



universität
wien

DIPLOMARBEIT

Titel der Arbeit

Figurale Analogien:
ein Test zur Messung des kristallisierten schlussfolgernden
Denkens mit Hilfe figuralen Materials

Verfasserin

Serra Ünal

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, in 2014

Studienkennzahl: 298

Studienrichtung: Psychologie

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

ABSTRACT (DEUTSCH)

Der neu entwickelte Test *Figurale Analogien* soll zur Erfassung des kristallisierten schlussfolgernden Denkens im figuralen Bereich dienen. Die präsentierten Analogie-Aufgaben sind durch Auswahl von einer der fünf Antwortmöglichkeiten zu lösen, wobei das Arbeitsmaterial aus inhaltlich sinnvollen Bildern besteht. Insgesamt wurden 27 Items konstruiert, die in der Diagnostik des Allgemeinwissens bei Erwachsenen Anwendung finden sollen.

An der Untersuchung, die in den Sicherheitsakademien der Polizei in Niederösterreich stattfand, nahmen 331 Auszubildende zwischen 18 und 32 Jahren teil. Diese bearbeiteten die neu konstruierten Items, sowie die Verfahren *AN-TOP* (Kubinger & Heuberger, in Vorb.) und *Gleichungen* (s. vorläufig bei Gamsjäger, 2012). Mit diesen beiden Tests soll logisch-schlussfolgerndes Denken erfasst werden. Die Ergebnisse von *AN-TOP* werden dabei zur Bestimmung der diskriminanten Validität von *Figurale Analogien* herangezogen.

Die anschließenden testtheoretischen Analysen wurden zur Skalierung nach dem dichotomen logistischen Modell von Rasch durchgeführt. Den Items von *Figurale Analogien* kann dabei a-priori und dem Itempool von *AN-TOP* a-posteriori, nach Ausscheiden von ca. vier Prozent der Items, Modellgeltung zugesprochen werden. Der Korrelationskoeffizient nach Pearson zwischen den Personenparametern dieser Verfahren beträgt $r = 0,217$.

Schlüsselwörter: kristallisiert, schlussfolgerndes Denken, Allgemeinwissen, figurale Analogien, Rasch-Modell

ABSTRACT (ENGLISH)

This paper presents a new psychological test, called *Figurale Analogien*, for the measurement of crystallized deductive reasoning skills in adults. 27 figural analogy questions were constructed that make use of non-abstract pictures about different subjects of general knowledge. To answer these questions one must first recognize the relationship between the first two illustrations and then complete the analogy by picking one of the five response alternatives.

A total of 331 trainees of the Federal Security Academy in Lower Austria between the ages of 18 and 32 participated in the study. They were administered *Figurale Analogien* as well as two logical reasoning tests: *AN-TOP* (Kubinger & Heuberger, in progress), and *Gleichungen* (for preliminary results see Gamsjäger, 2012). In this study, the results of *AN-TOP* are to be employed to determine the discriminant validity of *Figurale Analogien*.

Subsequent analyses were performed in regards to the fit of items to the Rasch model. This model showed to be an appropriate fit for the items of *Figurale Analogien*. The same was true for the itempool of *AN-TOP* after the exclusion of two items. The correlation between the person parameters of *Figurale Analogien* and *AN-TOP* is $r = 0.217$.

Keywords: crystallized, reasoning, general knowledge, figural analogies, rasch model

INHALTSVERZEICHNIS

I. EINLEITUNG	1
II. THEORETISCHER TEIL.....	2
1. Intelligenztheoretische Grundlage	2
1.1. Die Zwei-Faktoren-Theorie von Spearman.....	2
1.2. Hierarchische Faktorenmodelle	3
1.3. Das Modell mehrerer gemeinsamer Faktoren von Thurstone	4
1.4. Das Strukturmodell der Intelligenz von Guilford.....	5
1.5. Das Modell der fluiden und kristallisierten Intelligenz von Cattell.....	6
2. Kristallisiertes schlussfolgerndes Denken	8
2.1. Investment-Theorie	8
2.2. Analogien und logisch-schlussfolgerndes Denken	10
2.3. Definition von kristallisiertem schlussfolgerndem Denken.....	12
3. Wissensdiagnostik.....	14
3.1. Wissenspsychologie	14
3.2. Erfassung des Wissens	14
3.3. I-S-T 2000 R.....	16
4. Testtheoretische Grundlage.....	20
4.1. Item-Response-Theorie.....	20
4.2. Rasch-Modell	20

III. EMPIRISCHER TEIL.....	24
5. Ziele der Untersuchung und Hypothesen.....	24
6. Konstruktion von <i>Figurale Analogien</i>	25
6.1. Itemkonstruktion.....	25
6.1.1 Geltungsbereich und Vorgabe	25
6.1.2 Aufbau der Aufgaben.....	26
6.1.3 Antwortformat	27
6.1.4 Erstellung des Aufgabenmaterials	29
6.1.5 Themenbereiche und Schwierigkeitsstufen	31
6.2. Instruktion und Bearbeitungszeit.....	34
7. <i>Alpha Numeric Topologies (AN-TOP)</i>	36
8. Planung und Durchführung der Untersuchung	38
9. Stichprobe	40
10. Darstellung der Ergebnisse.....	42
10.1. Beschreibung des Vorgangs bei den Rasch-Modell Analysen	42
10.2. Ergebnisse der Rasch-Modell Analysen bei <i>Figurale Analogien</i>	43
10.3. Ergebnisse der Rasch-Modell Analysen bei <i>AN-TOP</i>	48
10.4. Deskriptive Ergebnisse zu <i>Figurale Analogien</i>	56
10.5. Deskriptive Ergebnisse zu <i>AN-TOP</i>	59
10.6. Korrelation zwischen <i>Figurale Analogien</i> und <i>AN-TOP</i>	60
IV. DISKUSSION UND AUSBLICK	62
V. ZUSAMMENFASSUNG.....	67
VI. LITERATURVERZEICHNIS	69
VII. ANHANG	75

ABBILDUNGSVERZEICHNIS

ABBILDUNG 1: Veranschaulichung der Investment-Theorie	9
ABBILDUNG 2: Beispiel einer einfachen figuralen Analogie-Aufgabe zu logisch-schlussfolgerndem Denken (Guilford & Hoepfner, 1976, S. 38)	12
ABBILDUNG 3: Grundstruktur der Fragestellung bei <i>Figurale Analogien</i>	26
ABBILDUNG 4: Präsentationsform der Antwortmöglichkeiten bei <i>Figurale Analogien</i>	27
ABBILDUNG 5: Instruktionsitem bei <i>Figurale Analogien</i>	30
ABBILDUNG 6: Itembeispiel von <i>AN-TOP</i> (Kubinger & Heuberger, in Vorb.).....	37
ABBILDUNG 7: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>Figurale Analogien</i> , alle Items)	45
ABBILDUNG 8: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (<i>Figurale Analogien</i> , alle Items).....	45
ABBILDUNG 9: Grafische Modellkontrolle für das Teilungskriterium „Alter“ (<i>Figurale Analogien</i> , alle Items)	46
ABBILDUNG 10: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>AN-TOP</i> , alle Items)	49
ABBILDUNG 11: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (<i>AN-TOP</i> , alle Items)	50
ABBILDUNG 12: Grafische Modellkontrolle für das Teilungskriterium „Alter“ (<i>AN-TOP</i> , alle Items)	50
ABBILDUNG 13: Grafische Modellkontrolle für das Teilungskriterium „Score“ (<i>AN-TOP</i> , nach Ausschluss von Items 11a und 20b)	53
ABBILDUNG 14: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“ (<i>AN-TOP</i> , nach Ausschluss von Items 11a und 20b).....	54
ABBILDUNG 15: Grafische Modellkontrolle für das Teilungskriterium „Alter“ (<i>AN-TOP</i> , nach Ausschluss von Items 11a und 20b)	54
ABBILDUNG 16: Verteilung der relativen Häufigkeiten der Rohwerte bei <i>Figurale Analogien</i> ($n = 331$)	57
ABBILDUNG 17: Verteilung der Itemschwierigkeitsparameter bei <i>Figurale Analogien</i>	57
ABBILDUNG 18: Verteilung der relativen Häufigkeiten der Rohwerte bei <i>AN-TOP</i> ($n = 331$)	59
ABBILDUNG 19: Verteilung der Itemschwierigkeitsparameter bei <i>AN-TOP</i>	60

TABELLENVERZEICHNIS

TABELLE 1: Jeweilige Position der Lösung bei jedem Item innerhalb der fünf Antwortmöglichkeiten	29
TABELLE 2: Häufigkeiten der für die Erstellung des Arbeitsmaterials verwendeten Quellen	31
TABELLE 3: Anzahl der Items pro Themenbereich und angenommene Schwierigkeitsstufe.....	33
TABELLE 4: Ausschnitt des Testdesigns von <i>AN-TOP</i>	36
TABELLE 5: Die Verteilung der Stichprobe in Bezug auf das Alter in Jahren ($n = 331$)	40
TABELLE 6: Die Verteilung der Stichprobe in Bezug auf die höchste abgeschlossene Ausbildung ($n = 331$)	41
TABELLE 7: Bei den Rasch-Modell Analysen nicht berücksichtigte Items von <i>Figurale Analogien</i>	44
TABELLE 8: Ergebnisse der Likelihood-Ratio-Tests (<i>Figurale Analogien</i> , alle Items).....	44
TABELLE 9: Schätzungen der Itemschwierigkeitsparameter bei <i>Figurale Analogien</i>	47
TABELLE 10: Bei den Rasch-Modell Analysen nicht berücksichtigte Items von <i>AN-TOP</i>	48
TABELLE 11: Ergebnisse der Likelihood-Ratio-Tests (<i>AN-TOP</i> , alle Items)	49
TABELLE 12: Ergebnisse des z-Tests für das Teilungskriterium „Score“ (<i>AN-TOP</i> , alle Items)	51
TABELLE 13: Ergebnisse der Likelihood-Ratio-Tests (<i>AN-TOP</i> , nach Ausschluss von Item 11a) .	52
TABELLE 14: Ergebnisse der Likelihood-Ratio-Tests (<i>AN-TOP</i> , nach Ausschluss von Items 11a und 20b).....	53
TABELLE 15: Schätzungen der Itemschwierigkeitsparameter bei <i>AN-TOP</i>	55
TABELLE 16: Häufigkeiten der Rohwerte bei <i>Figurale Analogien</i>	56
TABELLE 17: Lösungshäufigkeiten der Items von <i>Figurale Analogien</i>	58

I. EINLEITUNG

Analogien gehören zu den beliebtesten Arbeitsmaterialien in der Intelligenzdiagnostik. Mit ihnen wird, abhängig von der dahinterliegenden Intelligenztheorie, hauptsächlich logisch-schlussfolgerndes Denken erfasst. Das Ziel dieser Arbeit besteht darin, ausgehend von einer Idee von Univ.-Prof. Dr. Mag. K. D. Kubinger, figurale Analogie-Aufgaben zu entwickeln, die zur Erfassung des *kristallisierten schlussfolgernden Denkens* bei Erwachsenen eingesetzt werden können.

Laut dem Intelligenzmodell von Cattell (1961, 1963) werden mit kristallisierter Intelligenz jene Fähigkeiten widerspiegelt, die als Produkt von vorangegangenen und umweltbedingten Lern- und Sozialisationsprozessen zu betrachten sind. Die Erfassung dieser Fähigkeiten erfolgt in der Regel mit Hilfe von Wissenstests, die nach der Erhebungsmethode sowie dem Allgemeinheitsgrad und der Kodierung des erfassten Wissens voneinander unterschieden werden können (Beauducel & Süß, 2011). Die Anzahl deutschsprachiger Verfahren, die bei Erwachsenen das Allgemeinwissen erfassen, ist eher beschränkt. Als Beispiel für ein solches Verfahren ist der Intelligenz-Struktur-Test 2000 R (Amthauer, Brocke, Liepmann, & Beauducel, 2001) zu erwähnen, bei der die Möglichkeit der Wissenserfassung auch im figuralen Bereich angeboten wird. Der neu entwickelte Test *Figurale Analogien* soll sich von diesem darin unterscheiden, dass sowohl die gestellten Fragen als auch die Antwortmöglichkeiten zur Gänze aus Bildmaterial bestehen.

Von üblichen Analogie-Aufgaben sollen sich die Items von *Figurale Analogien* durch den sinnvollen Inhalt des Arbeitsmaterials abheben. Obwohl Analogien, bei denen die Fragestellung durch den Inhalt der Stimuli bestimmt wird, im verbalen Bereich sehr oft eingesetzt werden, dienen hingegen im figuralen Bereich in der Regel abstrakte, geometrische Objekte als Arbeitsmaterial.

Weiters soll die Skalierung des Verfahrens nach dem dichotomen logistischen Testmodell von Rasch erfolgen. Damit soll belegt werden, dass das neu entwickelte Verfahren *Figurale Analogien* eindimensional, d.h. nur kristallisiertes schlussfolgerndes Denken, misst.

II. THEORETISCHER TEIL

1. Intelligenztheoretische Grundlage

Dieser Abschnitt bietet einen kurzen Überblick über die Intelligenztheorien, die im Rahmen dieser Arbeit als Ausgangspunkt für die Definition der zu erfassenden Fähigkeit dienen sollen. Desweiteren wird auf die Rolle von Analogie-Aufgaben und des Allgemeinwissens bei diesen Theorien eingegangen.

1.1. Die Zwei-Faktoren-Theorie von Spearman

Spearman, der den Weg für die Anwendung der Faktorenanalyse in Psychologie bahnte, präsentierte mit seiner Zwei-Faktoren-Theorie das erste explizite Modell der Intelligenz (Stemmler, Hagemann, Amelang, & Bartussek, 2011). Er postulierte, dass allen Leistungen bei Tests, die der Kategorie der Intelligenz zugeordnet werden können, ein gemeinsamer Faktor (*g factor*, abgeleitet von *general intelligence*) zugrunde liege. Den Beweis für die Existenz dieses Generalfaktors sah er in der hohen positiven Korrelation, die zwischen den Leistungen einer Probandengruppe bei unterschiedlichen kognitiven Tests beobachtet werden kann (Spearman, 1904, 1914). Somit variere der Anteil von *g* zwar von Person zu Person, bleibe aber für ein Individuum hinsichtlich der korrelierten Leistungen gleich (Holling, Preckel, & Vock, 2004).

Die Varianz der Testleistungen, die nicht durch *g* erklärt werden kann, wurde von Spearman dem sogenannten spezifischen Faktor (*s factor*) attribuiert. Diese spezifische Komponente ist sowohl für einen bestimmten Test bzw. jede einzelne Aufgabenart charakteristisch, als auch, mit Ausnahme von Parallelförmigkeiten des gleichen Tests, bei jedem anderen Test unterschiedlich (Guilford & Hoepfner, 1976; Spearman, 1914).

Spearman (1904) berücksichtigte zur Formulierung seiner Theorie viele unterschiedliche Fähigkeiten (z.B. Fähigkeit zur Diskrimination von Tonstufen, Schulnoten in diversen Fächern wie Latein und Mathematik), die der Kategorie der Intelligenz zugeordnet werden könnten. Er vertritt jedoch die Meinung, dass zur Erfassung des *g*-Faktors am besten

solche Aufgaben geeignet sind, zur deren Lösung „*eduction of relations and correlates*“ benötigt werden (Spearman, 1927, S. 177). Diese Aufgaben erfordern die Fähigkeit, entweder die Beziehung zwischen zwei Stimuli schlussfolgernd herauszufinden (z.B. Haus zu Dach $\rightarrow$?; *Antwort*: Gesamtheit zu Teil) oder aus einer gegebenen Beziehung und einem Stimulus auf den zweiten zurückschließen zu können (z.B. kalt, ? $\rightarrow$ Gegenteil; *Antwort*: heiß) (Guilford, 1971, S. 57 f.). Aus diesem Grund werden vor allem Analogien, figurale Matrizenaufgaben und numerische Reasoningaufgaben zur Erfassung des *g*-Faktors sensu Spearman eingesetzt (Holling et al., 2004; Wechsler, 1961). Beispiele dafür befinden sich bei den Matrizentests von Raven wie SPM und APM, die logisch-schlussfolgerndes Denken messen (Heller, Kratzmeier, & Lengfelder, 1998a, 1998b).

Die Vorstellung eines Generalfaktors liegt heute noch vielen gängigen Intelligenztests zu Grunde, so dass der Punktwert, der bei einigen Testbatterien für den Gesamttest ausgegeben wird, mit dem *g*-Faktor von Spearman verglichen werden könnte (Stemmler et al., 2011). Jedoch wurde die Zwei-Faktoren-Theorie seit ihrer Publikation auch öfters kritisiert (siehe dazu z.B. Amelang & Bartussek, 1997; Guilford, 1971; Wechsler, 1961). Als besonders problematisch wurden Interkorrelationen zwischen Tests betrachtet, die darauf deuteten, dass Tests aus einer Gruppe zusätzlich zum *g*-Faktor auch etwas anderes gemeinsam hatten. Spearman bezeichnete diese als „Gruppenfaktoren“ und leitete somit andere Erklärungsversuche in die Wege (Guilford & Hoepfner, 1976; Spearman, 1927).

1.2. Hierarchische Faktorenmodelle

Mit den Restkorrelationen nach der Extraktion des Generalfaktors setzten sich Burt (1949, zitiert nach Guilford, 1971) und Vernon (1950) in Form von hierarchischen Faktorenmodellen auseinander. Die Annahme, dass intellektuelle Fähigkeiten unterschiedliche Grade an Allgemeinheit aufweisen, entstand durch Erhebungen, bei denen, im Vergleich zu Spearman, heterogenere Merkmalsbereiche erfasst wurden (Guilford & Hoepfner, 1976). Bei hierarchischen Faktorenmodellen stellt der *g*-Faktor die Ebene des höchsten Allgemeingrades dar, unter der die anderen Faktoren gruppenweise angesiedelt werden (Amelang & Bartussek, 1997).

In dem Modell von Vernon (1950) wird der *g*-Faktor in zwei „major group factors“ geteilt, *v:ed* (verbal-educational) und *k:m* (spatial-mechanical). Der erste Faktor (*v:ed*) wird in verbale und numerische Fähigkeiten, der zweite (*k:m*) in psycho-motorische, technische, räumliche und physikalische Fähigkeiten untergliedert. Auf diese sogenannten „minor group factors“ folgen in der Hierarchie die „specific factors“, die mit dem *s*-Faktor von Spearman vergleichbar sind und denen in hierarchischen Modellen eine eher geringe Bedeutung beigemessen wird.

Hierarchische Faktorenmodelle wurden durch die Positionierung eines varianzstarken *g*-Faktors als ihren Ausgangspunkt einer ähnlich scharfen Kritik wie Spearman ausgesetzt. Nichtsdestotrotz diente die Gruppierung von vielen spezifischen Intelligenzfaktoren in wenige und breitere Faktoren höherer Ordnung als Grundlage für die Entwicklung diverser Intelligenztests. Dabei stellen die Intelligenztestbatterien von Wechsler das im deutschen Sprachraum am häufigsten eingesetzte Beispiel dar (Stemmler et al., 2011). Sowohl WAIS-IV (*Wechsler Adult Intelligence Scale IV*, Petermann, 2012) als auch WISC-IV (*Wechsler Intelligence Scale for Children IV*, Petermann & Petermann, 2011) beinhalten den Untertest „Allgemeines Wissen“, der mit Fragen zu allgemein bekannten Ereignissen, Sachverhalten, Orten und Persönlichkeiten allgemeines Faktenwissen im verbalen Bereich erfasst.

1.3. Das Modell mehrerer gemeinsamer Faktoren von Thurstone

Thurstone (1932, zitiert nach Thurstone, 1961) vertrat in seinem Modell die Meinung, dass ein einziger Faktor für eine adäquate Erklärung der Intelligenz sowie für den prognostischen Einsatz von psychologischen Tests nicht ausreichend sei. Hingegen seien an der Bearbeitung von Denkaufgaben immer mehrere, im Gegensatz zu einer hierarchischen Anordnung, nebeneinanderstehende Gruppenfaktoren in unterschiedlichen Gewichtungsverhältnissen beteiligt.

Mit Hilfe der Multiplen Faktorenanalyse stellte Thurstone sieben Gruppenfaktoren fest, die er als „Primärfähigkeiten“ (engl. „*primary abilities*“) bezeichnete (Stemmler et al., 2011, vgl. dazu z.B. Guilford & Hoepfner, 1976; Thurstone, 1961):

1. *I (inductive reasoning)*: schlussfolgerndes Denken, Erkennen von Regeln
2. *M (memory)*: Merkfähigkeit, Kurzzeitgedächtnis
3. *N (number)*: Rechenfähigkeit, Fähigkeit, mit Zahlen umzugehen
4. *P (perceptual speed)*: Wahrnehmungsgeschwindigkeit
5. *S (space)*: räumliches Vorstellungsvermögen
6. *V (verbal)*: verbales Verständnis, Sprachbeherrschung
7. *W (word fluency)*: Wortflüssigkeit

An Thurstone wurde insbesondere für sein methodisches Vorgehen bei der Extraktion der Gruppenfaktoren Kritik geübt (siehe dazu Amelang & Bartussek, 1997; Guilford, 1976; Spearman, 1939; Vernon, 1950). Trotzdem wurde seine Vorstellung, dass die Leistung in einer bestimmten Aufgabe nicht von allen Primärfähigkeiten gleichzeitig determiniert werde, verschiedenen Intelligenztests zugrunde gelegt (Amelang & Bartussek, 1997; Thurstone, 1961). Als Beispiele sind das Leistungs-Prüf-System (L-P-S, Horn, 1983) und die älteren Versionen des Intelligenz-Struktur-Tests (I-S-T, Amthauer, 1953 und I-S-T 70, Amthauer, 1970) zu erwähnen.

1.4. Das Strukturmodell der Intelligenz von Guilford

Guilford (1956) postulierte, dass intellektuelle Prozesse nicht in Form einer hierarchischen Struktur, sondern in drei Dimensionen (*Operationen, Inhalte, Produkte*) beschrieben, klassifiziert und erklärt werden sollten. Diese drei Dimensionen lassen sich in dem Strukturmodell der Intelligenz graphisch in Form eines dreidimensionalen Würfels veranschaulichen (siehe dazu z.B. Guilford, 1971, S. 63) und beinhalten folgende Unterteilungen (Guilford & Hoepfner, 1976):

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

In diesem Modell wird jede Fähigkeit durch seine einzigartige Kombination aus jeweils einer Art der drei Dimensionen repräsentiert. Aus der Kombination von fünf Operationen

mit vier Inhalten und sechs Produkten resultieren somit 120 Faktoren (Guilford, 1971). Dabei wird jedoch nicht angenommen, dass diese „den ganzen Bereich intellektueller Züge oder Variablen abdecken“ (Guilford & Hoepfner, 1976, S. 33).

Obwohl eine standardisierte Testbatterie zur Erfassung von allen möglichen Kombinationen bis heute nicht existiert, können mit Hilfe dieses Modells bestehende Verfahren beschrieben und die untersuchten Fähigkeiten genauer klassifiziert werden (Funke & Vaterrodt-Plünnecke, 1998). So verglich Guilford (1956) die bei der Bearbeitung von Analogie-Aufgaben benötigte Fähigkeit mit dem Konzept von „*eduction of relations and correlates*“ von Spearman und ordnete sie bei figuralem Arbeitsmaterial der Operation „*Kognition*“ bzw. „*konvergente Produktion*“, dem Inhalt „*figural*“ und dem Produkt „*Beziehungen*“ zu.

Beim Betrachten einer Analogie wird in erster Linie die Operation „Kognition“ ausgelöst, die als das schnelle Entdecken bzw. Wiedererkennen von Informationen definiert wird. Erst nachdem das Produkt „Beziehung“, d.h. die Verbindung zwischen den präsentierten Informationen, aufgefunden wird, kommt es zu der Operation „Konvergente Produktion“, mit der definitionsgemäß die Entwicklung einer logischen Schlussfolgerung zum Erreichen der besten Lösung beschrieben wird. Der Inhalt „figural“ bezieht sich auf die Präsentationsform der Stimuli (siehe dazu z.B. Guilford, 1971, S. 85 f.; Guilford & Hoepfner, 1976, S. 37 ff.).

1.5. Das Modell der fluiden und kristallisierten Intelligenz von Cattell

Basierend auf der Annahme der hierarchischen Faktorenmodelle, dass *g* als einziger Faktor sowohl ungenügend zur Beschreibung der menschlichen Intelligenz als auch schwer zu bestimmen sei, ging Cattell (1961) von mehreren Faktoren aus, die zwar hierarchisch positioniert, jedoch nicht völlig voneinander unabhängig seien. Cattell (1963) zog die Arbeiten und Kritik von Thurstone, Vernon und Guilford in Erwägung und bestimmte durch seine Untersuchungen die „*fluide*“ und die „*kristallisierte*“ Intelligenz als die zwei wichtigsten Faktoren, auf denen Intelligenztests basieren (vgl. dazu Horn & Cattell, 1966, 1982).

Die fluide Intelligenz („*fluid intelligence*“, g_f^1) vereinigt kognitive Fähigkeiten, die benötigt werden, um sich neuen Situationen anpassen und neuartige Probleme lösen zu können. Ähnlich wie Spearman's Konzept der „*eduction of relations and correlates*“ bildet sie die Grundlage für wichtige mentale Prozesse wie abstraktes Denken, Problemlösen und logisch-schlussfolgerndes Denken. Diese Komponente basiert auf der bestimmenden Wirkung von biologischen und physiologischen Faktoren (Vererbung, Störungen des Zentralnervensystems oder der Sinnesorgane etc.) auf die intellektuelle Entwicklung und wird daher von früheren Lernerfahrungen wenig beeinflusst (Cattell, 1963, Horn & Cattell, 1966).

Die kristallisierte Intelligenz („*crystallized intelligence*“, g_c) spiegelt hingegen jene kognitiven Fertigkeiten wider, die als Produkt von vorangegangenen und umweltbedingten Lern- und Sozialisationsprozessen zu sehen sind. Sie wird hauptsächlich von Interessen und Routineaktionen der untersuchten Person und wenig von physiologischen Faktoren beeinflusst. Deswegen sollte die Fluktuation dieser Fähigkeiten während des Tages wegen z.B. Müdigkeit oder gesundheitlicher Umstände geringer ausfallen als bei fluider Intelligenz. Da der Schwerpunkt auf kulturellen und bildungsspezifischen Aspekten liegt, wird angenommen, dass g_c , im Gegensatz zu fluider Intelligenz, ihren Kulminationspunkt im individuellen Lebenslauf später erreiche und über die Zeit nur schwach abnehme (Cattell, 1961, 1963). Besonders diese Annahme hebt die Überlegungen von Cattell von den anderen erwähnten Intelligenztheorien hervor, da sie einen Rahmen für die Betrachtung und Erfassung menschlicher Intelligenz im Erwachsenenalter erlaubt (Ackerman, 1996).

Zur Erfassung fluider Intelligenz werden häufig zeitlich beschränkte Verfahren mit neuartigen, figuralen bzw. abstrakten Stimuli eingesetzt, die „*culture-fair*“² die logisch-schlussfolgernde Denkfähigkeit überprüfen (Cattell, 1961). Demgegenüber wird die kristallisierte Intelligenz meist ohne zeitliche Beschränkung durch Verfahren erfasst, die kulturspezifische Fähigkeiten erfordern (Cattell, 1987; Horn & Cattell, 1966).

¹ Basierend auf Publikationen von Cattell wird hier als Abkürzung für seine Intelligenzfaktoren der Kleinbuchstabe g mit tiefgestelltem f bzw. c verwendet. Es sei jedoch darauf hingewiesen, dass in moderner Literatur auch die Schreibweisen Gf und Gc bzw. $G(F)$ und $G(C)$ anzutreffen sind.

² „*Culture-fairness*“ beschreibt Intelligenztests mit dem Bestreben, die Möglichkeit einer Benachteiligung von Angehörigen bestimmter Sozialschichten in Bezug auf die geforderte Sprachkompetenz bzw. Thematik der Stimuli zu minimieren (Kubinger, 2009; Süß, 2003).

2. Kristallisiertes schlussfolgerndes Denken

Die vorliegende Arbeit stellt ein neues Verfahren zur Erfassung des *kristallisierten schlussfolgernden Denkens* vor. Die im letzten Abschnitt erwähnten Intelligenztheorien, insbesondere die Theorie der fluiden und kristallisierten Intelligenz von Cattell, legen die Vermutung nahe, dass *kristallisierte Intelligenz* und *logisch-schlussfolgerndes Denken* zwei gegensätzliche Dimensionen der menschlichen Intelligenz darstellen. Deswegen soll in diesem Abschnitt zuerst die Beziehung zwischen fluider und kristallisierter Intelligenz näher betrachtet werden. Anschließend wird auf die Tradition des Einsatzes von Analogie-Aufgaben zur Erfassung des logisch-schlussfolgernden Denkens eingegangen, um letztlich die Definition von kristallisiertem schlussfolgerndem Denken im Rahmen dieser Arbeit klar hervorzuheben.

2.1. Investment-Theorie

Mit der Investment-Theorie präsentiert Cattell (1987, S. 138 ff.) nicht nur eine Erweiterung der hierarchischen Struktur seines Intelligenzkonzepts (siehe Abschnitt 1.5), sondern setzt sich auch mit der Beziehung zwischen fluider (g_f) und kristallisierter (g_c) Intelligenz auseinander. Den Ausgangspunkt dieser Theorie bilden die Ergebnisse seiner Untersuchungen, die unabhängig vom Alter auf positive Korrelationen ($r = \text{ca. } 0.4 - 0.5$) zwischen g_f und g_c deuteten (Cattell, 1987, S. 116). Diese führten zu dem Erklärungsansatz, dass die intellektuelle Entwicklung eines Individuums mit einer einzigen, allgemeinen Fähigkeit zum Erkennen von Beziehungen beginne, die von der Funktionsweise des Gehirns sowie genetischen Faktoren abhängt. Diese Fähigkeit, die von Cattell als g_f bezeichnet wird, beeinflusse das Lerntempo im Kindesalter und diene auch als Grundlage für den Erwerb von komplexen und spezialisierten Fähigkeiten, die sich durch Übung und Erfahrung verfestigen bzw. „kristallisieren“. Demnach wird g_f in den Erwerb von g_c „investiert“ (Cattell, 1987). Gleichzeitig unterstützen diese komplexe und spezialisierte Fähigkeiten zusammen mit diversen anderen Faktoren (Motivation, Interessen, Lernumstände, sozioökonomischer Hintergrund etc.) die Entwicklung und das zielgerichtete Einsetzen von fluider Intelligenz. Somit hängen die Faktoren g_f und g_c theoretisch miteinander zu-

sammen (vgl. dazu Abbildung 1). In Cattells (1963) Worten zusammengefasst: „learning begets learning capacity“ (S. 5).

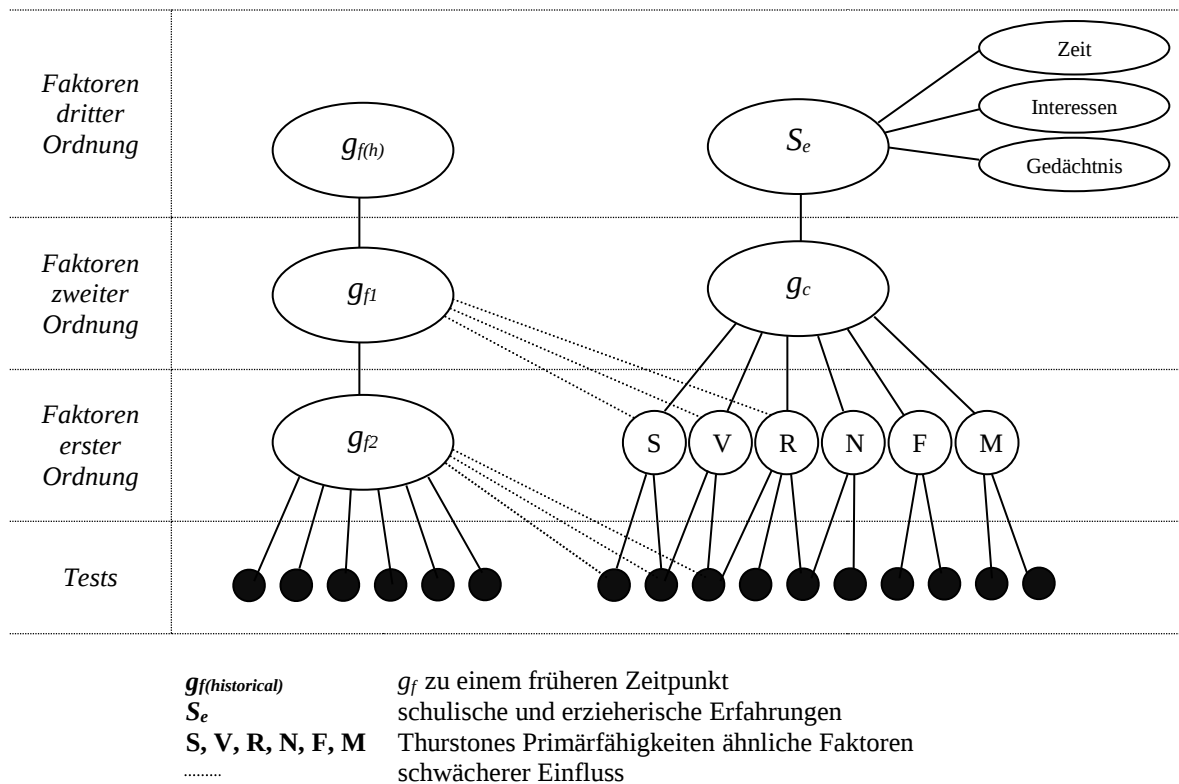


Abbildung 1: Veranschaulichung der Investment-Theorie
(vereinfachte Darstellung nach Cattell, 1987, S. 146)

Die Investment-Theorie veranlasste Cattell (1963, 1987) auch zu Annahmen über den Entwicklungsverlauf von g_f und g_c . Demnach entfalte sich fluide Intelligenz im Hinblick auf das Lebensalter rascher, erreiche ihren Höhepunkt früher und regrediere mit zunehmendem Alter. Hingegen zeige kristallisierte Intelligenz, die erst durch g_f zum Vorschein kommt, sowohl einen späteren Kulminationspunkt als auch ein kontinuierliches Wachstum über die Lebensspanne mit vernachlässigbarer Abnahme im Alter.

Viele Arbeiten der Gegenwart befassen sich mit der allgemeinen Theorie der fluiden und kristallisierten Intelligenz sowie der Investment-Theorie und berichten über theorieunterstützende Ergebnisse (vgl. dazu z.B. Nisbett, Aronson, Blair, Dickens, Flynn, Halpern, & Turkheimer, 2012). In einer Längsschnittstudie wurden g_f und g_c mit Hilfe Rasch-Modellkonformer Items untersucht und als voneinander trennbare Faktoren identifiziert, die Cat-

tells Annahmen entsprechende Entwicklungsverläufe zeigen (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002). Zusätzlich zu generell mit der Investment-Theorie übereinstimmenden Ergebnissen zeigten Kvist und Gustafsson (2008) den Einfluss kultureller und bildungsspezifischer Faktoren auf die kristallisierte Intelligenz, indem sie diese Fähigkeit bei heterogenen Populationen mittels verbaler Stimuli erfassten.

In Bezug auf das Allgemeinwissen („*general knowledge*“) stellten Chamorro-Premuzic, Furnham und Ackerman (2006) fest, dass es, wie andere kristallisierte Fähigkeiten, aus dem Einsatz von g_f über die Zeit resultiert. Darauf haben die Persönlichkeitsfaktoren „Offenheit für Erfahrung“, „Introversion“ und „Typical Intellectual Engagement“³ einen positiven und „Neurotizismus“ einen negativen Einfluss (vgl. Furnham, Swami, Arceche, & Chamorro-Premuzic, 2008). Die Komplexität und wissenschaftliche Relevanz der Investment-Theorie wurde zusätzlich durch die Identifizierung mehrerer Variablen unterstrichen, die die wechselseitige Beziehung zwischen g_f und g_c beeinflussen können, wie z.B. Lernfähigkeit (Schweizer & Koch, 2002), elterlicher sozio-ökonomischer Status und Bildungsniveau (Rindermann, Flores-Mendoza, & Mansur-Alves, 2010), Gedächtnis (McArdle, Hamagami, Meredith, & Bradway, 2000), Persönlichkeit und Interessen (Ackerman, 1996; Ackerman & Heggstad, 1997).

2.2. Analogien und logisch-schlussfolgerndes Denken

Eine Analogie ist ein Vergleich zwischen zwei Objekten in Bezug auf ihre wahrgenommene Ähnlichkeit. Da Analogien bei alltäglichen Denkprozessen eine wichtige Stellung einnehmen, spielen sie schon seit Beginn der Intelligenzforschung eine beachtliche Rolle bei der Entwicklung von Intelligenztests (Sternberg, 1977). Üblicherweise werden sie in der Form $A : a = B : ?$ präsentiert. Dabei besteht die Aufgabe darin, den passenden vierten Teil („b“) herauszufinden, der die selbe oder eine ähnliche Beziehung zum dritten Teil („B“) aufzeigt wie der zweite Teil („a“) zum ersten („A“) (Schiano, Cooper, Glaser, & Zhang, 1989).

³ Typical Intellectual Engagement (TIE) beschreibt den Umgang eines Individuums mit Wissen als eine Persönlichkeitseigenschaft. Personen mit höherem TIE tendieren dazu, beim Wissenserwerb neugieriger und getriebener zu handeln (Goff & Ackerman, 1992; vgl. auch Chamorro-Premuzic et al., 2006; Furnham et al., 2008).

Die Bedeutung verbaler Analogien bei der psychologisch-diagnostischen Erfassung von seinem *g*-Faktor erklärte Spearman (1927) wie folgend:

To answer correctly, the subjects must first educe the relation of Here to There, and then they must apply this relation to Now, so as to educe the correlate Then. Such tests as these always show correlations with all operations known to involve *g*; they therefore involve it themselves. (S.179)

In Anlehnung an Spearman berücksichtigte auch Vernon (1950) Analogien bei der Erfassung des *g*-Faktors und zählte verbale Analogien der unteren Ebene in seiner hierarchischen Anordnung, dem Faktor „*verbal*“, zu. In ähnlicher Weise werden Analogien in Bezug auf die Primärfähigkeiten von Thurstone dem Faktor „*inductive reasoning*“ zugeschrieben (Sternberg & Gardner, 1983). Zusätzlich zu verbalen Analogien beschäftigte sich Guilford (1971) auch explizit mit figuralen Analogien. Die Fähigkeit, die zur Lösung jeglicher Analogie-Aufgaben benötigt wird, ordnete er, wie schon in Abschnitt 1.4 beschrieben, den Operationen „*Kognition*“ und „*Konvergente Produktion*“ zu.

Da die fluide Intelligenz den oben erwähnten Faktoren (Spearmans *g*, Thurstones *inductive reasoning* und Guilfords *Konvergente Produktion*) konzeptuell viel ähnlicher ist als die kristallisierte Intelligenz, sah Cattell (1987) die Fähigkeit zu erfolgreicher Bearbeitung von Analogien als einen guten Indikator für fluide Intelligenz (vgl. dazu Sternberg & Gardner, 1983). Gleichzeitig machte er auch auf die Möglichkeit des Einsatzes von figuralen Analogie-Aufgaben bei *culture-fair* Verfahren aufmerksam.

Aufgrund dieser bedeutenden Rolle bei diversen Intelligenztheorien werden Analogien, insbesondere mit verbalen Stimuli, bei vielen bekannten Intelligenztests eingesetzt, z.B. bei I-S-T 2000 R (Amthauer, Brocke, Liepmann, & Beauducel, 2001), WIT-2 (Kersting, Althoff & Jäger, 2008) und IBF (ITB & Gittler, 2011). Wie schon erwähnt, erfordern sie die Fähigkeit, den logischen Zusammenhang zwischen zwei Stimuli zu erkennen und diese Regel zum Finden einer Lösung anzuwenden, wobei in Bezug auf Lösungsstrategien individuelle Differenzen bestehen (siehe dazu z.B. Gentile, Kessler, & Gentile, 1969; Schiano et al., 1989). Folglich werden sowohl verbale als auch figurale Analogien typischerweise zur Erfassung des *logisch-schlussfolgernden Denkens* eingesetzt (Sternberg, 1977).

2.3. Definition von kristallisiertem schlussfolgerndem Denken

Das Ziel dieser Arbeit besteht darin, ein neues psychologisch-diagnostisches Verfahren zu präsentieren, das mit Hilfe von figuralen Analogie-Aufgaben *kristallisiertes schlussfolgerndes Denken* erfassen soll. Der Unterschied zwischen den im letzten Abschnitt genannten Analogie-Aufgaben, mit denen logisch-schlussfolgerndes Denken gemessen wird, und dem neu entwickelten Test *Figurale Analogien* liegt am Inhalt des verwendeten Arbeitsmaterials.

Da hier der Fokus auf figuralem Material liegt, soll im Folgenden nicht näher auf verbale Analogien eingegangen werden. Sowohl in der weiter oben beschriebenen intelligenztheoretischen Forschung als auch bei gängigen Intelligenztests wird von figuralen Analogien Gebrauch gemacht, die aus geometrischen oder abstrakten Figuren mit unterschiedlichem Komplexitätsgrad bestehen. Abbildung 2 zeigt ein Beispiel einer solchen Analogie.

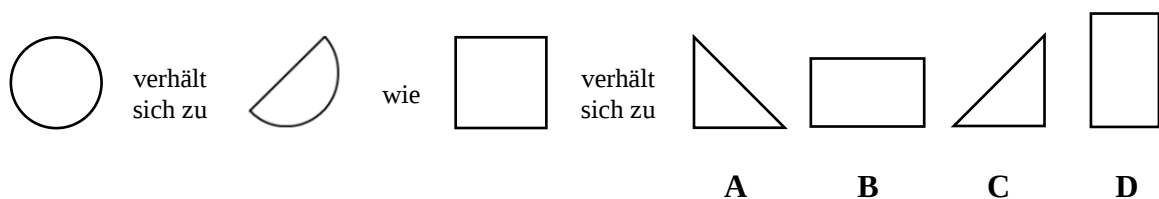


Abbildung 2: Beispiel einer einfachen figuralen Analogie-Aufgabe zu logisch-schlussfolgerndem Denken (Guilford & Hoepfner, 1976, S. 38)

Im Gegensatz dazu werden bei *Figurale Analogien* Bilder von alltäglichen Symbolen, bekannten Gegenständen sowie Lebewesen verwendet, d.h. das eingesetzte Arbeitsmaterial hat einen sinnvollen Inhalt. Bei der Bearbeitung der Aufgaben sollen Individuen zuerst erkennen, was mit diesen Bildern und Zeichnungen dargestellt wird und dann ihr Wissen über diese Objekte, also ihr Allgemeinwissen, einsetzen. Da dieses Wissen von vorangegangenen Lernprozessen beeinflusst wird, wird die Fähigkeit zur Aneignung des Allgemeinwissens sensu Cattell (1987) der *kristallisierten Intelligenz* zugeordnet. Zusätzlich zu der kristallisierten Intelligenz erfordert die Fragestellung durch die Form einer Analogie-Aufgabe die Fähigkeit zum *schlussfolgernden Denken*.

Die Bestimmung der erfassten Fähigkeit als kristallisiertes schlussfolgerndes Denken geschieht in Anlehnung an die Investment-Theorie von Cattell (1963, 1987; siehe dazu Abschnitt 2.1). Diese betrachtet g_f und g_c als sich gegenseitig beeinflussende Faktoren, wobei der Erwerb von kristallisierter Intelligenz durch den Einsatz fluider Intelligenz, d.h. des logisch-schlussfolgernden Denkens, ermöglicht wird. Deswegen soll *kristallisiertes schlussfolgerndes Denken* im Rahmen dieser Arbeit als die Fähigkeit definiert werden, in einem bestimmten Kulturkreis erworbenes Wissen durch das Erkennen von Zusammenhängen und die Anwendung der erfassten Regeln einsetzen zu können.

3. Wissensdiagnostik

Dieser Abschnitt liefert einen Überblick über die Wissensdiagnostik, genauer gesagt über die psychologisch-diagnostische Erfassung kristallisierter Intelligenz. Nach einer kurzen Vorstellung des Wissens aus psychologischer Perspektive werden die Besonderheiten der Wissensdiagnostik diskutiert, wobei der Schwerpunkt auf das Gütekriterium Validität gelegt wird. Abschließend wird das Erweiterungsmodul des I-S-T 2000 R als relevantes Beispiel für Wissenstests vorgestellt.

3.1. Wissenspsychologie

Aufgrund zunehmender Anforderungen an die Verarbeitung und Nutzung des Wissens in verschiedenen Kontexten wird den zu erwerbenden Wissensinhalten eine hohe gesellschaftliche und individuelle Relevanz zugeschrieben. Deswegen befassen sich viele Teildisziplinen der Psychologie (z.B. allgemeine, differentielle und Entwicklungspsychologie) mit unterschiedlichen Zielsetzungen mit der theoretischen Konzeption und Erfassung von Wissen (Beauducel & Süß, 2011).

Basierend auf allgemeinspsychologischen Kenntnissen wird zwischen deklarativem und prozeduralem Wissen unterschieden. *Deklaratives Wissen* bezieht sich auf mitteilbare, miteinander unverknüpfte Inhalte über Fakten, Handlungen, Verfahren oder Prozesse, während *prozedurales Wissen* durch den automatisierten und zeitlich optimierten Vollzug von Handlungen zum Vorschein kommt (Anderson, 1983, zitiert nach Beauducel & Süß, 2011; vgl. dazu z.B. Stemmler, Hagemann, Amelang, & Bartussek, 2011). Zusätzlich kann bei deklarativem Wissen zwischen *semantischem* (hauptsächlich verbal und symbolisch) und *episodischem* (über spezifische, autobiographische Erfahrungen) Wissen differenziert werden (Tulving, 1972; vgl. dazu Beauducel & Süß, 2011).

3.2. Erfassung des Wissens

Wissenstests, die im Rahmen der psychologischen Diagnostik entwickelt und angewendet werden, beschäftigen sich hauptsächlich mit deklarativem bzw. genauer mit semanti-

schem Wissen, da die Erfassung interindividueller Unterschiede in diesem Bereich sowohl mit einer hohen gesellschaftlichen Relevanz verbunden als auch in der Operationalisierung einfacher ist (Beauducel & Süß, 2011).

Ein wichtiger Unterschied zwischen Wissenstests und anderen psychologisch-diagnostischen Verfahren betrifft die Frage der Validierung. Wissenstests wird in erster Linie eine hohe *Inhaltsvalidität* zugeschrieben. In diesem Kontext wird davon ausgegangen, dass die verwendeten Fragen repräsentativ für die Gesamtheit des zu erfassenden Wissensbereichs sind (Beauducel & Süß, 2011; Stemmler et al., 2011). Für eine solche Annahme ist es von wesentlicher Bedeutung, dass die Items einen inhaltlichen Zusammenhang mit einem genau definier- bzw. begrenzten Wissensbereich aufweisen. Um dies zu gewährleisten, orientieren sich Wissenstests meist entweder an Experten und ihren Ratings des Testmaterials oder an gesellschaftlich definierten Rahmen, wie z.B. Lehrplänen von Schulen oder für einen bestimmten Beruf vorgeschriebenen Grundqualifikationen. Trotzdem gestaltet sich eine klare empirische Entscheidung über die Geltung der Inhaltsvalidität, besonders bei der Erfassung des Allgemeinwissens, als problematisch, da die Grenzen des (Allgemein-) Wissens schwierig zu definieren sind (Beauducel & Süß, 2011, S. 241 ff.).

Die Ermittlung der *Konstruktvalidität* bei Wissenstests erfolgt meist über die Untersuchung der diskriminanten und konvergenten Validität (vgl. z.B. Hossiep & Schulte, 2007; Amthauer, Brocke, Liepmann, & Beauducel, 2001). Zusätzlich wird eine positive Korrelation zwischen den Items einer bestimmten Wissensdomäne als Indikator für Konstruktvalidität gedeutet. Beauducel und Süß (2011, S. 244 ff.) machen in diesem Zusammenhang darauf aufmerksam, dass bei Wissenstests die Frage nach dem evaluierten Konstrukt nicht immer leicht beantwortet werden kann. Wie schon bei der Vorstellung der Investment-Theorie (Cattell, 1987) in Abschnitt 2.1 erwähnt wurde, wird die kristallisierte Intelligenz durch die fluide mitbestimmt. Gleichzeitig weisen diverse Untersuchungen darauf hin, dass auch andere Faktoren wie z.B. Lernstile (Furnham, Christopher, Garwood, & Martin, 2007), Persönlichkeit und Interessen (Ackerman, 1996, 2000; Ackerman & Heggestad, 1997; explizit für Allgemeinwissen: Chamorro-Premuzic, Furnham, & Ackerman, 2006; Furnham, Swami, Artech, & Chamorro-Premuzic, 2008) bei Erwerb und Messung der kristallisierten Intelligenz eine Rolle spielen. Deswegen bestehen Unsicherheiten dar-

über, welche Konstrukte in welchem Ausmaß bei einem Wissenstest genau erfasst werden (vgl. dazu Beauducel & Süß, 2011; Sternberg, 2003).

Der nächste wichtige Punkt in der Wissensdiagnostik betrifft die Gruppierung von Wissenstests. Ein einheitlicher Ordnungsrahmen für Wissenstests, wie es bei anderen kognitiven diagnostischen Verfahren üblich ist, hat sich in der Literatur bis jetzt noch nicht etabliert. Diesbezüglich schlagen Beauducel und Süß (2011, S. 253 ff.) vor, Wissenstests nach ihrem Allgemeinheitsgrad (allgemein/speziell), der Erhebungsmethode (deklarative Einzelfragen/Wissensnetzwerke/Reaktionen auf Situationen) und der Kodierung des Wissens (verbal/numerisch/figural) voneinander zu unterscheiden bzw. miteinander zu vergleichen (vgl. dazu z.B. de Jong & Ferguson-Hessler, 1996).

Als Beispiele für bekannte und in der Praxis oft eingesetzte Allgemeinwissenstests, die mittels deklarativen Einzelfragen kristallisierte Intelligenz erfassen, sind BOWIT (Hosiep & Schulte, 2007), LEWITE (Wagner-Menghin, 2004), LPS (Horn, 1983), WAIS-IV (verbale Untertests, Petermann, 2012) und WISC-IV (verbale Untertests, Petermann & Petermann, 2011) zu erwähnen (vgl. dazu Beauducel & Süß, 2011, S. 254). Im Folgenden wird der I-S-T 2000 R (Amthauer et al., 2001) genauer vorgestellt. Er beinhaltet, im Gegensatz zu den anderen genannten Tests, nicht nur verbale Stimuli und zeigt somit, zusammen mit seiner intelligenztheoretischen Grundlage, die größte Ähnlichkeit zu dem neu konstruierten Verfahren *Figurale Analogien*.

3.3. I-S-T 2000 R

Der Intelligenz-Struktur-Test 2000 R⁴ (Amthauer et al., 2001) ist ein Paper-Pencil-Verfahren, das eine Revision und Erweiterung von I-S-T 70 (Amthauer, 1970) darstellt. Er ist bei Personen zwischen 15 und 60 Jahren anwendbar und kann sowohl als Einzelvorgabe als auch in der Gruppe eingesetzt werden. Die Testentwicklung orientiert sich inhaltlich an modellübergreifend aufgewiesenen Konvergenzen in der Intelligenzstruktur-

⁴ Im Rahmen dieser Arbeit sollen die Gütekriterien von I-S-T 2000 R weitgehend unkommentiert bleiben. Für Interessierte sei auf das Manual des Verfahrens sowie die Testrezension von Schmidt-Atzert und Rauch (2008) hingewiesen.

forschung, insbesondere zwischen den Theorien von Thurstone (vgl. Abschnitt 1.3), Vernon (vgl. Abschnitt 1.2) und Cattell (vgl. Abschnitt 1.5), und strukturell an der hierarchischen Form von g_f und g_c in Anlehnung an die Investment-Theorie (vgl. Abschnitt 2.1).

Das *Grundmodul* dient zur Erfassung von fünf der sieben Primärfaktoren von Thurstone: verbale Intelligenz, numerische Intelligenz, figurale Intelligenz, Merkfähigkeit und „schlussfolgerndes Denken mit Wissensanteilen“. Dieser Wert für schlussfolgerndes Denken wird als Summenscore aus verbaler, numerischer und figuraler Intelligenz bestimmt, so dass ihm eine höhere Generalität zugesprochen wird (Amthauer et al., 2001, S. 16).

Für die vorliegende Arbeit ist besonders das *Erweiterungsmodul* von Relevanz, das aus einem einzigen Test mit einer Laufzeitbeschränkung von 40 Minuten besteht. Dieser Test umfasst 84 Wissensfragen, die ausbalanciert zu Kombinationen von sechs Themenbereichen (Geographie/Geschichte, Wirtschaft, Kunst/Kultur, Mathematik, Naturwissenschaften, Alltag) mit drei Kodierungsarten (verbal, numerisch, figural) zugeordnet werden.

Die Operationalisierung kristallisierter Intelligenz mit drei unterschiedlichen Materialtypen wird von den Testautoren als ein Entgegenwirken zu dem weit verbreiteten Einsatz von verbalem Arbeitsmaterial bei Wissenstests argumentiert (Amthauer et. al, 2001, S. 49 ff.). Kritisch ist dabei anzumerken, dass auch bei I-S-T 2000 R alle Fragen und ein Großteil der Antwortmöglichkeiten verbal formuliert sind. Dadurch spielt die verbale Kodierung bei Aufgaben zu numerisch sowie figural kodiertem Wissen ebenfalls eine bedeutende Rolle. Zum Beispiel wird bei einem Item verlangt, das richtige Symbol für ein wörtlich genanntes elektrisches Bauelement aus den präsentierten fünf Symbolen auszuwählen (Amthauer et al., 2001, Wissenstest Form B, Aufgabennummer 221). Um zu einer Lösung zu gelangen, muss die Testperson also im Besitz von sowohl des verbal (bei der Formulierung der Frage) als auch des figural (bei den Antwortmöglichkeiten) kodierten Wissens zu diesem Thema sein.

Bei der Auswertung des Erweiterungsmoduls werden folgende Intelligenzmaße geliefert: verbal kodiertes Wissen, numerisch kodiertes Wissen, figural kodiertes Wissen und Wissen als Summenwert. Mit dieser Gesamtskala Wissen „werden wesentliche Ausschnitte des Wissens erfasst, das jemand im Laufe seines Lebens in unserem Kulturkreis erworben hat“ (Amthauer et al., 2001, S. 93). Es besteht jedoch eine hohe Korrelation ($r = 0.70$)

zwischen der Gesamtskala Wissen und der Skala schlussfolgerndes Denken aus dem Grundmodul (Amthauer et al., 2001, S. 94).

Wenn die Ergebnisse des Grundmoduls miteinbezogen werden, können noch zusätzlich hierarchisch höhere Faktorwerte für „schlussfolgerndes Denken/ g_f “ und „Wissen/ g_c “ berechnet werden. Der Faktor „schlussfolgerndes Denken/ g_f “ werde im Sinne von fluider Intelligenz sensu Cattell, also als die von der Sozialisation unabhängige Fähigkeitskomponente, verstanden und komme durch Auspartialisierung der Wissensanteile zustande. Hingegen stehe der Faktor „Wissen/ g_c “ für kristallisierte Intelligenz, d.h. für Wissen ohne Einfluss des schlussfolgernden Denkens (Amthauer et al., 2001, S. 14). Diese Faktorenwerte zeigen eine geringere Korrelation ($r = 0.48$) miteinander als zwischen den Gesamtskalen Wissen und schlussfolgerndes Denken (Amthauer et al., 2001, S. 94).

Insgesamt gelingt es I-S-T 2000 R mit vier unterschiedlichen Werten (schlussfolgerndes Denken, Wissen, schlussfolgerndes Denken/ g_f und Wissen/ g_c), die zur Interpretation herangezogen werden können, sowohl eine praktische Umsetzung als auch eine Weiterentwicklung der Annahmen der Investment-Theorie. Außerdem wird die theoriegeleitete Teilung von g_f und g_c in verbale, numerische und figurale Intelligenz sowohl faktoren- als auch ähnlichkeitsstrukturanalytisch unterstützt (Amthauer et al. 2001; Beauducel, Brocke, & Liepmann, 2001). Zu bemängeln ist, dass bei dem Erweiterungsmodul Bemerkungen zur Validität in Form von Vergleichen mit anderen Tests und Außenkriterien nur für die Faktorwerte g_f und g_c angeführt werden; ähnliche Untersuchungen auf der Ebene der verbalen, numerischen und figuralen Facetten sind nicht vorhanden (vgl. Schmidt-Atzert & Rauch, 2008).

Ein weiterer Kritikpunkt betrifft den Einfluss des Lebensalters auf den Entwicklungsverlauf der kristallisierten Intelligenz. Im Manual werden die Erläuterungen der Investment-Theorie dazu (vgl. Abschnitt 2.1) nicht ausreichend und denkbare Auswirkungen dieser auf die Interpretation der Ergebnisse sowie ihre praktische Bedeutsamkeit gar nicht angesprochen. Zusätzlich lässt sich bei der Normierung des Erweiterungsmoduls nur eine relativ grobe Altersdifferenzierung mit unausgewogenen Stichprobengrößen erkennen, zum Beispiel werden Personen zwischen 14 und 25 Jahren zu einer einzigen Gruppe zusammengelegt (vgl. Schmidt-Atzert & Rauch, 2008). Bezüglich der Relevanz des Alters auf die Ergebnisse zeigte sich in einer Folgeuntersuchung, dass bei älteren Testpersonen (25- bis 50-Jährige) die numerischen Fähigkeiten mit verbal kodiertem Wissen über Naturwis-

senschaften und figural kodiertem Wissen über Mathematik hoch korrelieren (Beauducel et al., 2001).

Letztlich soll die fehlende Bezugnahme der Testautoren zur Skalierung⁵ von I-S-T 2000 R nach dem Rasch-Modell erwähnt werden. Eine Untersuchung von Bühner, Ziegler, Krumm und Schmidt-Atzert (2006) konkludiert dazu, dass sich das Verfahren großteils als „Rasch-skalierbar“ erweise und gezielte Verbesserungen zur Erfüllung der strengen Kriterien des Rasch-Modells hilfreich sein können. Jedoch bezieht sich diese Studie nur auf die Items des Grundmoduls, so dass Erkenntnisse über die Rasch-Modell-Konformität der Wissensfragen bislang noch fehlen.

⁵ Die Erfüllung des Gütekriteriums *Skalierung* ist dann anzunehmen, „wenn die laut Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen adäquat abbilden“ (Kubinger, 2009, S. 82). Zur Erklärung der Skalierung nach dem Rasch-Modell sei auf den nächsten Abschnitt verwiesen.

4. Testtheoretische Grundlage

Im Folgenden werden die Grundlagen der Item-Response-Theorie und insbesondere des dichotomen logistischen Testmodells von Georg Rasch (kurz: Rasch-Modell) vorgestellt. Dieses Modell wird im Rahmen dieser Arbeit zur Skalierung von *Figurale Analogien* herangezogen (siehe Abschnitt 10.1).

4.1. Item-Response-Theorie

In der psychologischen Testtheorie stehen sich zwei Ansätze mit unterschiedlichen Ausgangspunkten gegenüber. Während die *Klassische Testtheorie* hauptsächlich korrelationsstatistisch arbeitet und ein nicht prüfbares Modell darstellt, bedient sich die *Item-Response-Theorie*, die in der Literatur auch unter den Namen probabilistische Testtheorie und Latent-Trait-Theorie bekannt ist, als prüfbares Modell wahrscheinlichkeitstheoretischen Überlegungen (Kubinger, 2003).

Die Item-Response-Theorie veranschaulicht mit Hilfe von Wahrscheinlichkeitsfunktionen die Annahme, dass nicht beobachtbare (latente) Eigenschaften die Grundlage für die beobachtbaren (manifesten) Reaktionen von Testpersonen auf bestimmte Items bilden. In Bezug auf Leistungstests kann somit die Wahrscheinlichkeit für das Auftreten einer bestimmten Reaktion in Abhängigkeit von der zu erfassenden Fähigkeit der Person und von bestimmten Charakteristika der Aufgabe beschrieben werden (Kubinger, 2003; Kubinger, Rasch, & Yanagida, 2011).

Im Rahmen der Psychologischen Diagnostik spielt die Item-Response-Theorie eine wichtige Rolle. Dabei orientieren sich die meisten Testmodelle nach dem dichotomen logistischen Testmodell von Rasch (Kubinger, 2003).

4.2. Rasch-Modell

Die Wahrscheinlichkeitsfunktion, die dem Rasch-Modell zugrunde liegt, kann der Formel 4.1 entnommen werden. Dabei erfolgt die Darstellung der Lösungswahrscheinlichkeit

(„+“) als Funktion der Schwierigkeit des Items, modelliert mit dem Itemparameter σ , und der Fähigkeit der Testperson, modelliert mit dem Personenparameter ξ . Demzufolge wird mit der Komplementärwahrscheinlichkeit die Wahrscheinlichkeit abgebildet, dass Testperson v mit der Fähigkeit ξ_v Item i mit der Schwierigkeit σ_i nicht löst („–“) (Kubinger, 2003; Kubinger, Rasch, & Yanagida, 2011).

$$P (+|\xi_v, \sigma_i) = \frac{e^{\xi_v - \sigma_i}}{1 + e^{\xi_v - \sigma_i}} \quad (4.1)$$

Dass die Fähigkeit und die Schwierigkeit durch jeweils einen Parameter gekennzeichnet werden, widerspiegelt die dem Modell zugrunde liegende Annahme der *Eindimensionalität*. Alle Items aus einem Itempool messen folglich eine einzige und dieselbe latente Fähigkeit. Wie aus der Formel ersichtlich, wird die Lösungswahrscheinlichkeit bei steigender Fähigkeit ebenfalls größer, wenn die Itemschwierigkeit konstant bleibt. Gleichzeitig wird angenommen, dass die Anzahl gelöster Aufgaben, also der Rohwert, eine *erschöpfende Statistik* für den Fähigkeitsparameter ξ darstellt. Damit ist gemeint, dass die gesamte relevante Information über die interessierende Fähigkeit in dem Rohwert einbegriffen ist (Kubinger, 2003; Kubinger et al., 2011).

Für die Anwendung des Rasch-Modells müssen die Items entweder als „gelöst“ (+ bzw. 1) oder als „nicht gelöst“ (– bzw. 0) verrechnet werden. Folglich kann dieses Modell nur bei Vorliegen von dichotomen Daten eingesetzt werden. Die Kalibrierung der Aufgaben eines psychologischen Tests nach dem Rasch-Modell dient dazu, dass einerseits die Anzahl gelöster Aufgaben pro Person als Testkennwert herangezogen werden können, und andererseits es dabei zu keiner Wechselwirkung zwischen bestimmten Aufgaben und bestimmten Personengruppen kommt. Nur dann werden Verhaltensrelationen mit dem Rohwert adäquat abgebildet, sodass leistungsschwächere Personen tatsächlich einen niedrigeren Wert erzielen als leistungsstärkere Personen. Somit muss für jeden Test, der als Testkennwert die Anzahl gelöster Aufgaben heranzieht, *notwendigerweise* das Rasch-Modell (oder eine monotone Transformation dieses Modells) gelten, damit dieser Verrechnungsmodus als fair bezeichnet und das Gütekriterium Skalierung erfüllt werden

kann (Kubinger 2003, 2009; Kubinger et al., 2011, näheres zum Notwendigkeitsbeweis siehe z.B. Fischer, 1974; Kubinger, 2009).

Der Vorteil des Rasch-Modells zeigt sich dabei in der sogenannten „*lokalen stochastischen Unabhängigkeit*“. Das bedeutet, dass es während der Bearbeitung des Tests zu keinen Lerneffekten kommt und die erfolgreiche Bearbeitung eines früheren Items keine Voraussetzung für die Bearbeitung eines Folgeitems darstellt. Die Reaktion einer Testperson hängt also, abgesehen vom Zufall, nur von ihrer Fähigkeit und der Schwierigkeit der bearbeiteten Aufgabe ab (Kubinger 2003, 2005; Kubinger et al., 2011).

Weitere Besonderheiten des Rasch-Modells betreffen die *Stichprobenunabhängigkeit* und die *spezifische Objektivität* der Vergleiche. Bei Geltung des Rasch-Modells kann auf der einen Seite der Unterschied zwischen den Schwierigkeitsstufen der Items stichprobenunabhängig bestimmt werden. Die Wahl der Stichprobe hat also keinen Einfluss auf die Schätzungen der Itemparameter. Auf der anderen Seite können mit Hilfe des Personenparameters ξ die Fähigkeiten von zwei Testpersonen miteinander verglichen werden, unabhängig davon, welche Items aus dem selben Test diese Personen bearbeitet haben. Das bedeutet, dass spezifisch objektive Vergleiche möglich sind (Kubinger, 2003, 2009).

Diese speziellen Merkmale des Rasch-Modells bringen den Vorteil mit sich, dass die Gültigkeit des Modells geprüft werden kann. Bei Geltung des Rasch-Modells für die Items eines psychologischen Tests müssen die Schätzungen der Itemparameter gleich bleiben, wenn die Gesamtstichprobe willkürlich in zwei Subgruppen geteilt wird (Kubinger et al., 2011). Inferenzstatistisch kann das mittels des Likelihood-Ratio-Tests nach Andersen (1973) erfolgen. Mit diesem Modelltest werden die empirischen Daten darauf hin untersucht, ob sie durch die Itemparameterschätzungen in unterschiedlichen Teilstichproben besser erklärt werden können als durch denen in der Gesamtstichprobe (Kubinger, 2003; Rost, 2004). In der Regel geschieht dies durch die Teilung der Gesamtstichprobe in zwei Gruppen nach der Anzahl gelöster Aufgaben (internes Teilungskriterium). Als externe Teilungskriterien können weitere erfasste Merkmale der Testpersonen eingesetzt werden, wie z.B. Geschlecht, Alter, Bildungsniveau etc. (Kubinger, 2005; Kubinger et al., 2011).

Deskriptivstatistisch kann die Geltung des Rasch-Modells mit Hilfe eines Graphischen Modell-„Tests“ veranschaulicht werden, der auf die Eigenschaft der Stichprobenunabhängigkeit aufbaut. Bei dieser Modellkontrolle werden die Parameterschätzungen pro

Item in einem rechtwinkligen Koordinatensystem markiert, wobei die Gesamtstichprobe wiederum in Teilstichproben unterteilt wird. Die Items des untersuchten Tests entsprechen dann dem Rasch-Modell, wenn die aufgetragenen Punkte, die die Items repräsentieren, auf bzw. sehr nahe zu der durch den Ursprung gehenden 45° -Geraden liegen. So können Items einzeln beurteilt und, falls sie sich nicht als modellkonform erweisen, aus dem Itempool entfernt werden (Kubinger, 2009; Kubinger et al., 2011; Rost, 2004). Der z-Test nach Fischer und Scheiblechner (1970, zitiert nach Kubinger et al., 2011) wird auch häufig zur Beurteilung der Modellgeltung pro Item eingesetzt (genauer dazu siehe z.B. Fischer, 1974, S. 297 f.; Kubinger et al., 2011, S. 556 f.).

Wie oben erwähnt, werden Modelltests zur Kalibrierung eines Tests nach dem Rasch-Modell eingesetzt, um Eindimensionalität und Verrechnungsfairness zu gewährleisten und das Gütekriterium Skalierung zu erfüllen. Aus diesem Grund finden sie auch bei der Konstruktion von *Figurale Analogien* Anwendung, die im nächsten Kapitel näher beschrieben wird.

III. EMPIRISCHER TEIL

5. Ziele der Untersuchung und Hypothesen

Mit dieser Arbeit sollte in erster Linie die Konstruktion eines neuen Verfahrens realisiert werden, das mittels Analogie-Aufgaben kristallisiertes schlussfolgerndes Denken erfasst. Das Aufgabenmaterial sollte dabei gänzlich figural und inhaltlich sinnvoll sein. Die Details zur Entwicklung und Aufbau der Items werden im Abschnitt 6 genauer beschrieben.

Die empirische Untersuchung sollte das Ziel verfolgen, die neu erstellten Items von *Figurale Analogien* und die zusätzlich vorgegebenen Aufgaben von *AN-TOP* (siehe Abschnitt 7) auf ihre Konformität mit dem dichotomen logistischen Modell von Rasch zu überprüfen. Bei der Kalibrierung beider Tests sollten nicht Rasch-Modell-konforme Items identifiziert und gegebenenfalls ausgeschieden werden (siehe Abschnitte 10.1 - 10.3).

Weiters sollte die diskriminante Validität von *Figurale Analogien* in Bezug auf das Verfahren *AN-TOP* bestimmt werden, wofür die Korrelation der Personenparameter herangezogen wird. Dabei sollte sich die Festlegung des Grenzwerts für die maximal zu tolerierende Korrelation an theoretischen Überlegungen orientieren (Moosbrugger & Kelava, 2012) (siehe dazu Abschnitt 10.6).

Aus den genannten Zielen leiten sich folgende Hypothesen für die Untersuchung ab:

$H_0(1)$

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

$H_0(2)$

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

$H_0(3)$

Der Zusammenhang zwischen *Figurale Analogien* und *AN-TOP* ist kleiner als oder gleich dem festgelegten Grenzwert.

6. Konstruktion von *Figurale Analogien*

Im Folgenden wird das Vorgehen bei der Entwicklung von *Figurale Analogien* geschildert. Dabei werden aus urheberrechtlichen Gründen nicht alle Items des Verfahrens angeführt. Auf die Konstruktion der Items wird durch die Beschreibung des strukturellen Aufbaus der Aufgaben (siehe Abschnitt 6.1.2), die Begründung der Wahl des Antwortformats (siehe Abschnitt 6.1.3) und durch die Illustration der Erstellungsschritte (siehe Abschnitt 6.1.4) sowie der inhaltlichen Aspekte des Aufgabenmaterials (siehe Abschnitt 6.1.5) eingegangen. Anschließend werden Informationen über die eingesetzten Instruktionen angeboten (siehe Abschnitt 6.2). Die Definition und genauere Informationen bezüglich der zu erfassenden Fähigkeit können Abschnitt 2.3 entnommen werden; auf eine Wiederholung dieser wird hierbei verzichtet.

6.1. Itemkonstruktion

Wie schon im theoretischen Teil besprochen, wird mit dem neu entwickelten Verfahren, welches als Wissenstest positioniert wird (vgl. Abschnitt 3), die Erfassung des kristallisierten schlussfolgernden Denkens (vgl. Abschnitt 2.3) bezweckt. Als Arbeitsmaterial dienen dafür figurale Analogien. Dabei werden durch den Einsatz von Analogien Anteile des schlussfolgernden Denkens und durch die Verwendung von nicht abstraktem (sinnvollem) Bildmaterial Anteile der kristallisierten Intelligenz eingebaut. Dementsprechend bezieht sich das figurale Aufgabenmaterial auf das Allgemeinwissen und übt somit einen wesentlichen Einfluss auf die diversen Entscheidungen bei der Entwicklung der Items. Die Identifizierung und Handhabung der Besonderheiten des Aufgabenmaterials sowie die Vorgehensweise bei der Entwicklung der Items sollen im Folgenden genauer erläutert werden.

6.1.1 Geltungsbereich und Vorgabe

Bevor mit der Konstruktion der einzelnen Items angefangen wurde, wurde die Entscheidung getroffen, dass das Verfahren für den Einsatz bei Personen ab 18 Jahren entwickelt werden sollte. Als Antwortformat wurde das Multiple-Choice-Format gewählt (näheres

dazu siehe Abschnitt 6.1.3), infolgedessen kann der Test auch in einem Gruppensetting administriert werden.

Bei *Figurale Analogien* handelt es sich um ein Paper-Pencil-Verfahren, d.h. die Aufgaben werden den Testpersonen in gedruckter Form mit dem vorbereiteten Testheft vorgelegt. Die Vorgabe bei der empirischen Untersuchung erfolgte nach einer kurzen, schriftlichen Instruktion mit einer Zeitbeschränkung von 18 Minuten (siehe auch Abschnitt 6.2).

6.1.2 Aufbau der Aufgaben

Das Aufgabenmaterial besteht aus jeweils acht Bildern pro Item, die in zwei Zeilen angeordnet sind. Die ersten drei Bilder werden für die Formulierung der Frage und die anderen fünf bei der Veranschaulichung der Antwortmöglichkeiten angewendet. Abbildung 3 zeigt die Grundstruktur, nach der alle Analogien des Verfahrens, folglich alle Fragenstellungen, aufgebaut sind. Auf die Struktur der Antwortmöglichkeiten wird im nächsten Abschnitt genauer eingegangen.

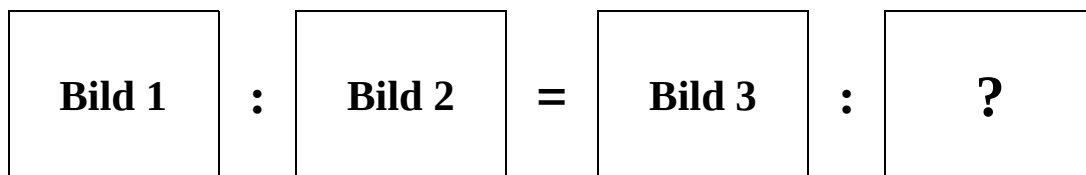


Abbildung 3: Grundstruktur der Fragestellung bei *Figurale Analogien*

Die Frage wird, wie es für Analogie-Aufgaben im Rahmen der Psychologischen Diagnostik üblich ist, in Form einer Gleichung formuliert. Die ersten zwei Bilder der Frage werden mit einem Doppelpunkt, der die Verbindung zwischen den Bildern symbolisieren soll, miteinander verknüpft. Als der erste Teil der Analogie-Aufgabe dienen sie zur Präsentation der Beziehung, die von der Testperson für die weitere Bearbeitung des Items herausgefunden werden soll. Das dritte Bild wird mittels Doppelpunkt mit einem Fragezeichen verbunden, wodurch die tatsächliche Fragestellung stattfindet. Diese Struktur soll die Testpersonen darauf hinweisen, dass zwischen dem dritten Bild und dem Bild, das als richtige Antwort gelten soll, eine ähnliche Beziehung besteht wie zwischen den ersten zwei Bildern.

6.1.3 Antwortformat

Unterhalb der Fragestellung werden die weiteren fünf Bilder positioniert, die mit den Buchstaben A bis E gekennzeichnet sind. Bei der Bearbeitung soll jene Antwortmöglichkeit, die die Analogie-Aufgabe am besten vervollständig, ausfindig gemacht und durch die Markierung des zugehörigen Buchstaben ausgewählt werden (siehe Abbildung 4).

Bild 4	Bild 5	Bild 6	Bild 7	Bild 8
A	B	C	D	E

Abbildung 4: Präsentationsform der Antwortmöglichkeiten bei *Figurale Analogien*

Wie daraus ersichtlich, wird bei *Figurale Analogien* als Antwortformat ein Multiple-Choice-Format mit fünf Antwortmöglichkeiten eingesetzt. Von diesen Antwortmöglichkeiten stellt eine einzige die korrekte Antwort dar, wobei die anderen vier als Distraktoren, d.h. als Ablenkung der Testperson von der Lösung, dienen („1 aus 5“). Das Item wird nur dann als gelöst gewertet, wenn das richtige Bild und keine Distraktoren markiert werden.

Die Wahl für das Multiple-Choice-Format statt eines offenen Antwortformats, bei dem die Testperson die Lösung selber produzieren soll, kann auf die figurale Natur des Aufgabenmaterials zurückgeführt werden. Da dieser Test nicht ein Maß für die zeichnerischen oder kreativen Fähigkeiten der Testpersonen darstellen soll, wäre die Anwendung eines offenen Antwortformats unangebracht. Gleichzeitig gewährleistet der Einsatz des Multiple-Choice-Formats sowohl Ökonomie bei der Durchführung sowie bei der Auswertung als auch die Geltung des Gütekriteriums Objektivität, genauer der Verrechnungssicherheit. Da es durch das Antwortformat genau vorgegeben wird, wie die Reaktionen der Testpersonen zu verrechnen sind, liefert die schnelle Auswertung unabhängig von der auswertenden Person immer dasselbe Ergebnis (vgl. dazu Lienert & Raatz, 1998; Kubinger, 2009, S. 43 f.). Zusätzlich ermöglicht dieses Antwortformat die Vorgabe des Verfahrens in der Gruppe.

Die Vorgabe von fünf Antwortmöglichkeiten anstelle von mehr lässt sich wiederum durch die Besonderheit des Aufgabenmaterials begründen. Auf der einen Seite wurde darauf geachtet, die Testhefte so zu formatieren, dass alle Antwortmöglichkeiten innerhalb einer Zeile präsentiert werden können, um die Aufmerksamkeit der Testpersonen nicht zu stark zu belasten. Gleichzeitig sollte für die Bilder eine bestimmte Größe behalten werden, um ihr Erkennungspotenzial zu maximieren. Auf der anderen Seite gestaltet sich das Auffinden der Analogie unpassenden Bildern, die plausibel genug erscheinen, um von den Testpersonen nicht zu schnell ausgeschlossen zu werden, und somit ihre Aufgabe als Distraktoren erfüllen können, bei einem Wissenstest besonders schwierig. Das liegt daran, dass abstrakte bzw. selbst konstruierte Bilder nicht zum Einsatz kommen, so dass nur eine begrenzte Anzahl von themenrelevanten Bildern als Distraktoren zur Verfügung steht. Als Resultat dieser Überlegungen zeigte sich, dass mit fünf Antwortmöglichkeiten sowohl eine optimale Größe für die Bilder als auch eine realisierbare Anzahl passender Distraktoren erreicht werden können.

Der Einsatz von einer einzigen richtigen Antwortmöglichkeit aus den fünf gegebenen (statt z.B. „2 aus 5“ oder „x aus 5“) ist durch die Form der Fragestellung zu erklären. Eine Analogie setzt definitionsgemäß den Vergleich der Beziehung zwischen zwei vorgegebenen Objekten mit der Beziehung zwischen einem vorgegebenen und einem gesuchten Objekt voraus. Deswegen besteht bei *Figurale Analogien* die als richtig zu verrechnende Reaktion der Testperson nur aus einem einzigen, die Analogie genau vervollständigenden Bild („1 aus 5“).

Obwohl das Multiple-Choice-Format „1 aus 5“ aus den oben dargelegten Gründen die beste Alternative für den entwickelten Test darstellt, bringt es auch einen nicht unwichtigen Nachteil mit sich, nämlich das Problem des Rateeffekts (vgl. Kubinger, 2009, S. 130ff.). Bei *Figurale Analogien* muss die Testperson die richtige Antwort aus fünf Antwortmöglichkeiten auswählen, so dass die a-priori Ratewahrscheinlichkeit, d.h. die Wahrscheinlichkeit durch die zufällige Auswahl einer Antwortmöglichkeit die Aufgabe zu lösen, mit $\frac{1}{5} = 0,20$ relativ hoch ausfällt.

Zur Minimierung von Rateeffekten ist bei der Auswahl von Distraktoren und bei der Positionierung der Lösung innerhalb der Antwortmöglichkeiten Vorsicht geboten. Deswegen wurde bei der Konstruktion der Items von *Figurale Analogien* die Lösung etwa gleich häufig auf die Positionen A bis E aufgeteilt und darauf geachtet, dass die Positio-

nierung in Bezug auf den gesamten Test nach keinem bestimmten Antwortmuster (z.B. A-A-B-B oder A-B-C-A-B-C) erfolgt. Tabelle 1 gibt einen Überblick darüber, wie die Lösung bei jedem Item positioniert wurde und wie häufig die Lösung eine der fünf Stellen der Antwortmöglichkeiten einnimmt. Kombiniert mit der Auswahl von möglichst plausiblen Distraktoren soll dieses Vorgehen auf formaler Ebene dabei helfen, die Ratewahrscheinlichkeit über alle Items konstant und so gering wie möglich zu halten.

Tabelle 1: Jeweilige Position der Lösung bei jedem Item
innerhalb der fünf Antwortmöglichkeiten

	Itemnummer	Häufigkeit der Position
1. Stelle (A)	8, 14, 16, 20, 23	5
2. Stelle (B)	2, 5, 9, 12, 13, 26	6
3. Stelle (C)	1, 4, 7, 10, 22	5
4. Stelle (D)	6, 15, 19, 25, 27	5
5. Stelle (E)	3, 11, 17, 18, 21, 24	6

6.1.4 Erstellung des Aufgabenmaterials

Das Testheft von *Figurale Analogien* umfasst 27 Items auf zehn A4-Seiten, von denen die erste für die Vorgabe der Instruktion (siehe Abschnitt 6.2) reserviert ist. Pro Seite werden somit nur drei Items präsentiert, was auf die für hohe Erkennbarkeit optimierte Größe der Bilder zurückzuführen ist.

Insgesamt besteht das Aufgabenmaterial, inklusive des Beispielitems, aus 224 Bildern. Sechs Items (zwei Seiten des Testhefts) beinhalten farbige Bilder, die restlichen 21 Aufgaben sind schwarz-weiß. Abbildung 5 gibt die Beispielaufgabe (B) wieder, die wegen ihrer Einfachheit für die Instruktion ausgewählt wurde.

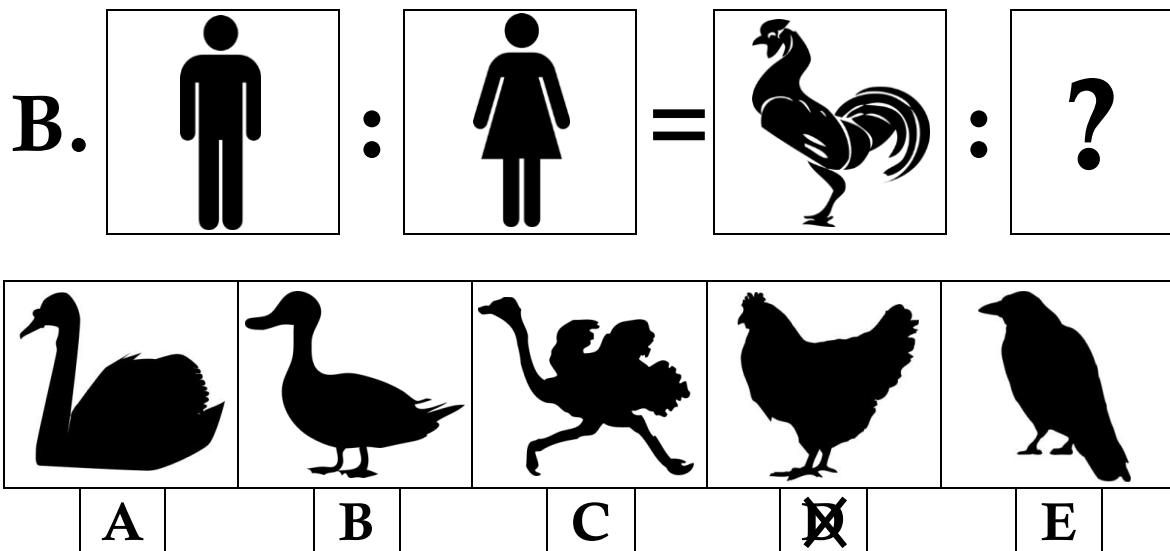


Abbildung 5: Instruktionsitem bei *Figurale Analogien*

Das Aufgabenmaterial ist, abgesehen von den verbalen Anteilen der Instruktion, ausschließlich figural. Wie schon öfters erwähnt, besteht es aus Bildern, genauer aber aus Abbildungen, Zeichnungen, Piktogrammen⁶ bzw. Fotos. Inhaltlich beziehen sich die Abbildungen auf geographische bzw. symbolische Repräsentationen von Ländern und die Zeichnungen auf bildliche, zum Teil stark vereinfachte Darstellungen von Lebewesen sowie diversen, vorwiegend alltäglichen Gegenständen. Mit Piktogrammen werden allgemein gebräuchliche Informationssymbole bzw. Orientierungshilfen, mit Fotos bekannte Persönlichkeiten abgebildet.

Die Bilder wurden teils dem Internet entnommen, teils sind sie eigens für das Verfahren erstellt worden. Tabelle 2 verschafft einen Überblick darüber, wie viele Bilder aus welchen Quellen stammen. Genauere Angaben über die Herkunft der einzelnen Bilder werden hier nicht angeführt, um den exakten Inhalt der Items nicht bekannt zu geben. Etliche Retuschen der Bilder wie z.B. Farbwechsel, Größenanpassung, Schärfeveränderung, Entfernung von Hintergrundelementen etc. wurden mit Hilfe eines unentgeltlichen online

⁶ Ein Piktogramm ist eine grafische, stilisierte Darstellung, die eine bestimmte Information vermittelt, z.B. das international bekannte Rauchverbotszeichen oder die Symbole für Herren und Damen, die beim Instruktionsitem vorkommen.

Foto-Editors⁷ durchgeführt. Die finalen Formen der Bilder wurden dann entweder im PNG- oder JPEG-Grafikformat komprimiert und anschließend in Form von Analogien im Testheft organisiert.

Tabelle 2: Häufigkeiten der für die Erstellung des Arbeitsmaterials verwendeten Quellen

Quelle	Anzahl der Bilder
http://www.clker.com/	81
http://commons.wikimedia.org	75
http://openclipart.org/	31
für den Test erstellt (mittels Microsoft Office und/oder Pixlr)	29
sonstige Webseiten	8

Alle aus dem Internet übernommenen Bilder stehen in der sogenannten *Public Domain*, d.h. sie sind urheberrechtlich nicht geschützt und somit gemeinfrei. Die Anordnung der Bilder als Analogie-Aufgaben bei dem Verfahren *Figurale Analogien* unterliegt jedoch dem Urheberrecht, weswegen im Rahmen dieser Arbeit auf eine vollständige Wiedergabe der entwickelten Items verzichtet wird⁸.

6.1.5 Themenbereiche und Schwierigkeitsstufen

Inhaltlich bestehen die Aufgaben von *Figurale Analogien* aus Wissensfragen, zu deren Lösung das Vorhandensein des erworbenen Wissens über mehrere unterschiedliche Themenbereiche benötigt wird. Während der Ideensammlung in der ersten Phase der Itemkonstruktion wurde ersichtlich, dass sich Wissensinhalte aus manchen Themengebieten

⁷ <http://pixlr.com/editor/>

⁸ Bei Interesse an der Ansicht der Items ist der Betreuer dieser Arbeit Univ. - Prof. Dr. Mag. Kubinger zu kontaktieren.

viel leichter figural darstellen lassen als andere. Um zu verhindern, dass bei dem entwickelten Verfahren die Erfassung der Wissensaneignung über bestimmte Themen über- bzw. unterrepräsentiert wird, wurden drei Themenbereiche festgelegt, zu denen jeweils die gleiche Anzahl von Items zugeordnet werden sollte. Gleichzeitig wurden bei der Formulierung der Items Fragen vermieden, die nur von Personen mit sehr spezifischen Interessen oder besonderen Lerngelegenheiten richtig beantwortet werden könnten. Die dadurch entstandenen Themenbereiche beinhalten jeweils neun Aufgaben und umfassen folgende Wissensgebiete:

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

Diese Unterteilung lässt erkennen, dass die eingesetzten Wissensaufgaben einen hohen Allgemeinheitsgrad aufweisen. Da der Test *Figurale Analogien* für den Einsatz bei Personen im Erwachsenenalter konstruiert wurde, orientieren sich die Wissensfragen nicht nur an schulischen Curricula, sondern beziehen sich auch auf Wissen über andere Aspekte des alltäglichen Lebens, das nicht zwingend in einem institutionellen Bildungssetting erworben wird. Dabei soll wiederholt angemerkt werden, dass das zu erfassende Wissen in hohem Maße kulturgebunden ist, so dass die Items für Personen außerhalb des deutschen Sprachraums ein anderes Schwierigkeitsniveau aufweisen bzw. sogar unlösbar sein könnten.

In Bezug auf die Verteilung der Items zu den unterschiedlichen Schwierigkeitsstufen wurde eine ähnliche Ausbalancierung vorgenommen. Es sollte nämlich so weit wie möglich vermieden werden, dass z.B. alle Fragen eines Themenbereichs besonders leicht gelöst werden können, während die schwierigeren Items aus einem anderen Themenbereich stammen. Es ergibt sich aus der inhaltlichen Beschaffenheit des Aufgabenmaterials, dass die jeweiligen Schwierigkeitsniveaus der Aufgaben bei einem Wissenstest während der Itemkonstruktion nicht genau bestimmt bzw. miteinander verglichen, sondern nur von den TestautorInnen geschätzt werden können. Die Aufstellung von Regeln zur Definition von Schwierigkeitsstufen, wie es bei vielen Leistungstests mit abstraktem Itemmaterial üblich

ist, ist hier nicht durchführbar. Dies trägt auch dazu bei, dass der Itempool nicht beliebig durch ein a-priori definiertes Konstruktionsrational erweitert werden kann.

Entsprechend dieser Überlegungen wurde bei der Entwicklung der Wissensfragen versucht, abzuschätzen, als wie schwierig sich die Lösung jedes Items für eine Testperson mit durchschnittlicher Fähigkeit darstellen würde⁹. Die Items wurden anschließend so konstruiert, dass jedem Themenbereich eine gleich große Anzahl von Wissensfragen aus angenommenen Schwierigkeitsstufen zugeordnet werden können.

Über die Kombinationen aus Themengebieten und Schwierigkeitsniveaus sowie die Anzahl der Fragen bei jeder Kombination werden in Tabelle 3 berichtet, die bei der Itemkonstruktion als Orientierungshilfe diente. Um die Motivation der Testpersonen durch abwechslungsreiches Material zu fördern, wurden diese Kombinationen randomisiert über das Testheft verteilt. Weder die Themenbereiche noch die Schwierigkeitsstufen haben somit einen Einfluss auf die Reihenfolge der Aufgaben.

Tabelle 3: Anzahl der Items pro Themenbereich und angenommene Schwierigkeitsstufe

Themen	Schwierigkeitsstufe					Summe
	leicht	leicht - mittelschwierig	mittelschwierig	mittelschwierig - schwierig	schwierig	
<i>Menschen</i>	2	1	3	1	2	9
<i>Naturwissens.</i>	2	1	3	1	2	9
<i>Alltag</i>	2	1	3	1	2	9
Summe	6	3	9	3	6	27

⁹ Die Itemschwierigkeitsparameter, die nach der Datenerhebung berechnet wurden, sind Abschnitten 10.2 und 10.4 zu entnehmen

6.2. Instruktion und Bearbeitungszeit

Die Informationen, die von den Testpersonen zur Bearbeitung von *Figurale Analogien* benötigt werden, befinden sich auf der ersten Seite des Testhefts. Bei der empirischen Untersuchung wurde für das Testheft die Überschrift „Allgemeinwissen“ gewählt, um einerseits den inhaltlichen Unterschied des Verfahrens im Vergleich zu dem davor zu bearbeitenden Test *Gleichungen* zu unterstreichen, und andererseits die Testpersonen auf den Inhalt der folgenden Items kognitiv vorzubereiten. Bei der Instruktion wurde gezielt eine leicht verständliche sprachliche Ausdrucksweise eingesetzt, um die Notwendigkeit ergänzender Erklärungen durch den/die TestleiterIn zu minimieren. Der Erfolg dieser Vorgehensweise zeigte sich bei der Datenerhebung. Die Testpersonen der Stichprobe stellten keine zusätzlichen Fragen bezüglich der Instruktion.

Die Instruktion beginnt mit einer kurzen Erklärung über den Aufbau der Items, gefolgt von dem Beispielitem (siehe Abbildung 5), bei dem die richtige Antwort schon markiert ist. Anschließend wird der Grund für die Wahl der markierten Antwortmöglichkeit angegeben und auf die Suche nach einer einzigen richtigen Antwort sowie auf die begrenzte Bearbeitungszeit hingewiesen. Die Instruktion, die bei der Datenerhebung verwendet wurde, wird in ihrer Gesamtheit im Anhang unter A wiedergegeben.

Ein vollständiger Verzicht auf die sprachlichen Anteile der Instruktion wäre auch denkbar. Um die Machbarkeit dieser Möglichkeit zu überprüfen, wurden vor der Datenerhebung fünf Personen gebeten, sich zuerst nur das Instruktionsitem anzuschauen und dann die Items zu bearbeiten. Diese Personen berichteten über keine Verständnisschwierigkeiten in Bezug auf die Methode der Bearbeitung des Verfahrens. Im Rahmen dieser Arbeit wurde diese Idee jedoch nicht realisiert, da bei der empirischen Untersuchung die Überprüfung der Lösbarkeit der Items sowie die strukturelle Ähnlichkeit der Instruktion der angewendeten Verfahren (*Figurale Analogien*, *AN-TOP*, *Gleichungen*) priorisiert wurde.

Bei der Datenerhebung wurde die Bearbeitungszeit für alle drei vorgegebenen Verfahren auf je 18 Minuten beschränkt. Obwohl *Figurale Analogien* als reiner Power-Test, d.h. ohne Begrenzung der Bearbeitungszeit, konzipiert wurde, zwingt die Vorgabe in einer Gruppensituation zur Festlegung einer Zeitbeschränkung. Zur Standardisierung des Testablaufs wurde die Zeitbeschränkung, die für die anderen beiden Verfahren bestimmt wurde, auch für *Figurale Analogien* übernommen.

In diesem Zusammenhang wurden die Testleiterinnen auf die Möglichkeit hingewiesen, dass die Testung vor Ablauf der vorgeschriebenen Zeit beendet werden könne, falls alle Testpersonen einer Gruppe mit der Bearbeitung der Aufgaben fertig werden sollten. Diese Situation ist für das Verfahren *Figurale Analogien* in allen Testpersonengruppen eingetreten, so dass alle Testungen nach sieben bis 13 Minuten abgeschlossen werden konnten (durchschnittliche Bearbeitungszeit: zehn Minuten, genaueres siehe im Anhang unter B). Es konnte auch beobachtet werden, dass manche Testpersonen, die besonders schnell arbeiteten, schon nach vier Minuten fertig wurden (mehr dazu siehe Kapitel IV).

7. Alpha Numeric Topologies (AN-TOP)

Alpha Numeric Topologies (Kubinger & Heuberger, in Vorb.) (Abkürzung: *AN-TOP*) dient zur Erfassung des logisch-schlussfolgernden Denkens mit Hilfe von figuralem Arbeitsmaterial. Dabei werden die Testpersonen mit der Aufgabe konfrontiert, Gesetzmäßigkeiten zu erkennen und entsprechend anzuwenden. Intelligenztheoretisch ist dieses Verfahren dem Modell der fluiden Intelligenz von Cattell (vgl. Abschnitt 1.5) zuzuordnen.

Im Rahmen dieser Arbeit wird die Vorgabe von *AN-TOP* zur Bestimmung der diskriminanten Validität herangezogen. Auf der einen Seite besteht das figurale Arbeitsmaterial von *AN-TOP* aus Zahlen, Buchstaben und geometrischen Figuren und ist somit, im Gegensatz zu dem von *Figurale Analogien*, inhaltlich abstrakt. Auf der anderen Seite ist die erfasste logisch-schlussfolgernde Fähigkeit unabhängig von früheren Lernerfahrungen, so dass sie im Kontrast zu der kristallisierten Fähigkeitskomponente bei *Figurale Analogien* steht.

Tabelle 4: Ausschnitt des Testdesigns von *AN-TOP*

	Itemnummer													
	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b
Testheft 1	1			2	3		4		5			6	7	
Testheft 2		1		2		3		4	5		6			7
Testheft 3	1		2			3	4			5	6		7	

Insgesamt umfasst *AN-TOP* 56 Items, die über drei Testhefte verteilt wurden. Die Hälfte der Aufgaben stellen Parallelitems dar und kommen jeweils bei zwei unterschiedlichen Testheften vor. Ein Ausschnitt dieses verlinkten Testdesigns wird in Tabelle 4 vorgestellt. Dabei gibt die Itemnummer (1a, 1b, 2a etc.) Auskunft über die Nummerierung einer Aufgabe in Bezug zu allen verfügbaren Items. Die anderen Zahlen bilden die Positionen der

Items im jeweiligen Testheft ab, z.B. ist Item 1a die erste Aufgabe sowohl im ersten als auch im dritten Testheft, wobei Item 1b nur im zweiten Testheft erscheint. Das gesamte Testdesign ist dem Anhang unter C zu entnehmen.

Jedes Testheft beinhaltet nur 28 Aufgaben, für deren Bearbeitung 18 Minuten zur Verfügung gestellt werden. Bei jedem Item werden den Testpersonen drei Kästchen mit einer bestimmten Anzahl von Symbolen präsentiert. Ihre Aufgabe ist, die Gesetzmäßigkeit, die sich zwischen den Symbolen der ersten zwei Kästchen befindet, zu identifizieren und dementsprechend das fehlende Symbol beim dritten Kästchen einzutragen. Eines der Instruktionsitems wird als Beispiel in Abbildung 6 wiedergegeben: da die Beziehung zwischen den ersten zwei Rahmen durch den Kreis definiert wird, soll beim dritten Rahmen als Lösung das Symbol „○“ eingetragen werden.

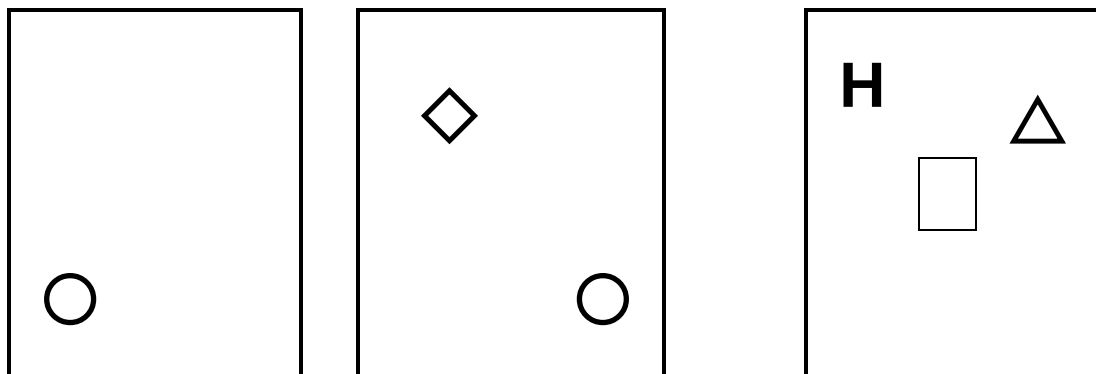


Abbildung 6: Itembeispiel von *AN-TOP* (Kubinger & Heuberger, in Vorb.)

8. Planung und Durchführung der Untersuchung

Zu Beginn der Stichprobenakquirierung wurde entschieden, dass für die Kalibrierung der Items nach dem Rasch-Modell eine Stichprobengröße von 250 bis 300 Testpersonen angestrebt werden sollte. Zusätzlich sollten die Daten bei Personen ab 18 Jahren sowie in Gruppentestungen erhoben und auf die Verwendung einer rein studentischen Stichprobe verzichtet werden. Diesen Überlegungen zufolge wurde versucht, Testungen in den Sicherheitsakademien der österreichischen Polizei zu verwirklichen, insbesondere weil bei diesen Bildungsinstitutionen für die Aufnahme ein Mindestalter von 18 Jahren vorausgesetzt wird.

Um die intendierte Stichprobengröße zu erreichen, wurden Gespräche mit den Verantwortlichen der Sicherheitsakademien in Traiskirchen sowie in Ybbs (beide in Niederösterreich) geführt und entsprechende Anträge an das Bundesministerium für Inneres gestellt (siehe im Anhang unter D). Nach Erhalt offizieller Genehmigungen wurden für die Testungen insgesamt fünf Termine vereinbart. Dies geschah nach Absprache mit dem jeweiligen Schulkommandanten, da der Ablauf des regulären Unterrichts so wenig wie möglich beeinträchtigt werden sollte. Mit Hilfe eines Informationsschreibens wurden die Testpersonen sowie zuständige LehrerInnen schon im Voraus über die geplanten Testungen und ihr Ablauf in Kenntnis gesetzt (siehe im Anhang unter E).

Als Testleiterinnen dienten neben den beiden Diplomandinnen (Gamsjäger [zuständig für *Gleichungen*] und Ünal) zwei weitere Studentinnen aus dem Arbeitsbereich der Psychologischen Diagnostik der Universität Wien, die zuvor über den allgemeinen Ablauf und ihre Aufgaben während den Testungen geschult wurden. Die Testungen wurden während den regulären Unterrichtszeiten in den jeweiligen Räumlichkeiten der Klassen durchgeführt. Pro Klasse waren durchschnittlich 24 Personen anwesend. Die Datenerhebungen in Traiskirchen fanden am 17.04. und 19.04.2012 statt, wobei jeweils vier Klassen parallel getestet wurden. Die Testungen in Ybbs erfolgten an drei weiteren Terminen (20.04., 27.04. und 09.05.2012) für insgesamt sechs Klassen. Obwohl für die Testungen jeweils zwei Unterrichtseinheiten mit insgesamt 100 Minuten verfügbar waren, wurde im Durchschnitt weniger als eine Stunde in Anspruch genommen. Die Testleiterinnen hatten nämlich die Möglichkeit, zum nächsten Verfahren zu wechseln bzw. die Testung vorzeitig zu

beenden, sobald beobachtet werden konnte, dass alle Testpersonen mit der Bearbeitung fertig waren.

Bei den Testungen wurden die TeilnehmerInnen zuerst ausführlich über den Testungsablauf informiert. Sie wurden darauf hingewiesen, dass ihre Daten anonymisiert und nur für wissenschaftliche Zwecke genutzt werden würden und dass ihre Leistungen bei der Testung ihre schulischen Bewertungen in der Sicherheitsakademie nicht beeinflussen würden. Um die Anonymität zu gewährleisten, wurde jeder Testperson ein fünfstelliger Probandencode erteilt, bei dem die ersten zwei Zeichen den Testungsort kennzeichneten. Danach bekam jede Testperson ein Blatt zur Erhebung der demographischen Daten und jeweils ein Testheft für *Gleichungen*, *Figurale Analogien* und *AN-TOP*, welche auch in dieser Reihenfolge bearbeitet werden sollten. Der Test *Gleichungen* wurde von Gamsjäger (2012) im Rahmen ihrer Diplomarbeit konstruiert und soll zur Erfassung von logisch-schlussfolgerndem Denken mit Hilfe von numerischem Aufgabenmaterial dienen. Bei den Verfahren *Gleichungen* und *AN-TOP* waren jeweils drei unterschiedliche Testformen verfügbar, von denen jede Testperson jeweils nur eine erhielt. Dabei wurde darauf geachtet, dass die neun Kombinationen von Testheften ungefähr gleich oft eingesetzt wurden. Bei *Figurale Analogien* bearbeiteten alle Testpersonen dieselbe Testform. Die Instruktionen für die Verfahren lasen die Testpersonen selbstständig; bei Nachfragen standen die Testleiterinnen zur Verfügung. Für die Bearbeitung der Verfahren waren jeweils 18 Minuten geplant, die für *Figurale Analogien* nie und für *AN-TOP* häufig, jedoch nicht immer zur Gänze gebraucht wurden.

Die Testleiterinnen wurden zusätzlich noch gebeten, sich Notizen über die benötigten Bearbeitungszeiten und das beobachtete Verhalten der TeilnehmerInnen während den Testungen zu machen. Dabei wurde dokumentiert, dass sowohl durch die starke Variation der Bearbeitungsgeschwindigkeiten der Testpersonen als auch durch die zu kleinen Testungsräume die Versuche der Testpersonen, bei ihren SitznachbarInnen abzuschauen bzw. die Aufgaben gemeinsam zu bearbeiten, zum Teil nicht verhindert werden konnten. Zusätzlich berichteten die Testleiterinnen über Unruhen, die in den Klassenräumen entstanden, wenn der Großteil der Testpersonen mit der Bearbeitung der Items fertig wurde und auf die „langsameren“ TeilnehmerInnen warten musste. Diese Verhaltensweisen konnten während der gesamten Testung, aber insbesondere bei der Bearbeitung von *Figurale Analogien* beobachtet werden (mehr dazu im Kapitel IV).

9. Stichprobe

Die Stichprobe setzt sich insgesamt aus 331 Auszubildenden der Sicherheitsakademien zusammen. Davon bestehen 77,9 % aus männlichen und 22,1 % aus weiblichen Testpersonen. Wie aus der Tabelle 5 ersichtlich, variiert das Alter der Testpersonen zwischen 18 und 32 Jahren ($Md_{Alter} = 23$ Jahre). Dabei ist davon auszugehen, dass dies eine repräsentative Stichprobe für ÖsterreicherInnen im jungen Erwachsenenalter darstellt, die eine Ausbildung bei der Polizei anstreben.

Tabelle 5: Die Verteilung der Stichprobe in Bezug auf das Alter in Jahren ($n = 331$)

Alter in Jahren	absolute Häufigkeit	Prozent
18	2	0,6
19	15	4,5
20	52	15,7
21	57	17,2
22	39	11,8
23	32	9,7
24	39	11,8
25	19	5,7
26	18	5,4
27	23	6,9
28	12	3,6
29	7	2,1
30	12	3,6
31	3	0,9
32	1	0,3

Nahezu alle TeilnehmerInnen (96,4 %) gaben Deutsch als ihre Muttersprache an, wobei hier angemerkt werden sollte, dass die österreichische Staatsbürgerschaft als eine der Voraussetzungen für die Aufnahme in die Sicherheitsakademien gilt. In Bezug auf die höchste abgeschlossene Ausbildung hatten 40,8 % eine berufsbildende höhere Schule

beendet. Nähere Details zu den Bildungsniveaus der Testpersonen können Tabelle 6 entnommen werden.

Tabelle 6: Die Verteilung der Stichprobe in Bezug auf die höchste abgeschlossene Ausbildung ($n = 331$)

höchste abgeschlossene Ausbildung	absolute Häufigkeit	Prozent
Pflichtschule	9	2,7
Lehre	88	26,6
Allgemeinbildende Höhere Schule (AHS)	94	28,4
Berufsbildende Höhere Schule (z.B. BHS, HTL, HAK)	135	40,8
Hochschule (Universität, FH, PH)	5	1,5

Als Wohnort wurde am häufigsten Niederösterreich (72,8 %) genannt, was darauf zurückgeführt werden kann, dass die Testungen in diesem Bundesland stattfanden. 15,4 % wohnten in Wien, die anderen österreichischen Bundesländer waren noch seltener vertreten (Steiermark: 3,6 %, Burgenland: 2,7 %, Oberösterreich: 1,2 %, Kärnten und Salzburg: jeweils 0,6 %). Zehn TeilnehmerInnen (3 %) machten keine Angaben bezüglich ihres Wohnorts.

10. Darstellung der Ergebnisse

Im Folgenden beschäftigen sich die ersten drei Abschnitte mit den Analysen zur Geltung des Rasch-Modells in Bezug auf die Items von *Figurale Analogien* und *AN-TOP*. Genauere theoretische Erklärungen zum Rasch-Modell befinden sich in Abschnitt 4.2 und werden hier nicht wieder angeführt. Danach werden für beide Verfahren deskriptive Ergebnisse zu den Rohwerten, Lösungshäufigkeiten sowie Itemschwierigkeitsparametern präsentiert. Im letzten Abschnitt wird das Ergebnis der Korrelation zwischen *Figurale Analogien* und *AN-TOP* besprochen.

Für die Rasch-Modell Analysen und die Berechnung der Personenparameter wurde die Open-Source-Statistik-Software R (3.0.1) mit dem zusätzlichen Programmpaket *eRm* (extended Rasch modeling)¹⁰, und für die weiteren Analysen das Programm IBM SPSS Statistics (Version 21) eingesetzt.

10.1. Beschreibung des Vorgangs bei den Rasch-Modell Analysen

Zur Prüfung der Geltung des Rasch-Modells bei den Items von *Figurale Analogien* und *AN-TOP* werden folgende Teilungskriterien verwendet:

- **Anzahl gelöster Aufgaben (Score):** Rohwert $\leq$ Median vs. Rohwert $>$ Median (internes Teilungskriterium)
- **Alter:** 18- bis 23-Jährige vs. 24- bis 32-Jährige
- **Bildungsniveau:** niedrig (Pflichtschule oder Lehre) vs. hoch (AHS, BHS oder Hochschule)

Die Teilungskriterien „Bildungsniveau“ und „Alter“ spielen besonders für die Skalierung von *Figurale Analogien* eine wichtige Rolle, da theoretisch davon auszugehen ist, dass die erfasste Fähigkeit sowohl von den vorangegangenen Lernerfahrungen als auch vom Lebensalter beeinflusst wird (vgl. Abschnitt 2.1). Mit diesem Verfahren soll eindimensio-

¹⁰ Genauerer zu *eRm* kann dem Manual von Poinstingl, Mair und Hatzinger (2007) entnommen werden.

nal die Fähigkeit zu kristallisiertem schlussfolgerndem Denken und somit weder das Alter noch das Bildungsniveau der Testpersonen erfasst werden.

Um zu prüfen, ob die Schätzungen der Itemparameter in den jeweiligen zwei Teilstichproben statistisch gleich sind, wird der Likelihood-Ratio-Test nach Andersen (1973) angewendet. Dabei wird das akzeptierte untersuchungsbezogene Risiko 1. Art auf $\alpha = 0,01$ festgesetzt, da mehrere Signifikanztests durchgeführt werden (vgl. dazu z.B. Kubinger, Rasch & Yanagida, 2011, S. 270 f.).

Es ist von Modellabweichungen auszugehen, falls es zu signifikanten Ergebnissen bei den durchgeführten Likelihood-Ratio-Tests kommen sollte. In einem solchen Fall erfolgt die Modellkontrolle grafisch, um das auffällige Item bzw. Items zu identifizieren. Jene Items, die weit entfernt von der 45°-Geraden liegen, sind nicht modellkonform. Diese sollen schrittweise aus dem Itempool entfernt werden, bis die Ergebnisse der Likelihood-Ratio-Tests für alle Teilungskriterien nicht mehr signifikant sind und somit a-posteriori Modellgeltung erreicht wird.

Außerdem wird der z-Test nach Fischer und Scheiblechner (1970, zitiert nach Kubinger et al., 2011) gerechnet, der jedoch wegen der Überhöhung des Risikos 1. Art nur deskriptiv interpretiert wird. Bei der Ausscheidung der Items sollen in erster Linie inhaltliche Gründe berücksichtigt werden, welche vor allem bei *Figurale Analogien* eine wichtige Rolle spielen.

Falls ein Item von allen Testpersonen oder keiner Testperson einer Teilstichprobe gelöst wurde, ist die Schätzung dieses Items für das jeweilige Teilungskriterium nicht möglich. Die Anzahl der schätzbaren Items ist immer um eins mehr als die Anzahl der Freiheitsgrade (df) und kann für alle Teilungskriterien den jeweiligen Tabellen über die Ergebnisse der Likelihood-Ratio-Tests entnommen werden.

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

Da Item 19 von allen Testpersonen richtig beantwortet wurde, kann es durch das Rasch-Modell nicht geschätzt werden. Somit beträgt die Anzahl der Items von *Figurale Analogien*, die auf Modellgeltung geprüft werden, 26.

Die Itemnummern der Aufgaben, die in den Analysen nicht berücksichtigt wurden, sind der Tabelle 7 zu entnehmen. Diese Items sind nicht schätzbar, weil sie von allen Testpersonen einer oder beider Teilstichproben gelöst wurden. Dabei treten Item 1 und Item 15 bei allen Teilungskriterien auf, da diese Items jeweils nur von einer einzigen Testperson in der gesamten Stichprobe nicht gelöst wurden.

Tabelle 7: Bei den Rasch-Modell Analysen nicht berücksichtigte Items von *Figurale Analogien*

Teilungskriterium	Itemnummern
Anzahl gelöster Aufgaben	1, 9, 11, 15, 20, 24, 26, 27
Bildungsniveau	1, 15
Alter	1, 15

Wie aus der Tabelle 8 ersichtlich, sind die Ergebnisse der durchgeführten Likelihood-Ratio-Tests nach Andersen für die ausgewählten Teilungskriterien nicht signifikant. Die dazugehörigen graphischen Modellkontrollen werden in den Abbildungen 7-9 präsentiert.

Tabelle 8: Ergebnisse der Likelihood-Ratio-Tests (*Figurale Analogien*, alle Items)

Teilungskriterium	χ^2	df	χ^2_{krit}	p
Anzahl gelöster Aufgaben	19,343	17	33,409	0,309
Bildungsniveau	26,408	23	41,638	0,282
Alter	24,554	23	41,638	0,374

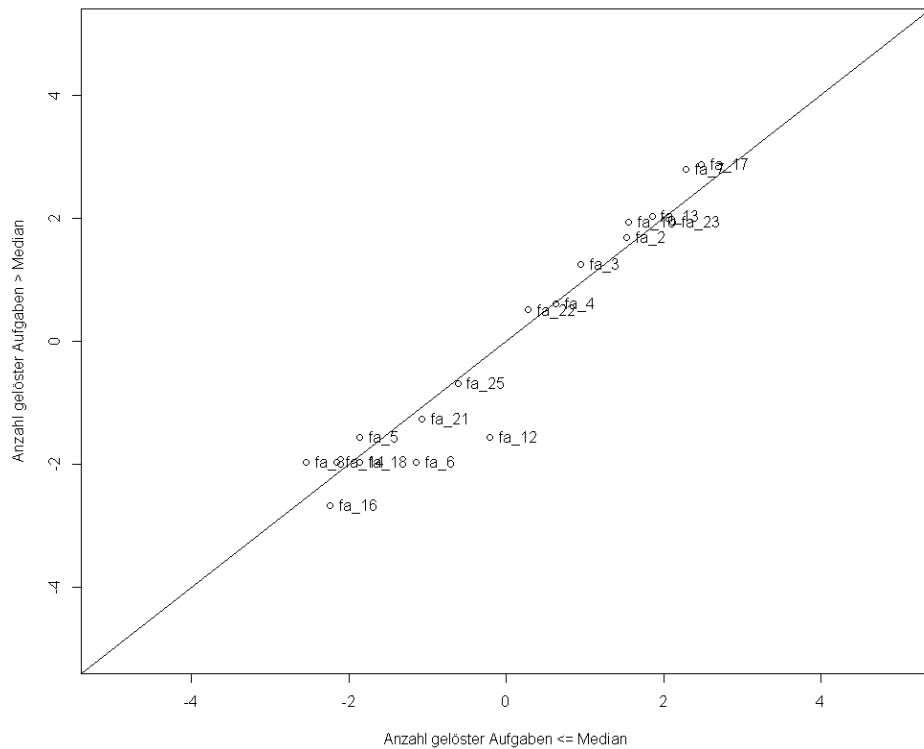


Abbildung 7: Grafische Modellkontrolle für das Teilungskriterium „Score“
(*Figurale Analogien*, alle Items)

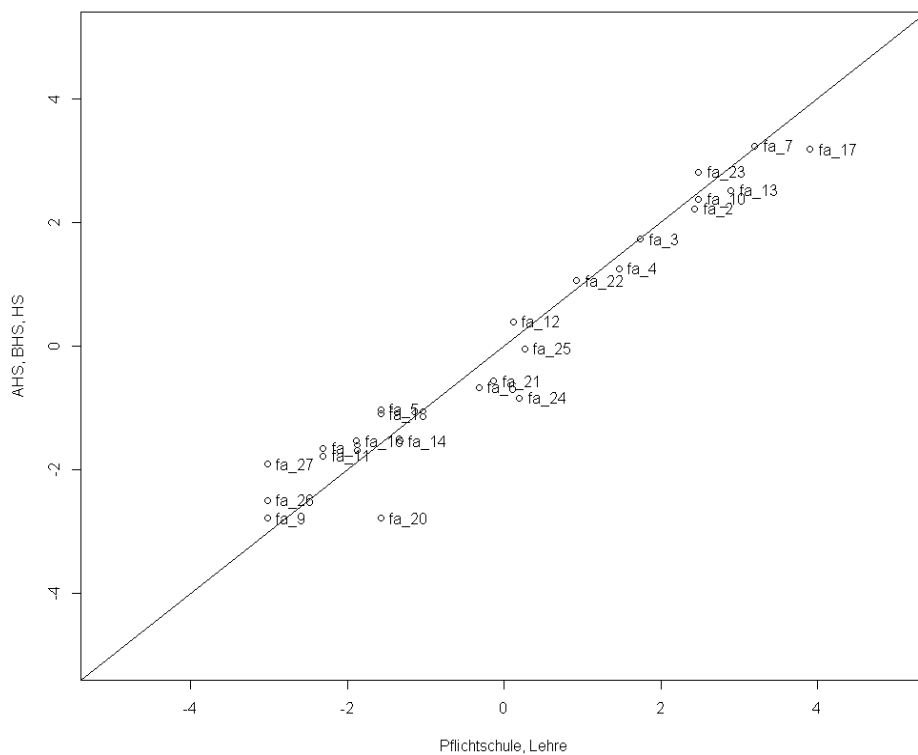


Abbildung 8: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“
(*Figurale Analogien*, alle Items)

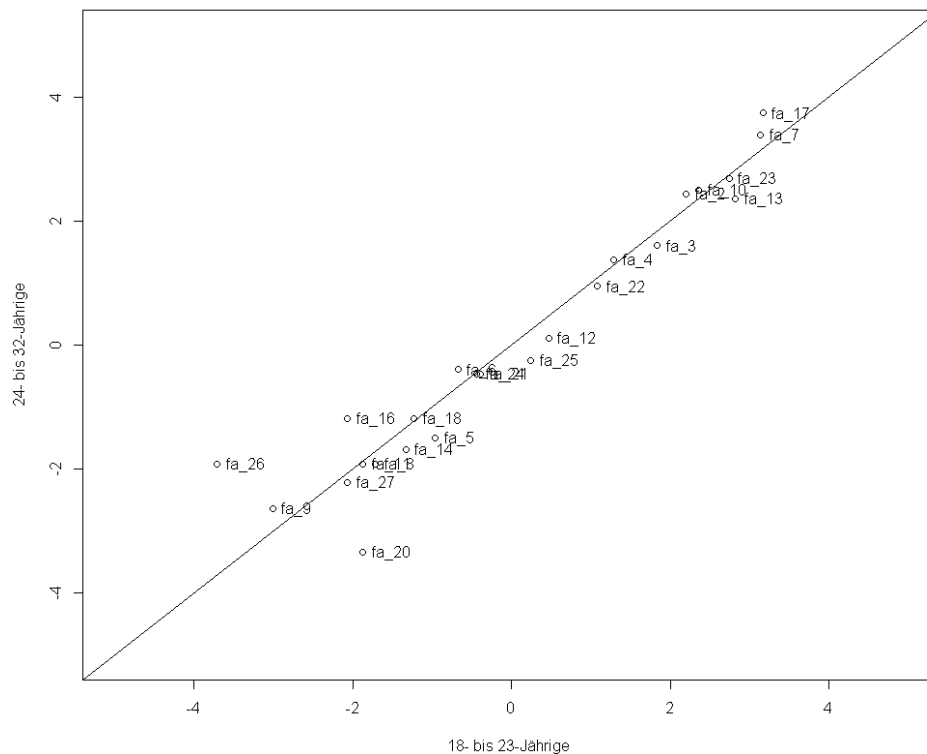


Abbildung 9: Grafische Modellkontrolle für das Teilungskriterium „Alter“
(*Figurale Analogien*, alle Items)

Auch die grafischen Modellkontrollen sind zufriedenstellend. Obwohl Item 12 beim Teilungskriterium „Anzahl gelöster Aufgaben“ und Item 20 bei den Teilungskriterien „Bildungsniveau“ und „Alter“ grafisch auffällig sind, wurden diese Items nicht eliminiert, da dafür keine inhaltlichen Gründe gefunden werden konnten und, noch wichtiger, bei den Likelihood-Ratio-Tests keine signifikanten Modellabweichungen beobachtet wurden.

Somit wird kein Item aus dem Itempool ausgeschieden und alle 27 Items des Verfahrens *Figurale Analogien* können als mit dem dichotomen logistischen Modell von Rasch konform angesehen werden. Folglich kann die Anzahl der gelösten Aufgaben als Testkennwert herangezogen werden, und das Verfahren misst eindimensional die Fähigkeit zu kristallisiertem schlussfolgerndem Denken.

Die Ergebnisse der Schätzungen der Itemschwierigkeitsparameter sind Tabelle 9 zu entnehmen. Dabei signalisiert ein höherer geschätzter Wert einen Anstieg in der Schwierigkeit.

Tabelle 9: Schätzungen der Itemschwierigkeitsparameter bei *Figurale Analogien*

Itemnummer	Itemschwierigkeitsparameter
2	2,584
3	2,036
4	1,617
5	-0,852
6	-0,253
7	3,526
8	-1,493
9	-2,539
10	2,707
11	-1,594
12	0,628
13	2,922
14	-1,164
15	-3,940
16	-1,317
17	3,669
18	-0,908
20	-1,965
21	-0,122
22	1,324
23	3,012
24	-0,154
25	0,354
26	-2,311
27	-1,827

10.3. Ergebnisse der Rasch-Modell Analysen bei AN-TOP

Der Itempool der auf Modellgeltung zu prüfenden Items besteht bei AN-TOP aus 54 Aufgaben, da Item 1a und Item 1b von allen Testpersonen der Stichprobe gelöst wurden und deswegen nicht geschätzt werden können.

Auch bei den Analysen zu den Items von AN-TOP gibt es mehrere nicht schätzbare Items, die von keiner oder allen Testpersonen von zumindest einer der beiden Teilstichproben gelöst wurden. Diese werden für jedes Teilungskriterium in Tabelle 10 präsentiert. Sofern nicht anders notiert, beziehen sich die angegebenen Itemnummern auf alle Berechnungsdurchgänge.

Tabelle 10: Bei den Rasch-Modell Analysen nicht berücksichtigte Items von AN-TOP

Teilungskriterium	Itemnummern
Anzahl gelöster Aufgaben	2a, 2b, 3a, 4a, 4b, 5a, 5b, 6a, 6b, 9a*, 13b, 19a*, 19b, 21a♦
Bildungsniveau	2a, 2b, 5a
Alter	2a, 2b, 5a

* Diese Items konnten beim zweiten und dritten Berechnungsdurchgang geschätzt werden.

♦ Dieses Item konnte beim zweiten Berechnungsdurchgang geschätzt werden.

Bei der Durchführung von Likelihood-Ratio-Tests nach Andersen mit allen schätzbaren Items kam es, wie aus Tabelle 11 ersichtlich, in dem Teilungskriterium „Anzahl gelöster Aufgaben“ zu signifikanten Modellabweichungen.

Die grafischen Modellkontrollen der drei Teilungskriterien sind den Abbildungen 10-12 zu entnehmen. Dabei wurden im Teilungskriterium „Anzahl gelöster Aufgaben“ Items 11a, 12b und 20b grafisch auffällig.

Bevor Items ausgeschlossen wurden, wurde für das Teilungskriterium „Anzahl gelöster Aufgaben“ auch der z-Test nach Fischer und Scheiblechner berechnet, bei dem die Items 10a, 11a und 20b auffällig wurden (siehe Tabelle 12).

Tabelle 11: Ergebnisse der Likelihood-Ratio-Tests (AN-TOP, alle Items)

Teilungskriterium	χ^2	df	χ^2_{krit}	p
Anzahl gelöster Aufgaben	79,903	39	62,428	0
Bildungsniveau	63,469	50	76,154	0,096
Alter	64,574	50	76,154	0,081

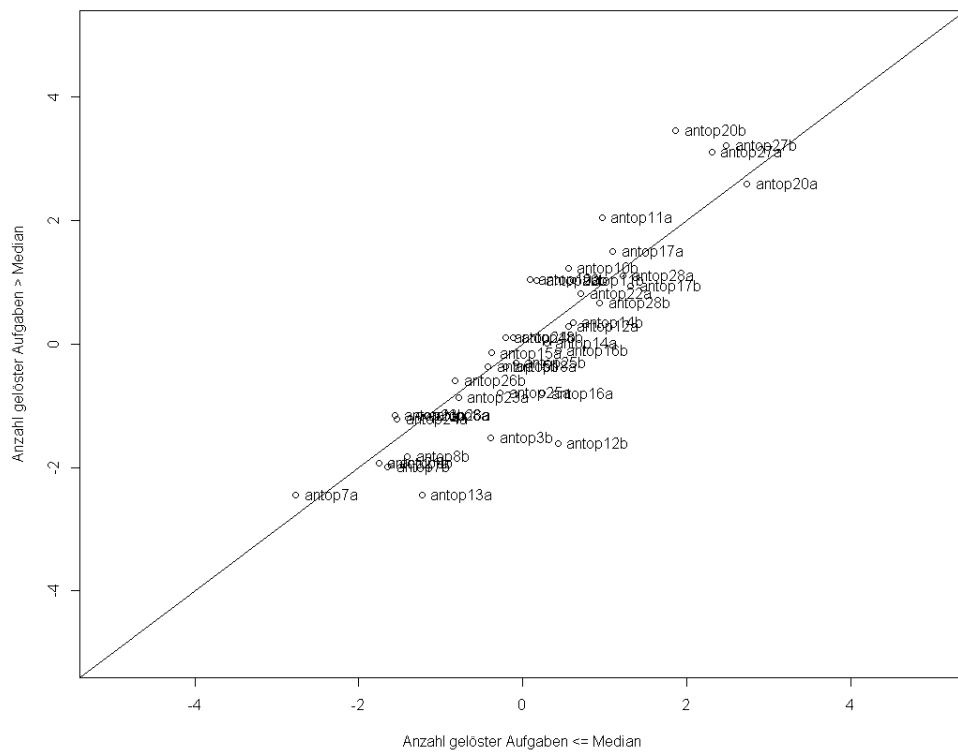


Abbildung 10: Grafische Modellkontrolle für das Teilungskriterium „Score“ (AN-TOP, alle Items)

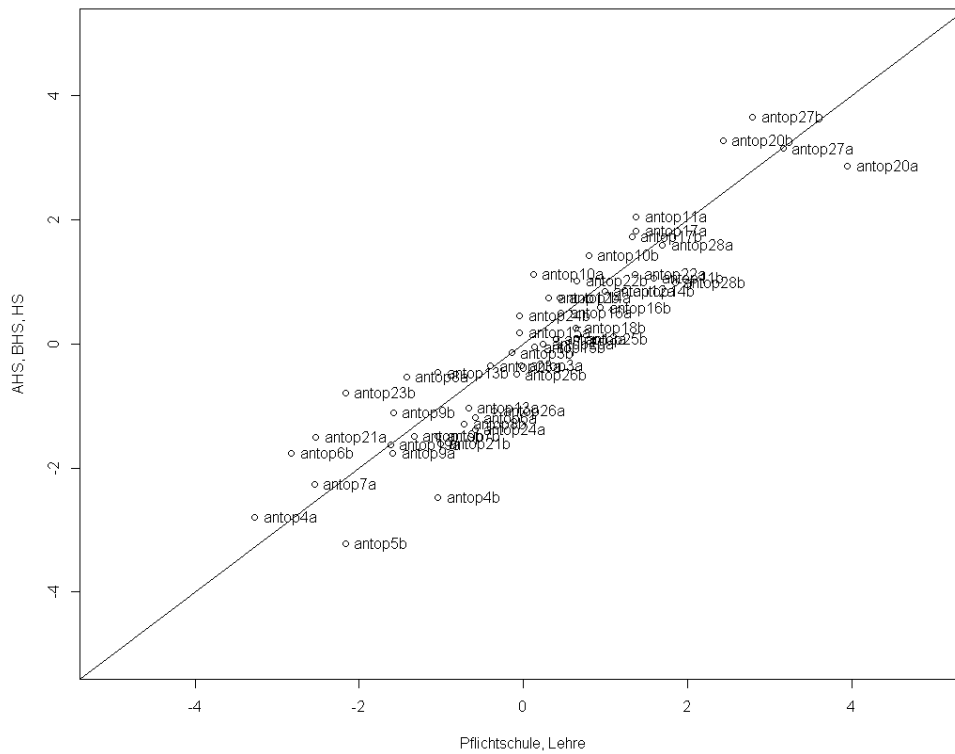


Abbildung 11: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“
(AN-TOP, alle Items)

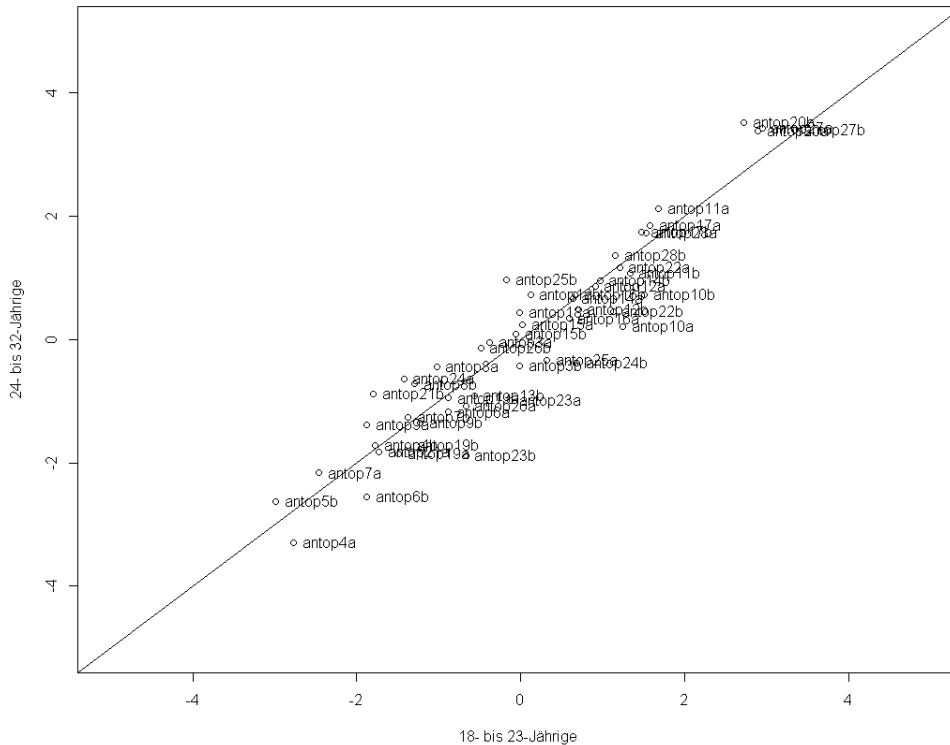


Abbildung 12: Grafische Modellkontrolle für das Teilungskriterium „Alter“
(AN-TOP, alle Items)

Tabelle 12: Ergebnisse des z-Tests für das Teilungskriterium „Score“
(AN-TOP, alle Items)

Itemnummer	z-Wert	<i>p</i>
3b	-1,847	0,065
7a	0,298	0,766
7b	-0,321	0,748
8a	0,012	0,990
8b	-0,550	0,582
9b	-0,431	0,667
10a	3,118	0,002
10b	1,598	0,110
11a	3,606	0,000
11b	0,975	0,329
12a	-0,804	0,422
12b	-1,977	0,048
13a	-1,203	0,229
14a	-0,535	0,593
14b	-0,766	0,444
15a	0,401	0,689
15b	0,158	0,874
16a	-2,095	0,036
16b	-1,026	0,305
17a	1,309	0,191
17b	-0,853	0,393
18a	-0,386	0,699
18b	0,410	0,682
20a	-0,264	0,792
20b	4,079	0,000
21b	-0,231	0,817
22a	0,359	0,720

22b	2,041	0,041
23a	-0,170	0,865
23b	0,493	0,622
24a	0,561	0,575
24b	0,579	0,563
25a	-1,021	0,307
25b	-0,397	0,691
26a	0,134	0,893
26b	0,489	0,625
27a	1,481	0,139
27b	1,845	0,065
28a	-0,385	0,700
28b	-0,605	0,545

Für die auffälligen Items konnten keine inhaltlichen Gründe gefunden werden. Beim nächsten Berechnungsdurchgang wurde das Item 11a, das in dem Teilungskriterium „Anzahl gelöster Aufgaben“ sowohl beim z-Test nach Fischer und Scheiblechner als auch grafisch auffällig war, aus dem Itempool entfernt. Es kam wiederum zu signifikanten Modellabweichungen im selben Teilungskriterium (siehe Tabelle 13).

Tabelle 13: Ergebnisse der Likelihood-Ratio-Tests
(AN-TOP, nach Ausschluss von Item 11a)

Teilungskriterium	χ^2	df	χ^2_{krit}	p
Anzahl gelöster Aufgaben	82,349	41	64,950	0
Bildungsniveau	60,106	49	74,920	0,133
Alter	64,423	49	74,920	0,069

Basierend auf Grafischen Modellkontrollen und z-Tests wurde anschließend das Item 20b ausgeschieden. Nach Ausschluss von Items 11a und 20b fallen die Likelihood-Ratio-Tests nach Andersen in allen Teilungskriterien nicht signifikant aus (siehe Tabelle 14). Die Ergebnisse der Graphischen Modellkontrollen sind auch zufriedenstellend und befinden sich in den Abbildungen 13-15. Die beiden ausgeschlossenen Items können dem Anhang unter F entnommen werden.

Tabelle 14: Ergebnisse der Likelihood-Ratio-Tests
(AN-TOP, nach Ausschluss von Items 11a und 20b)

Teilungskriterium	χ^2	<i>df</i>	χ^2_{krit}	<i>p</i>
Anzahl gelöster Aufgaben	57,674	39	62,428	0,027
Bildungsniveau	56,695	48	73,683	0,182
Alter	60,739	48	73,683	0,103

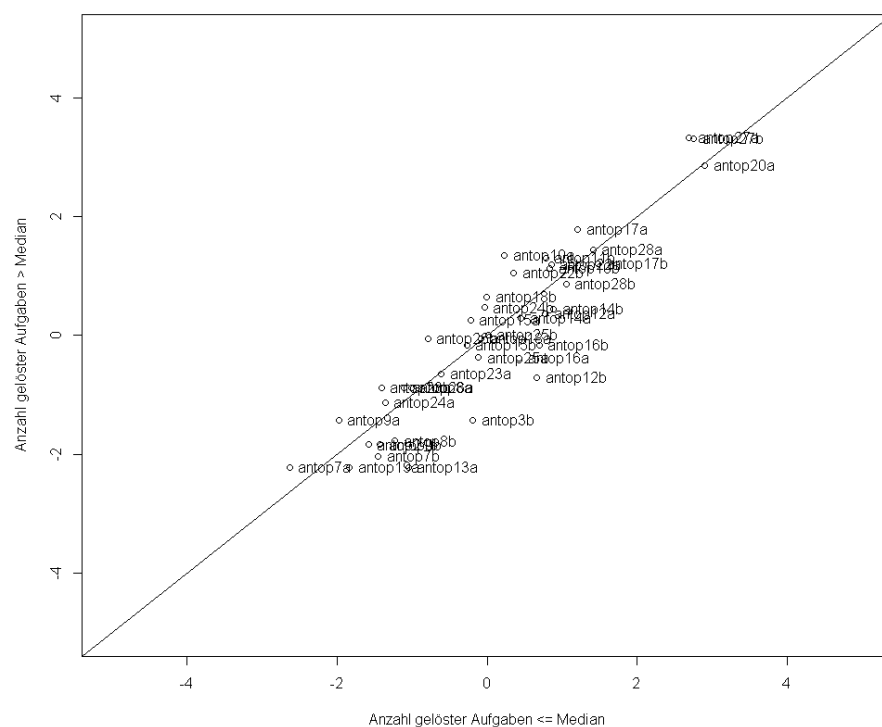


Abbildung 13: Grafische Modellkontrolle für das Teilungskriterium „Score“
(AN-TOP, nach Ausschluss von Items 11a und 20b)

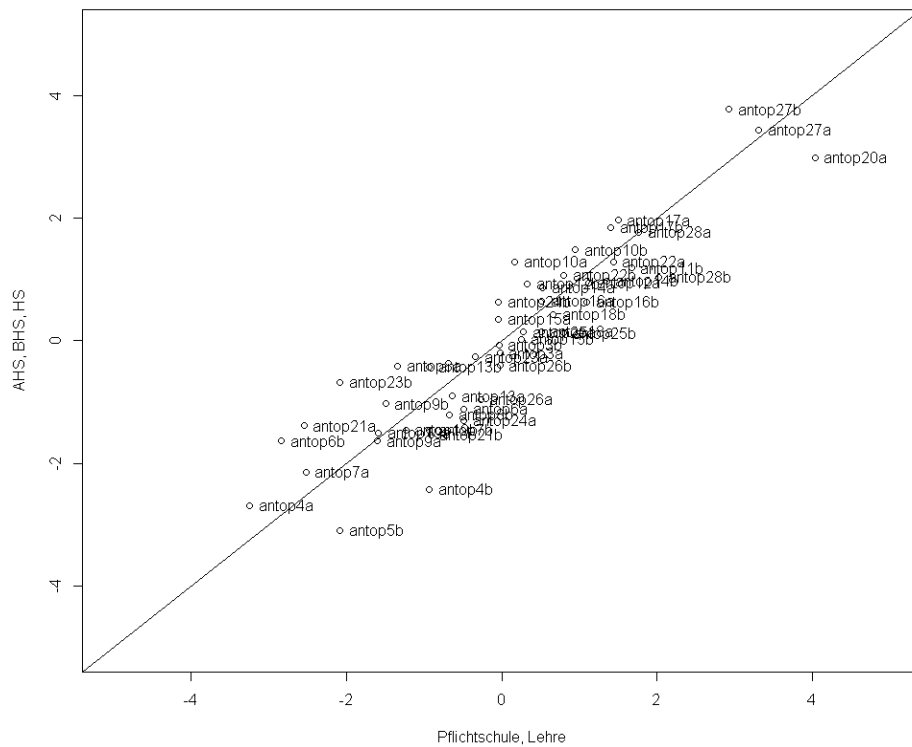


Abbildung 14: Grafische Modellkontrolle für das Teilungskriterium „Bildungsniveau“
(AN-TOP, nach Ausschluss von Items 11a und 20b)

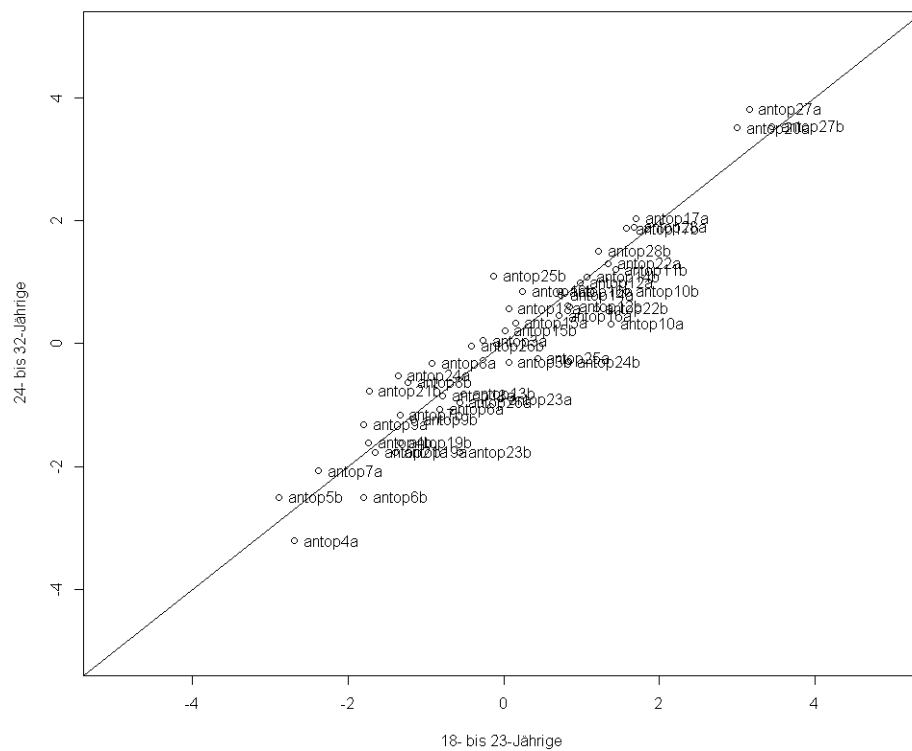


Abbildung 15: Grafische Modellkontrolle für das Teilungskriterium „Alter“
(AN-TOP, nach Ausschluss von Items 11a und 20b)

Diesen Ergebnissen zufolge kann den übrigen 54 Items a-posteriori Modellgültigkeit zugeschrieben werden. Folglich misst das Verfahren *AN-TOP* eindimensional logisch-schlussfolgerndes Denken.

Die Ergebnisse der Schätzungen der Itemschwierigkeitsparameter von *AN-TOP* sind in Tabelle 15 zu finden.

Tabelle 15: Schätzungen der Itemschwierigkeitsparameter bei *AN-TOP*

Itemnummer	Itemschwierigkeitsparameter	Itemnummer	Itemschwierigkeitsparameter
2b	-4,211	16a	0,826
3a	0,082	16b	0,997
3b	0,162	17a	2,052
4a	-2,623	17b	1,921
4b	-1,469	18a	0,486
5a	-4,211	18b	0,720
5b	-2,489	19a	-1,307
6a	-0,672	19b	-1,173
6b	-1,765	20a	3,427
7a	-2,033	21a	-1,473
7b	-1,046	21b	-1,088
8a	-0,415	22a	1,539
8b	-0,780	22b	1,217
9a	-1,386	23a	-0,071
9b	-0,956	23b	-0,732
10a	1,186	24a	-0,779
10b	1,553	24b	0,647
11b	1,550	25a	0,403
12a	1,211	25b	0,567
12b	0,973	26a	-0,496
13a	-0,590	26b	-0,058
13b	-0,367	27a	3,613
14a	0,975	27b	3,700
14b	1,302	28a	1,974
15a	0,446	28b	1,550
15b	0,315		

10.4. Deskriptive Ergebnisse zu *Figurale Analogien*

Da der Test *Figurale Analogien* im Rahmen dieser Arbeit neu konstruiert wurde, werden im Folgenden deskriptive Ergebnisse zu den Rohwerten, Lösungshäufigkeiten und Itemschwierigkeitsparametern angeboten.

Bei der Auswertung dieses Verfahrens werden die Markierungen von korrekten Antworten zusammengezählt. Somit beträgt der maximal erreichbare Rohwert in Bezug auf die Rasch-Modell-konformen Items von *Figurale Analogien* 27. Der Tabelle 16 sowie dem Histogramm in Abbildung 16 ist zu entnehmen, dass die Testpersonen der Stichprobe ($n = 331$) im Durchschnitt einen Rohwert von 21 (Standardabweichung: 2,225) erzielt haben, d.h. durchschnittlich wurden ca. 78 % (Standardabweichung: 8,24) der Items gelöst. Mindestens wurden zwölf Items richtig beantwortet und keine Testperson hat alle Items gelöst.

Tabelle 16: Häufigkeiten der Rohwerte bei *Figurale Analogien*

Rohwert	absolute Häufigkeit ($n = 331$)	relative Häufigkeit (in Prozent)
12	1	0,30
14	3	0,91
15	6	1,81
16	10	3,02
17	10	3,02
18	31	9,37
19	39	11,78
20	60	18,13
21	54	16,31
22	67	20,24
23	29	8,76
24	17	5,14
25	4	1,21

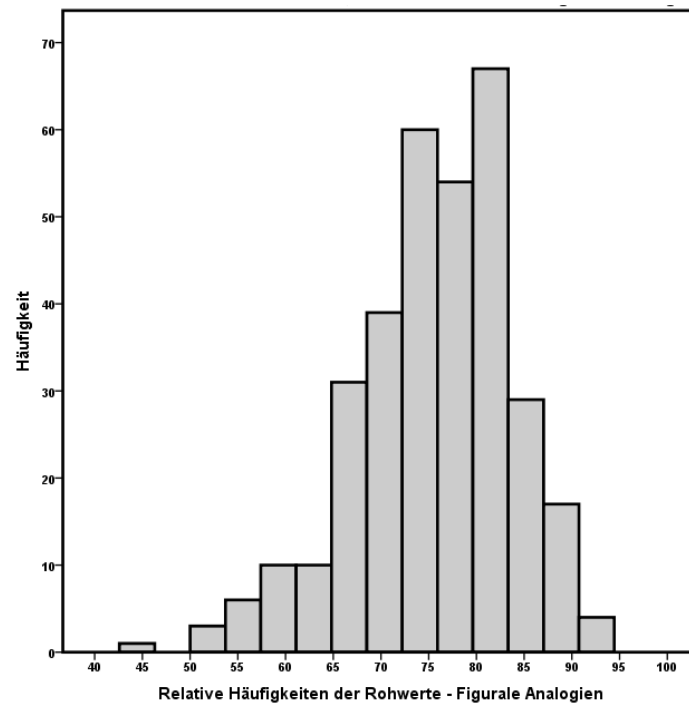


Abbildung 16: Verteilung der relativen Häufigkeiten der Rohwerte bei *Figurale Analogien* ($n = 331$)

Die Schwierigkeitsparameter der Items von *Figurale Analogien* erstrecken sich dabei auf einem Kontinuum von $-3,940$ bis $+3,669$ (vgl. dazu Tabelle 9). Ihre Verteilung ist der Abbildung 17 zu entnehmen.

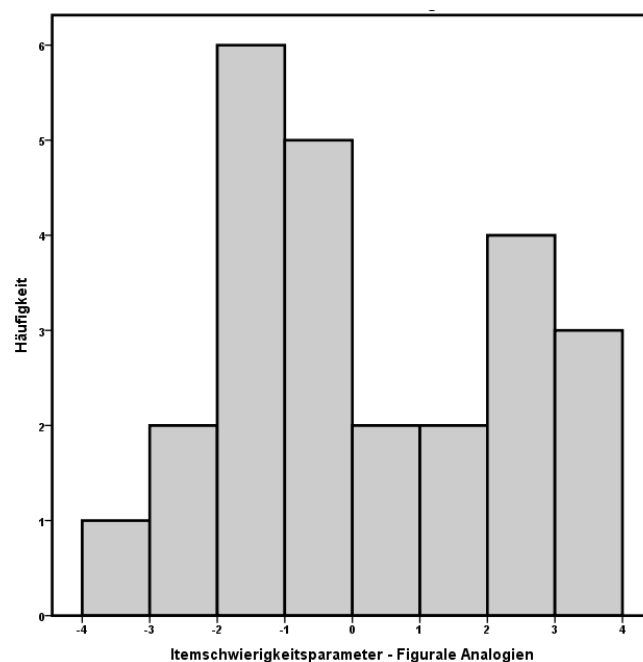


Abbildung 17: Verteilung der Itemschwierigkeitsparameter bei *Figurale Analogien*

Über die Lösungshäufigkeiten der Items in der Stichprobe werden in Tabelle 17 berichtet. Ein Item wurde von allen Testpersonen gelöst; zehn weitere Items wurden von mindestens 95 % der Testpersonen richtig beantwortet. Diese sind vor allem diejenigen Items, die während der Itemkonstruktion als „leicht“ oder „leicht bis mittelschwierig“ eingestuft wurden (vgl. dazu Tabelle 3).

Tabelle 17: Lösungshäufigkeiten der Items von *Figurale Analogien*

Lösungshäufigkeit (in Prozent) (absolut, $n = 331$)		Schwierigkeitsstufe bei der Itemkonstruktion	Itemnummer
17,52	58	schwierig	17
19,64	65	mittelschwierig-schwierig	7
28,70	95	schwierig	23
30,51	101	mittelschwierig	13
35,05	116	mittelschwierig	10
37,76	125	schwierig	2
50,45	167	schwierig	3
60,12	199	schwierig	4
66,47	220	mittelschwierig	22
79,15	262	schwierig	12
83,08	275	mittelschwierig	25
88,52	293	mittelschwierig	21
88,82	294	mittelschwierig	24
89,73	297	mittelschwierig-schwierig	6
93,96	311	leicht	5
94,26	312	mittelschwierig	18
95,47	316	mittelschwierig-schwierig	14
96,07	318	leicht	16
96,68	320	mittelschwierig	8
96,98	321	leicht-mittelschwierig	11
97,58	323	leicht	27
97,89	324	leicht-mittelschwierig	20
98,49	326	leicht-mittelschwierig	26
98,79	327	mittelschwierig	9
99,70	330	leicht	1
99,70	330	leicht	15
100	331	leicht	19

10.5. Deskriptive Ergebnisse zu AN-TOP

Um Modellgeltung zu erreichen, wurden in Abschnitt 10.3 zwei auffällige Items aus dem Itempool von AN-TOP ausgeschlossen. Da diese Items nicht bei allen Testheften vorkamen, enthalten Testformen 1 und 2 jeweils 26 und Testform 3 28 Items der insgesamt 54 Rasch-Modell-konformen Aufgaben.

Um einen Vergleich des prozentualen Anteils der richtig beantworteten Items unabhängig von dem bearbeiteten Testheft zu ermöglichen, werden hier nur die relativen Häufigkeiten der Rohwerte dargestellt (siehe dazu das Histogramm in Abbildung 18). Im Durchschnitt lösten die Testpersonen der Stichprobe 79 % der Items (Standardabweichung: 13,28). Die Rohwerte variierten bei den ersten zwei Testformen zwischen 6 und 26 (max. = 26) und bei Testform 3 zwischen 10 und 26 (max. = 28). Der maximale Rohwert wurde dabei nur von einer einzigen Testperson bei der Bearbeitung von Testform 1 erreicht.

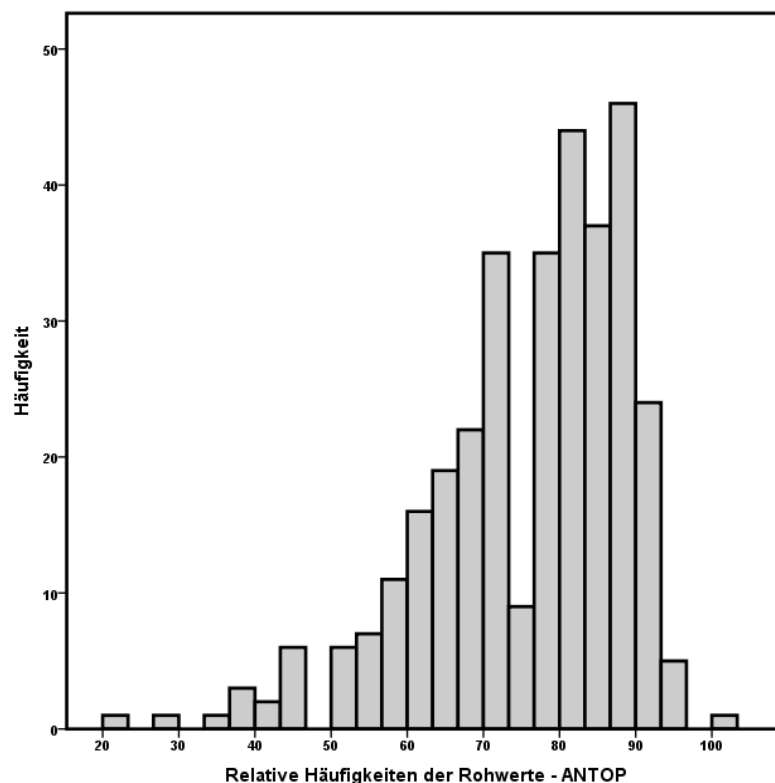


Abbildung 18: Verteilung der relativen Häufigkeiten der Rohwerte bei AN-TOP ($n = 331$)

Die Schwierigkeitsparameter der Rasch-Modell konformen Items von *AN-TOP* bewegen sich zwischen $-4,211$ und $+3,700$ (vgl. dazu Tabelle 15). Das Histogramm ihrer Verteilung wird in Abbildung 19 dargestellt.

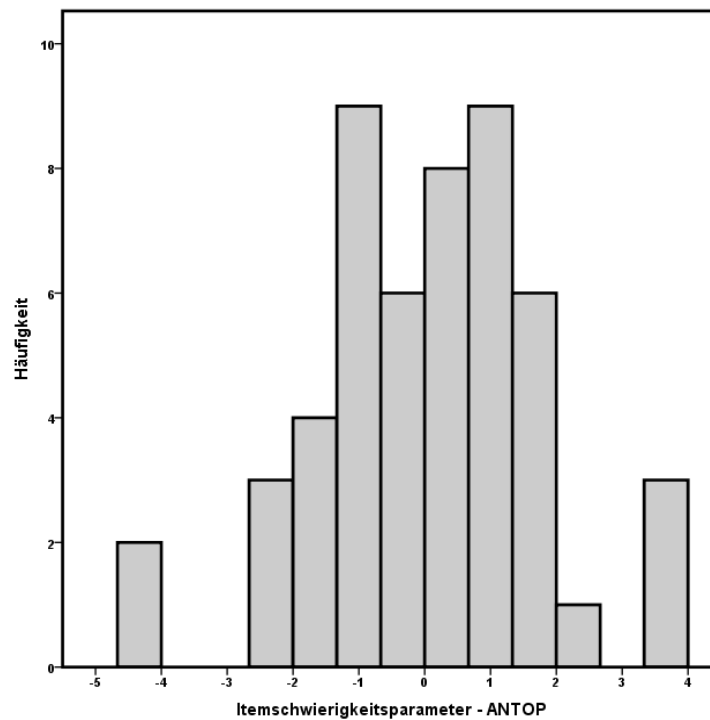


Abbildung 19: Verteilung der Itemschwierigkeitsparameter bei *AN-TOP*

10.6. Korrelation zwischen *Figurale Analogien* und *AN-TOP*

Die Berechnung der Korrelation zwischen den Personenparametern von *Figurale Analogien* und *AN-TOP* soll zur Untersuchung der diskriminanten Validität dienen, da mit diesen beiden Verfahren unterschiedliche Konstrukte erfasst werden (vgl. dazu z.B. Lienert & Raatz, 1998). Dafür soll zuerst mit Hilfe von theoretischen Überlegungen ein Grenzwert für die maximal zu tolerierende Korrelation festgelegt werden (Moosbrugger & Kelava, 2012).

Mit *Figurale Analogien* wird kristallisiertes schlussfolgerndes Denken, mit *AN-TOP* logisch-schlussfolgerndes Denken erfasst. Diese Konstrukte werden laut dem Intelligenzmodell von Cattell zwei unterschiedlichen Faktoren der Intelligenz zugeordnet, nämlich der kristallisierten und der fluiden Intelligenz.

Bei der Formulierung der Investment-Theorie stellte Cattell (1987) fest, dass diese Faktoren voneinander nicht völlig unabhängig sind, sondern miteinander korrelieren. Dieser positive Zusammenhang variierte in seinen Untersuchungen meist zwischen 0,40 und 0,50 (Cattell, 1987, S. 116). Deswegen wird hier $r = 0,40$ als Grenzwert festgelegt, der von der Korrelation zwischen den Personenparametern von *Figurale Analogien* und *AN-TOP* für die Annahme der Hypothese H_0 (3) nicht überschritten werden sollte. Dabei ist weiters zu beachten, dass beide Verfahren von figuralem Arbeitsmaterial Gebrauch machen. Aus diesem Grund ist eine geringe Korrelation nicht ganz auszuschließen.

Die Bestimmung der Korrelation erfolgte über die Personenparameter, die nach dem Ausscheiden der nicht modellkonformen Items berechnet wurden. Da die Personenparameter durch die Geltung des Rasch-Modells als normalverteilt und intervallskaliert angesehen werden können, wird zur Berechnung ihres Zusammenhangs der Pearson-Korrelationskoeffizient herangezogen.

Es besteht ein geringer, positiver Zusammenhang zwischen den Personenparametern von *Figurale Analogien* und *AN-TOP*. Dabei beträgt der Korrelationskoeffizient nach Pearson $r = 0,217$ und das Bestimmtheitsmaß $B = 0,047$. Somit wird ca. 5 % der Varianz der Leistung in einem der beiden Tests durch das Ergebnis in dem jeweils anderen Test erklärt.

Der Zusammenhang zwischen den beiden Verfahren liegt unter dem aufgrund theoretischer Überlegungen bestimmten Grenzwert. Dieses Ergebnis erlaubt, die diskriminante Validität als gegeben zu betrachten. Folglich werden mit *Figurale Analogien* und *AN-TOP* zwei unterschiedliche Konstrukte erfasst.

IV. DISKUSSION UND AUSBLICK

Im Rahmen dieser Arbeit ist es gelungen, ein neues Verfahren zur Erfassung des kristallisierten schlussfolgernden Denkens zu entwickeln, das im Rahmen der Wissensdiagnostik eingesetzt werden kann. Zur Lösung der Rasch-Modell-konformen Aufgaben müssen die Testpersonen inhaltlich sinnvolle, figurale Analogien durch Auswahl einer der fünf Antwortmöglichkeiten vervollständigen.

Analogie-Aufgaben werden schon seit den Anfängen der Intelligenzforschung bei psychologisch-diagnostischen Verfahren eingesetzt. Jedoch geschah dies bis jetzt immer nur zur Erfassung von logisch-schlussfolgerndem Denken, das sensu Cattell der fluiden Intelligenz zugeordnet wird. Mit dem neu entwickelten Test *Figurale Analogien* konnte gezeigt werden, dass mittels inhaltlich sinnvollem Aufgabenmaterial Fragestellungen in Form von Analogien auch bei der Messung des anderen Intelligenzfaktors Anwendung finden können. Die verwendeten Items verlangen nicht nur die Fähigkeit zum schlussfolgernden Denken, sondern durch bildungs- und kulturspezifische Komponenten wird die kristallisierte Intelligenz ebenfalls angesprochen.

Von vielen anderen Wissenstests hebt sich das Verfahren *Figurale Analogien* durch die Erfassung des Allgemeinwissens im figuralen Bereich ab. Da Bilder und Symbole in unserer Kultur im alltäglichen Leben eine wichtige Rolle spielen und der Wissenserwerb nicht nur im verbalen Bereich stattfindet, sollte auch seine Erfassung nicht auf diesen Bereich limitiert werden. Die als Aufgabenmaterial verwendeten Bilder waren für die Testpersonen leicht verständlich, so dass eine weitere Verbalisierung der dargestellten Objekte nicht notwendig war. Aus diesem Grund erfolgt die Bearbeitung der Items unabhängig von sprachlichen Fähigkeiten der Testpersonen. Durch entsprechende Überarbeitung der Instruktionen wäre es auch möglich, den Test in eine ganz sprachfreie Version umzugestalten.

Für die 27 neu konstruierten Items von *Figurale Analogien* wurde die a-priori Geltung des dichotomen logistischen Modells von Rasch erreicht. Mit diesem Test wird eindimensional das Konstrukt kristallisiertes schlussfolgerndes Denken gemessen. Im Vergleich dazu misst der Test *AN-TOP*, der ebenfalls bei der empirischen Untersuchung eingesetzt

wurde, eindimensional logisch-schlussfolgerndes Denken. Die Modellkonformität seiner Items wurde a-posteriori, nach Ausschluss von zwei Items, erzielt. Bei beiden Verfahren kann daher die Anzahl gelöster Aufgaben pro Person als Testkennwert herangezogen werden. Durch diesen fairen Verrechnungsmodus wird auch das Gütekriterium Skalierung erfüllt.

Zur Bestimmung der diskriminanten Validität wurde die Korrelation zwischen den Personenparametern von *Figurale Analogien* und *AN-TOP* berechnet. Da die beiden Verfahren jeweils unterschiedliche Faktoren der Intelligenz (kristallisierte vs. fluide) messen sollen, entspricht das Ergebnis einer geringen Korrelation den Erwartungen. Weitere Validierungsstudien, die sich mit der prognostischen Validität und/oder mit der Übereinstimmungsvalidität zwischen *Figurale Analogien* und anderen Wissenstests befassen, wären wünschenswert.

Weiters kann dem Test Augenscheinvalidität zugesprochen werden, d.h. für die Testpersonen ist die zu erfassende Fähigkeit offensichtlich (vgl. dazu Kubinger, 2009). Es wird nämlich schon auf der Instruktionsseite darauf hingewiesen, dass sich die Items auf Allgemeinwissen beziehen. Zusätzlich kann durch den sinnvollen und alltagsnahen Inhalt der Aufgaben eine gesteigerte Akzeptanz bezüglich der Testung seitens der TeilnehmerInnen erwartet werden. Bei der Durchführung der Untersuchung wurde beobachtet, dass die TeilnehmerInnen die Bearbeitung von *Figurale Analogien* interessanter fanden als der anderen beiden Verfahren. Auch nach der Testung haben viele Testpersonen ihr Interesse an dem Verfahren verbalisiert. Möglicherweise liegt dies daran, dass die numerischen bzw. abstrakt-figuralen Items der anderen Verfahren eher den Charakter einer (schulischen) Leistungssituation besitzen. Wenn keine Skepsis gegenüber der Messintention besteht und die Testpersonen eine positive Einstellung gegenüber der Bearbeitung der Items zeigen, ist folglich auch mit höherer Motivation bei der Durchführung von *Figurale Analogien* zu rechnen.

In der Phase der Itemkonstruktion wurden die Aufgaben von *Figurale Analogien* zu jeweils einem der drei übergreifenden Themenbereiche (*Menschen, Naturwissenschaften, Alltag*) zugeordnet, um eine Überrepräsentation von Fragen aus einem bestimmten Bereich zu vermeiden. Die Existenz dieser Themenbereiche als voneinander trennbare Faktoren wurde im Rahmen dieser Arbeit nicht untersucht. Statistische Analysen diesbezüglich könnten in künftigen Validierungsuntersuchungen, vor allem in Bezug zur Kon-

struktvalidität durchgeführt werden, jedoch vorzugsweise erst nach einer Erweiterung des Itempools. Für den Fall der Nachweisbarkeit dieser Themenbereiche ist jedoch zu notieren, dass dabei eine Interpretation der Rohwerte getrennt nach Themengebieten nicht intendiert bzw. empfohlen wird.

In Bezug auf das Gütekriterium Objektivität ist die standardisierte Form der Durchführung sowie der Auswertung zu erwähnen. Bei der Vorgabe von *Figurale Analogien* brauchen die TestleiterInnen und die Testpersonen nur minimal miteinander zu interagieren. Die Instruktionen, die von den Testpersonen selbständig gelesen werden sollen, wurden möglichst kurz und präzise gehalten und enthalten alle notwendigen Informationen zum Bearbeitungsvorgang. Verrechnungssicherheit (Auswertungsobjektivität) wird durch die Bestimmung einer einzigen richtigen Antwort aus fünf Antwortmöglichkeiten gewährleistet.

Die Auswertung, bei der lediglich die Anzahl der korrekten Antworten gerechnet werden, ist durch die relativ geringe Itemanzahl einfach und schnell. Die Durchführung beansprucht ebenfalls vergleichsweise wenig Zeit. In der Untersuchung wurde festgestellt, dass die Bearbeitungszeit im Durchschnitt zehn Minuten beträgt. Zusätzlich zu geringen Zeitanprüchen der Bearbeitung und Auswertung trägt außerdem die Möglichkeit der Gruppenvorgabe zur Erfüllung des Gütekriteriums Ökonomie bei.

Das ausgewählte Multiple-Choice-Antwortformat „1 aus 5“ ermöglicht die Vorgabe in der Gruppe und eine kurze Bearbeitungsdauer, dafür muss aber eine a-priori Ratewahrscheinlichkeit von 0,20 in Kauf genommen werden. Durch die Geltung des Rasch-Modells wird jedoch sichergestellt, dass die Rateeffekte bei allen Items im gleichen Ausmaß sind, so dass die Anzahl gelöster Aufgaben als Testkennwert herangezogen werden kann (vgl. dazu Kubinger, 2009). Trotzdem zeigt sich beim Betrachten der Lösungshäufigkeiten der Items sowie der Verteilung der relativen Häufigkeiten der Rohwerte, dass einige Items eher leichter zu lösen sind als bei der Bestimmung der Schwierigkeitsstufen während der Itemkonstruktion angenommen wurde. Dabei sollte untersucht werden, ob die unbeabsichtigten niedrigen Itemschwierigkeiten durch Rateeffekte beim Multiple-Choice-Antwortformat verursacht werden.

Eine weitere Erklärung für die unerwarteten Lösungshäufigkeiten der Items könnte an den Umständen der durchgeführten Untersuchung liegen. Die Testleiterinnen berichteten

nämlich über Unruhen während der Vorgabe, die teilweise durch die relativ kleinen Testungsräume nicht vermieden werden konnten. Da dieses Problem vor allem das Verfahren *Figurale Analogien* betraf, ist es höchstwahrscheinlich auf den Inhalt und das Layout des Tests zurückzuführen. Der abwechslungsreiche Inhalt der Aufgaben könnte die Testpersonen zum sofortigen Informationsaustausch und das Layout, das pro Seite nur drei großgedruckte Items anbietet, zu unehrlichen Arbeitstendenzen geleitet haben. Positiv zu erwähnen ist dabei, dass diese Punkte gleichzeitig auch die Motivation der Testpersonen während der Bearbeitung der Aufgaben gesteigert zu haben scheinen.

Falls das Antwortformat und/oder die Unruhen während der Untersuchung zu höheren Lösungshäufigkeiten beigetragen haben sollten, sollte nach Verbesserungsmöglichkeiten gesucht werden. Bezüglich des Antwortformats wäre in erster Linie die Effektivität der Distraktoren als plausible Antwortmöglichkeiten zu prüfen. Eine sequentielle Vorgabe der Antwortmöglichkeiten, bei der jedes Bild einzeln nacheinander bewertet werden soll, kann die Ratewahrscheinlichkeit, abhängig von der Position der korrekten Antwort, reduzieren. Da jedes Item nur eine einzige Lösung hat, die nicht von den anderen Antwortmöglichkeiten abhängt, wäre die Änderung der Vorgabe in ein sequentielles Format ohne Überarbeitung der Items zulässig. Jedoch kann dieses Format nur bei computergestützten Tests angewendet werden. Eine Umwandlung von *Figurale Analogien* in ein computergestütztes Verfahren würde außerdem unehrliche Arbeitstendenzen während der Vorgabe minimieren, da jede Testperson vor eigenem Bildschirm und im eigenen Tempo arbeiten würde. So könnten auch Unruhen vermieden werden, die dabei entstehen, wenn Testpersonen, die die Items schneller bearbeiten, auf die anderen warten müssen. Zusätzlich kann auch die monetäre Ökonomie durch Computertestungen gesteigert werden, weil im Paper-Pencil-Format insgesamt zehn A4-Seiten für 27 Items gebraucht werden.

Weiters ist im Hinblick auf die Untersuchungsstichprobe zu bemängeln, dass nur Personen zwischen 18 und 32 Jahren, die in den Sicherheitsakademien der Polizei ausgebildet werden, getestet wurden. Da beim Intelligenzmodell von Cattell der kristallisierten im Vergleich zur fluiden Intelligenz eine schwächere Abnahme über die Zeit vorausgesagt wird, wäre es für künftige Untersuchungen interessant, auch Personen aus anderen Altersgruppen zu testen. Außerdem bestand die Untersuchungsstichprobe größtenteils aus männlichen Teilnehmern. Um bei weiteren Rasch-Modell-Analysen auch das Geschlecht als Teilungskriterium heranziehen zu können, sollten in der Zukunft Stichproben mit ei-

ner ausgewogeneren Geschlechterverteilung angestrebt werden. Ferner wären Untersuchungen bei Personen aus unterschiedlichen Bildungs- und Karrierewegen wünschenswert, so dass die Einsatzmöglichkeiten des Verfahrens bestimmt bzw. erweitert werden können.

Zusätzlich zu weiteren Prüfungen der Skalierung anhand des Rasch-Modells bei neuen, repräsentativen Stichproben sind auch Normierungsuntersuchungen notwendig. Diese würden eindeutige Interpretationen der Testkennwerte ermöglichen. Bei der Erstellung von Normierungstabellen ist zu berücksichtigen, dass die erfasste Fähigkeit kulturgebunden ist, so dass von einer Verallgemeinerung der Ergebnisse für den Einsatz bei anderen Ländern als dem Untersuchungsland abzuraten ist. Da Erwachsene ab 18 Jahren im deutschsprachigen Raum den intendierten Geltungsbereich des Verfahrens darstellen, sind vor allem diesbezüglich weitere Validierungsuntersuchungen notwendig.

Die Konstruktion von weiteren, insbesondere schwierigen Items wäre für die Zukunft des Verfahrens auch wünschenswert. Die Erweiterung des Itempools könnte auf der einen Seite die Aufnahme von Wissensfragen aus anderen Themengebieten ermöglichen, und auf der anderen Seite zur Entwicklung einer Parallelversion bzw. mehrerer Testformen beitragen. Jedoch ist anzumerken, dass die Aufstellung von itemgenerierenden Regeln nicht möglich ist, so dass die Entwicklung weiterer Items durch das Angebot an figuralen Stimuli mit Bezug zum Allgemeinwissen stark beschränkt wird.

Zusammenfassend ist festzuhalten, dass im Rahmen dieser Arbeit ein gelungener Konstruktionsvorgang von Aufgaben vorgestellt wird, bei denen sinnvolle figurale Analogien als Arbeitsmaterial Anwendung finden. Für die entwickelten Items konnte die Geltung des Rasch-Modells präsentiert werden. Die Untersuchung der diskriminanten Validität durch den Vergleich mit *AN-TOP* lieferte auch zufriedenstellende Ergebnisse. *Figurale Analogien* stellt somit für die Erwachsenenendiagnostik einen neuartigen und ökonomischen Test zur Erfassung des kristallisierten schlussfolgernden Denkens im figuralen Bereich dar.

V. ZUSAMMENFASSUNG

In der vorliegenden Arbeit wurde ein neues Verfahren zur Erfassung des kristallisierten schlussfolgernden Denkens bei Erwachsenen entwickelt. Die theoretische Grundlage für *Figurale Analogien* bildeten hauptsächlich das Intelligenzmodell und die Investment-Theorie von Cattell (1963, 1987). Als Arbeitsmaterial wurden figurale Analogie-Aufgaben herangezogen, mit denen traditionellerweise fluide Intelligenz gemessen wird. Durch den sinnvollen Inhalt dieser Aufgaben wird der Einsatz von figuralen Analogien auch bei der Wissensdiagnostik ermöglicht.

Insgesamt wurden 27 Items mit ausschließlich figuralem Arbeitsmaterial konstruiert, deren Inhalt drei unterschiedliche Themengebiete des Allgemeinwissens widerspiegeln. Dabei ist der Einsatz des Multiple-Choice-Antwortformats „1 aus 5“ durch die Beschaffenheit des Arbeitsmaterials zu begründen, mit der einerseits die Anzahl plausibler Distraktoren beschränkt, und andererseits die Anwendung des freien Antwortformats verhindert wird.

An der empirischen Untersuchung nahmen 331 Auszubildende der Sicherheitsakademien der Polizei in Niederösterreich teil. Den TeilnehmerInnen wurden der neu entwickelte Test sowie *AN-TOP* (Kubinger & Heuberger, in Vorb.) und *Gleichungen* (s. vorläufig bei Gamsjäger, 2012) vorgegeben, welche zur Erfassung des logisch-schlussfolgernden Denkens dienen. Die durchschnittliche Bearbeitungszeit von *Figurale Analogien* in der Gruppe betrug dabei zehn Minuten.

Die testtheoretischen Analysen wurden anhand des dichotomen logistischen Testmodells von Rasch unternommen. Die Modellkontrollen erfolgten graphisch, sowie mit Hilfe des Likelihood-Ratio-Tests nach Andersen (1973) und des z-Tests nach Fischer und Scheiblechner (1970, zitiert nach Kubinger, Rasch, & Yanagida, 2011). Als Teilungskriterien wurden die Anzahl gelöster Aufgaben sowie das Alter und Bildungsniveau der Testpersonen herangezogen. Für alle 27 Items von *Figurale Analogien* zeigte sich eine a-priori Geltung des Rasch-Modells. Für *AN-TOP* wurde ein Rasch-Modell-konformer Itempool erst nach sukzessivem Ausscheiden von zwei Items erreicht. Somit ist davon

auszugehen, dass beide Tests das Gütekriterium Skalierung erfüllen und eindimensional messen.

Zur Bestimmung der diskriminanten Validität wurde die Korrelation zwischen den Personenparametern von *Figurale Analogien* und *AN-TOP* gerechnet. Dabei beträgt der Korrelationskoeffizient nach Pearson $r = 0,217$ und liegt unter dem theoretisch festgelegten Grenzwert.

In der Zukunft sollte die Geltung des Rasch-Modells bei den Items von *Figurale Analogien* anhand neuer Stichproben untersucht werden. Weitere Validierungsstudien und eine Erweiterung des Itempools wären auch wünschenswert.

VI. LITERATURVERZEICHNIS

- Ackerman, P. L. (1996). A theory of adult intellectual development: process, personality, interests, and knowledge. *Intelligence*, 22 (2), 227-257.
- Ackerman, P. L. (2000). Domain-specific knowledge as the „dark matter“ of adult intelligence: Gf/Gc, personality and interest correlates. *Journal of Gerontology: Psychological Sciences*, 55B (2), P69-P84.
- Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, 121 (2), 219-245.
- Amelang, M., & Bartussek, D. (1997). *Differentielle Psychologie und Persönlichkeitsforschung* (4. überarbeitete und erweiterte Auflage). Stuttgart: Kohlhammer.
- Amthauer, R. (1953). *Intelligenz-Struktur-Test (I-S-T)*. Göttingen: Hogrefe.
- Amthauer, R. (1970). *Intelligenz-Struktur-Test 70 (I-S-T 70)*. Göttingen: Hogrefe.
- Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)*. Göttingen: Hogrefe.
- Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, 38 (1), 123-140.
- Beauducel, A., Brocke, B., & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: facets for verbal, numerical and figural intelligence. *Personality and Individual Differences*, 30 (6), 977-994.
- Beauducel, A., & Süß, H-M. (2011). Wissensdiagnostik: Allgemeine und spezielle Wissenstests. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik* (Enzyklopädie der Psychologie, Themenbereich B, Serie II, Band 3) (S. 235-273). Göttingen: Hogrefe.
- Bühner, M., Ziegler, M., Krumm, S., & Schmidt-Atzert, L. (2006). Ist der I-S-T 2000 R Rasch-skalierbar? *Diagnostica*, 52 (3), 119-130.

- 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.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: Elsevier Science Publishers.
- Chamorro-Premuzic, T., Furnham, A., & Ackerman, P. L. (2006). Ability and personality correlates of general knowledge. *Personality and Individual Differences*, 41 (3), 419-429.
- de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31 (2), 105-113.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Hans Huber.
- Funke, J., & Vaterrodt-Plünnecke, B. (1998). *Was ist Intelligenz?*. München: Beck.
- Furnham, A., Christopher, A. N., Garwood, J., & Martin, G. N. (2007). Approaches to learning and the acquisition of general knowledge. *Personality and Individual Differences*, 43 (6), 1563-1571.
- Furnham, A., Swami, V., Arteché, A., & Chamorro-Premuzic, T. (2008). Cognitive ability, learning approaches and personality correlates of general knowledge. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 28 (4), 427-437.
- Gamsjäger, C. (2012). *Entwicklung des Tests „Gleichungen“ zur Erfassung von Reasoning*. Unveröffentlichte Diplomarbeit, Universität Wien, Wien.
- Gentile, J. R., Kessler, D. K., & Gentile, P. K. (1969). Process of solving analogy items. *Journal of Educational Psychology*, 60 (6), 494-502.

- Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of Typical Intellectual Engagement. *Journal of Educational Psychology*, 84 (4), 537-552.
- Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53, 267-293.
- Guilford, J. P. (1971). *The nature of human intelligence*. (International student edition). New York: McGraw-Hill.
- Guilford, J. P., & Hoepfner, R. (1976). *Analyse der Intelligenz*. Weinheim: Beltz.
- Heller, K. A., Kratzmeier, H., & Lengfelder, A. (1998a). *Matrizen Test Manual, Bd. 2. Ein Handbuch mit deutschen Normen zu den Standard Progressive Matrices*. Göttingen: Beltz Test GmbH.
- Heller, K. A., Kratzmeier, H., & Lengfelder, A. (1998b). *Matrizen Test Manual, Bd. 1. Ein Handbuch mit deutschen Normen zu den Advanced Progressive Matrices*. Göttingen: Beltz Test GmbH.
- Holling, H., Preckel, F., & Vock, M. (2004). *Intelligenzdiagnostik*. Göttingen: Hogrefe.
- Horn, W. (1983). *Leistungsprüfsystem (LPS)* (2. erweiterte und verbesserte Auflage). Göttingen: Hogrefe.
- 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.
- Hossiep, R., & Schulte, M. (2007). *Bochumer Wissenstest (BOWIT)*. Göttingen: Hogrefe.
- ITB (Institut für Test- und Begabungsforschung GmbH, Bonn), & Gittler, G. (2011). *Intelligenz-Basis-Funktionen (IBF) (Version 25 – Revision 2)*. Mödling: Schuhfried.
- Kersting, M., Althoff, K., & Jäger, A. O. (2008). *Wilde-Intelligenztest 2 (WIT-2)*. Göttingen: Hogrefe.

- Kubinger, K. D. (2003). Probabilistische Testtheorie. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 415-423). Weinheim: Beltz.
- Kubinger, K. D. (2005). Psychological test calibration using the rasch model - Some critical suggestions on traditional approaches. *International Journal of Testing*, 5 (4), 377-394.
- Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2. überarbeitete und erweiterte Auflage). Göttingen: Hogrefe.
- Kubinger, K. D., Rasch, D., & Yanagida, T. (2011). *Statistik in der Psychologie: Vom Einführungskurs bis zur Dissertation*. Göttingen: Hogrefe.
- Kvist, A. V., & Gustafsson J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence*, 36 (5), 422-436.
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Beltz.
- McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, 38 (1), 115-142.
- McArdle, J. J., Hamagami, F., Meredith, W., & Bradway, K. P. (2000). Modeling the dynamic hypotheses of Gf-Gc theory using longitudinal life-span data. *Learning and Individual Differences*, 12 (1), 53-79.
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2. Auflage). Berlin: Springer-Verlag.
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American Psychologist*, 67 (2), 130-159.
- Petermann, F. (Hrsg.). (2012). *WAIS-IV. Wechsler Adult Intelligence Scale – Revision IV*. Frankfurt am Main: Pearson Assessment.

- Petermann, F., & Petermann, U. (Hrsg.). (2011). *WISC-IV. Wechsler Intelligence Scale for Children-IV*. Frankfurt am Main: Pearson Assessment.
- Poinstingl, H., Mair, P., & Hatzinger, R. (2007). *Manual zum Softwarepackage eRm (extended Rasch modeling) Anwendung des Rasch-Modells (1-PL-Modell)*. Lengerich: Pabst Science Publishers.
- Rindermann, H., Flores-Mendoza, C., & Mansur-Alves, M. (2010). Reciprocal effects between fluid and crystallized intelligence and their dependence on parents' socio-economic status and education. *Learning and Individual Differences*, 20 (5), 544-548.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (Zweite, vollständig überarbeitete und erweiterte Auflage). Bern: Hans Huber.
- Spearman, C. (1904). "General Intelligence" objectively determined and measured. *The American Journal of Psychology*, 15 (2), 201-292.
- Spearman, C. (1914). The theory of two factors. *Psychological Review*, 21 (2), 101-115.
- Spearman, C. (1927). *The abilities of man*. London: Macmillan & Co.
- Spearman, C. (1939). Thurstone's work re-worked. *The Journal of Educational Psychology*, 30 (1), 1-16.
- Schiano, D. J., Cooper, L. A., Glaser, R., & Zhang, H. C. (1989). Highs are to lows as experts are to novices: Individual differences in the representation and solution of standardized figural analogies. *Human Performance*, 2 (4), 225-248.
- Schmidt-Atzert, L., & Rauch, W. (2008). TBS-TK Rezension: Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). 2., erweiterte und überarbeitete Auflage. *Report Psychologie*, 33 (6), 303-304.
- Schweizer, K., & Koch, W. (2002). A revision of Cattell's investment theory: cognitive properties influencing learning. *Learning and Individual Differences*, 13 (1), 57-82.

- Stemmler, G., Hagemann, D., Amelang, M., & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung* (7. vollständig überarbeitete Auflage). Stuttgart: Kohlhammer.
- Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, 84 (4), 353-378.
- Sternberg, R. J. (2003). Construct validity of the theory of successful intelligence. In R. J. Sternberg, J. Lautrey & T. I. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 55-77). Washington, D.C.: American Psychological Association.
- Sternberg, R. J., & Gardner M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General*, 112 (1), 80-116.
- Süß, H.-M. (2003). Culture fair. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 82-86). 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.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.
- Vernon, P. E. (1950). *The structure of human abilities*. London: Methuen & Co.
- Wechsler, D. (1961). Cognitive, Conative, and Non-Intellective Intelligence. In J. J. Jenkins & D. G. Paterson (Eds.), *Studies in individual differences. The search for intelligence* (pp. 651-660). London: Methuen & Co.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised (WAIS-R)*. New York: Psychological Corporation.
- Wagner-Menghin, M. (2004). *Lexikon-Wissen-Test (LEWITE)*. Mödling: Schuhfried.

VII. ANHANG

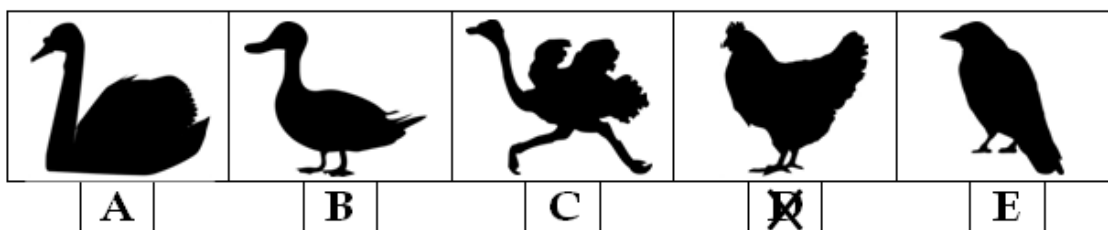
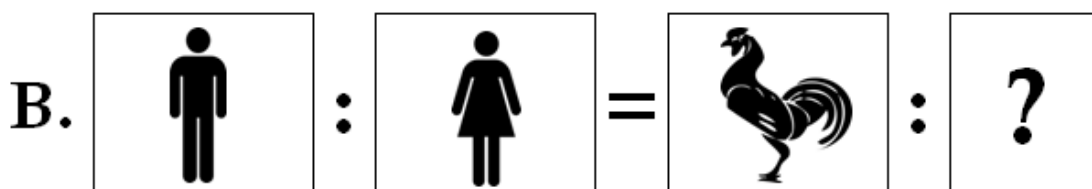
A. Instruktionsseite von *Figurale Analogien*

ALLGEMEINWISSEN

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

Sehen Sie sich 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, markieren Sie bitte Ihre Antwort durch Ankreuzen des Buchstabens unter dem Bild. Beachten Sie bitte, dass immer *nur eine Lösung* richtig ist.

Für die Testbearbeitung stehen Ihnen 18 Minuten zur Verfügung. Falls Sie vor Ablauf der Zeit mit der Bearbeitung der Aufgaben fertig sind, warten Sie bitte auf ein Zeichen des Testleiters.

Gibt es von Ihrer Seite noch Fragen?

Bitte noch nicht umblättern.

Warten Sie auf ein Zeichen des Testleiters.

B. Bei der empirischen Untersuchung beobachtete Bearbeitungszeiten für
Figurale Analogien

Bearbeitungszeit	Anzahl der Klassen
7 Minuten	1
8 Minuten 30 Sekunden	1
9 Minuten	4
9 Minuten 30 Sekunden	2
10 Minuten	1
11 Minuten 30 Sekunden	2
12 Minuten	1
12 Minuten 30 Sekunden	1
13 Minuten	1

C. Testdesign von AN-TOP (TH: Testheft)

Itemnummer																				
	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b	7a	7b	8a	8b	9a	9b	10a	10b
TH 1	1			2	3		4		5			6	7			8	9		10	
TH 2		1		2		3		4	5		6			7		8		9		10
TH 3	1		2			3	4			5	6		7		8			9	10	

	11a	11b	12a	12b	13a	13b	14a	14b	15a	15b	16a	16b	17a	17b	18a	18b	19a	19b	20a	20b
TH 1	11			12	13			14	15		16		17			18	19			20
TH 2	11		12			13		14		15		16	17		18			19		20
TH 3		11	12		13		14			15	16			17	18		19		20	

	21a	21b	22a	22b	23a	23b	24a	24b	25a	25b	26a	26b	27a	27b	28a	28b
TH 1	21		22		23			24	25			26	27		28	
TH 2		21		22	23		24			25		26		27		28
TH 3		21	22			23	24		25		26			27	28	

D. Ansuchen für die Datenerhebung

Serra Ünal

Klimschgasse 16/2/13
1030 Wien

Tel: +43 676 90 43 415

E-Mail: a0442724@unet.univie.ac.at

Bundesministerium für Inneres

Abt I/9 (SIK)

zH. Sg. Hr. Direktor Dr. Norbert Leitner

Bildungszentrum Wien

Marokkanergasse 4

1030 Wien

Wien, am 05.03.2012

Betreff: Ansuchen um eine Bewilligung für die Datenerhebung im Rahmen einer Diplomarbeit an den Ausbildungszentren der Bundespolizei

Ich studiere Psychologie an der Universität Wien und schreibe meine Diplomarbeit am Institut für Angewandte Psychologie: Gesundheit, Entwicklung und Förderung, Fakultät für Psychologie, unter der Leitung von Univ.-Prof. Dr. Mag. Klaus D. Kubinger.

In meiner Diplomarbeit beschäftige ich mich mit einem neu konstruierten psychologisch-diagnostischen Verfahren mit dem Arbeitstitel „Figurale Analogien“ zur Erfassung des kristallisierten schlussfolgernden Denkens („Allgemeinwissen“). Die Testpersonen sollen dabei aus Bildmaterial bestehende Analogien bearbeiten, um Rückschlüsse über ihren Wissensumfang in verschiedenen Themengebieten zu erhalten.

Untersuchungsablauf:

Es ist geplant, das psychologisch-diagnostische Verfahren in der Gruppe zu administrieren. Somit ist es möglich, mehrere Auszubildende zu einem Zeitpunkt zu testen. Die Auszubildenden müssten sich in einem von der Bundespolizei zur Verfügung gestellten Raum einfinden, in dem die Testungen stattfinden können (z.B. Klassenzimmer, Schulungsraum). Der Raum muss, abgesehen von ausreichend Tischen und Stühlen, über keine weitere Ausstattung verfügen.

Zusätzlich zu dem neu konstruierten Verfahren „Figurale Analogien“ wird der „AN-TOP“ (*Alpha Numeric Topologies*, Kubinger & Heuberger, in Vorb.) vorgegeben. Dieser erfasst das logisch-schlussfolgernde Denken und soll zur diskriminanten Validierung der „Figurale Analogien“ dienen.

Der dritte Teil der Testung besteht aus dem von Fr. Gamsjäger im Rahmen ihrer Diplomarbeit ebenfalls neu konstruiertem Verfahren „Gleichungen“. Dieses erfasst kristalline, numerische Intelligenz („Reasoning“).

Die Testung dauert insgesamt maximal sechzig Minuten.

Sobald eine offizielle Genehmigung für die Testungen vorliegt, werden Termine fixiert, an denen Testleiter(innen) zum Testen kommen dürfen. Bei den Terminvereinbarungen legen wir großen Wert darauf, den Ablauf des regulären Unterrichts so wenig wie möglich zu unterbrechen.

Als Testleiter(innen) fungieren neben mir selbst, Diplomanden(innen) des Arbeitsbereichs Psychologische Diagnostik der Fakultät für Psychologie, die sowohl über theoretische als auch praktische Erfahrung beim Testen verfügen.

Insgesamt sollen die Daten von 250 – 300 Personen erhoben werden, um entsprechende statistische Analysen vornehmen zu können.

Die gewonnenen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke verwendet. Jedem Auszubildenden wird zu Beginn der Testung ein Probanden-Code zugewiesen, um größtmögliche Anonymität zu gewährleisten.

Folgende Ausbildungszentren haben sich bereits zur Teilnahme an der Studie, vorbehaltlich der Bewilligung durch das Ministerium, bereit erklärt:

- Bildungszentrum Traiskirchen (GenMjr Mag. Arthur Reis)

Über eine positive Entscheidung Ihrerseits würde ich mich sehr freuen.

Mit der Bitte um Unterstützung und freundlichen Grüßen,



Serra Ünal

E. Beispiel eines Informationsblatts für Testpersonen und LehrerInnen

INFORMATION ZU DER TESTUNG AM 20. APRIL 2012

Liebe LehrerInnen und liebe Auszubildende!

Am 20. April finden in Ihrer Klasse in der 5. Stunde Testungen statt. Zu diesem Zweck kommen Psychologie-Studentinnen der Universität Wien in die Schule.

Worum geht's?

Die Studentinnen haben im Rahmen ihrer Diplomarbeit neue psychologisch-diagnostische Verfahren konstruiert, die sie nun an ca. 300 Personen erproben müssen.

Die Testung besteht aus drei aufeinanderfolgenden Teilen, mit denen logisch-schlussfolgerndes Denken in verschiedenen Bereichen sowie Allgemeinwissen erfasst werden soll, und dauert in etwa 60 Minuten.

Bringen Sie bitte ein Schreibgerät (zB. Kugelschreiber) zur Testung mit.

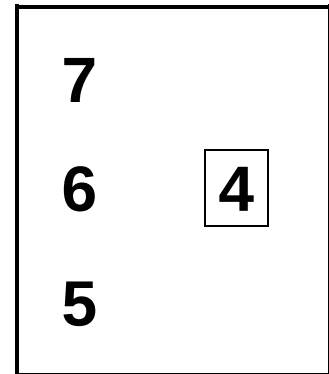
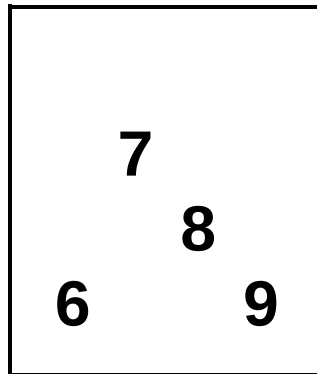
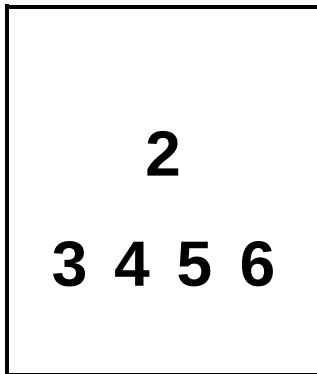
Was passiert damit?

Die Ergebnisse der Testungen haben keinen Einfluss auf Ihre schulischen Bewertungen. Die gewonnen Daten werden ausschließlich für wissenschaftliche Zwecke benutzt und sind anonymisiert.

Für weitere Fragen stehen Ihnen die Testleiterinnen gerne zur Verfügung.

F. Infolge der Rasch-Modell Analysen ausgeschlossene Items von AN-TOP

Item 11a



Item 20b

