte Schlussfolgerungen zu ziehen. Diese Fähigkeit ist bei der Testperson unterdurchschnittlich ausgeprägt.
FÄRBIGER MATRIZENTEST (<i>Undeutsch, 2010</i>)		
In einer 3x3-Matrix bestehend aus farbigen Figuren ist das fehlende Element in der Mitte der Matrix zu ergänzen.	PR = 59	Dieses Ergebnis spiegelt die Fähigkeit wider, aus vorgegebenen Elementen Gesetzmäßigkeiten herauszufinden und diese anzuwenden. Die Testperson hat hier durchschnittlich abgeschnitten.

Probandencode: GSE01
Datum der Testung: 30.10.2014

NUMERISCHE TOPOLOGIEN (Kubinger & Heuberger, in Vorb.)		
In diesem Verfahren werden drei Rahmen mit Zahlen präsentiert, wobei jeweils im letzten Rahmen eine fehlende Ziffer einzutragen ist.	PR = 31	Hiermit wird logisches Denken mithilfe von Zahlenmaterial erfasst. Die Testperson hat eine durchschnittlich Leistung im logisch-schlussfolgernden Denken im numerischen Bereich erbracht.
GLEICHUNGEN (Kubinger & Gamsjäger, in Vorb.)		
Die Aufgaben dieses Tests bestehen aus mathematischen Gleichungen, für deren Lösung die korrekten Werte für x und y gefunden werden müssen.	PR = 99	Dieser Wert zeigt die Fähigkeit, mathematische Gleichungen durch logisches Denken zu lösen. Diese Fähigkeit ist bei der Testperson überdurchschnittlich ausgeprägt.
SPRACHLICHE ANALOGIEN (Kubinger & Ünal, in Vorb.)		
Die Aufgaben bestehen aus jeweils drei Begriffen, wobei zwischen den ersten beiden eine Beziehung besteht. Diese Beziehung ähnelt jener zwischen dem dritten vorgegebenen sowie dem gesuchten Begriff.	PR = 10	Dieser Wert kann als die Fähigkeit interpretiert werden, durch die Anwendung von Allgemeinwissen zu logisch korrekten Lösungen zu kommen. Die Leistung der Testperson ist hier unterdurchschnittlich ausgefallen.
FIGURALE ANALOGIEN (Ünal & Kubinger, in Vorb.)		
Bei diesem Test sind jeweils drei Bilder oder Symbole vorgegeben, wobei zwischen den ersten beiden eine Beziehung besteht. Diese Beziehung soll auf das dritte Bild und eines der fünf Auswahlbilder übertragen werden.	PR = 47	Mit diesem Verfahren wird einerseits Allgemeinwissen, andererseits auch logisch-schlussfolgerndes Denken mithilfe von Bildmaterial erhoben. Die Leistung der Testperson liegt hier im durchschnittlichen Bereich.

Curriculum Vitae

BERNADETTE SCHNAIT, BA

Geburtsdatum: 07.03.1987
E-Mail-Adresse: bernadette.schnait@gmx.at

AUSBILDUNG

seit März 2015 **Psychotherapeutisches Propädeutikum**
Postgraduate Center der Universität Wien

seit März 2010 **Diplomstudium Psychologie**
Universität Wien, Fakultät für Psychologie

Jan. – Dez. 2010 **Ausbildung zum Personal Coach im tiergestützten Setting**
Department für Tiergestützte Medizin Mauritiusshof

Okt. 2006 – April 2010 **Bachelorstudium Transkulturelle Kommunikation**
Universität Wien, Zentrum für Translationswissenschaft
Sprachen: Deutsch, Englisch, Französisch

Sept. 2001 – Juni 2006 **Bundeshandelsakademie Zwettl (NÖ)**

BERUFSERFAHRUNG UND PRAKTIKA

Okt. 2014 – Juni 2015 **Studienassistentin am Arbeitsbereich Psychologische Diagnostik**
Institut für Angewandte Psychologie: Gesundheit, Entwicklung und Förderung, Fakultät für Psychologie, Universität Wien

seit Aug. 2010 **Übersetzerin und Korrekturleserin**
ROMA Friseurbedarf – Robert Maurer GmbH

Mai – Sept. 2013 **Praktikum in der Praxisgemeinschaft Salvatorgasse**
Klinische Psychologie, Gesundheitspsychologie, Klinische Neuropsychologie