



universität
wien

Diplomarbeit

Titel der Arbeit

Studie zur konvergenten Validität des
Untertests *Formale Folgerichtigkeit* des AID 3

Verfasserin

Carina Heuschober

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Mai 2012

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuerin: Mag. Dr. Stefana Holocher-Ertl

Abstract

In der vorliegenden Arbeit wird der von Hagenmüller (2011) im Zuge der Erweiterung der Intelligenztestbatterie AID 2 (Adaptives Intelligenzdiagnostikum 2, Version 2.2, Kubinger, 2009) entwickelte Untertest *Formale Folgerichtigkeit*, der der Erfassung von logisch-schlussfolgerndem Denken anhand figuraler Folgen bei Kindern und Jugendlichen im Alter von sechs bis 15 Jahren dienen soll, einer konvergenten Validierung unterzogen.

Zur Überprüfung der Validität des neuen Verfahrens wurden Reasoning-Untertests aus der *Grundintelligenztest Skala 1 (CFT 1, Cattell et al., 1997)* und dem *Prüfsystem für Schul- und Bildungsberatung* (Horn, 1969) in einer von Ecker (2011) adaptierten Version eingesetzt sowie Schulnoten in Deutsch und Mathematik erhoben. Um zudem differenziertere Aussagen bezüglich des Antwortformats treffen zu können, wurden der Test *Formale Folgerichtigkeit* und der adaptierte Untertest drei des *PSB* einerseits mit Multiple-Choice und andererseits mit freiem Antwortformat vorgegeben.

Die anhand einer Stichprobe von 84 Kindern einer Wiener Volksschule im Alter von sechs bis 10 Jahren gewonnenen Ergebnisse zeigen zufriedenstellende Korrelationen des Untertests *Formale Folgerichtigkeit* mit den oben erwähnten Reasoning-Verfahren, die für eine gelungene konvergente Validierung von *Formale Folgerichtigkeit* sprechen.

Abstract

In this study a convergent validation of the subtest *Formale Folgerichtigkeit*, which was developed by Hagenmüller (2011) in the course of the expansion of the intelligence test battery AID 2 (Adaptives Intelligenzdiagnostikum 2, Version 2.2, Kubinger, 2009) to measure figural reasoning in children and adolescents aged six to 15 years, is executed.

To check the validity of the new instrument several reasoning subtests from the *Grundintelligenztest Skala (CFT 1, Cattell et al., 1997)* and the *Prüfsystem für Schul- und Bildungsberatung (Horn, 1969)* in an adapted version of Ecker (2011) were used as well as school grades in German and Mathematics. To make more differentiated statements regarding the response format, *Formale Folgerichtigkeit* and the adapted subtest three of the *PSB* were applied both using a multiple-choice and a free response format.

The obtained results, based on a sample of 84 children of a Viennese school between the ages of six to 10 years, show satisfactory correlations of the subtest *Formale Folgerichtigkeit* with the reasoning tests mentioned above, that should provide an indication for a successful convergent validation of *Formale Folgerichtigkeit*.

Inhaltsverzeichnis

I. THEORETISCHER TEIL.....	7
1 Einleitung	7
2 Logisch-schlussfolgerndes Denken als Konstrukt.....	8
2.1 Der Begriff des schlussfolgernden Denkens.....	8
2.2 Induktion und Deduktion als wesentliche kognitive Komponenten des logischen Schlussfolgers.....	9
3 Logisches Schlussfolgern und der theoretische Bezug zum Intelligenzbegriff....	10
3.1 Spearman's Zwei-Faktoren-Theorie.....	10
3.2 Thurstones Primary Mental Abilities.....	11
3.3 Cattell & Horns Theorie der fluiden und kristallinen Intelligenz.....	11
3.4 Guilfords Würfelmodell.....	12
3.5 Jägers Berliner Intelligenzstrukturmodell.....	13
3.6 Carrolls Three Stratum Theory.....	14
4 Messung von logisch-schlussfolgerndem Denken bei Kindern und Jugendlichen anhand figural-anschaulicher psychologisch-diagnostischer Verfahren.....	15
4.1 Matrizentests.....	16
4.1.1 Matrizentests nach Raven.....	16
4.2 Grundintelligenztests nach Cattell.....	18
4.2.1 Grundintelligenztest Skala 1	18
4.2.2 Grundintelligenzskala 2	21
4.3 Der Untertest 3 des Prüfsystems für Schul- und Bildungsberatung.....	21
4.4 Der Untertest Formale Folgerichtigkeit	23

5	Das Gütekriterium Validität.....	25
5.1	Inhaltliche Gültigkeit	26
5.2	Konstruktvalidität	27
5.2.1	Die Faktorenanalyse.....	28
5.2.2	Die Mehr-Konstrukt-mehr-Methoden-Matrix.....	28
5.2.3	Nicht-korrelative Ansätze.....	28
5.3	Kriteriumsvalidität.....	29
5.4	Das vorliegende Validierungskonzept.....	30
II.	EMPIRISCHER TEIL.....	30
6	Ziel der Untersuchung.....	30
7	Fragestellungen und Hypothesen.....	31
7.1	Unterschiede zwischen freiem Antwortformat und Multiple-Choice Antwortformat..	31
7.2	Hypothesen zum Zusammenhang von Formale Folgerichtigkeit und Reasoning Verfahren	32
7.2.1	Hypothesen zum Zusammenhang von Formale Folgerichtigkeit und PSB.....	32
7.2.2	Hypothesen zum Zusammenhang von Formale Folgerichtigkeit und CFT 1.....	34
7.2.2.1	Untertest 3 – Klassifikationen.....	34
7.2.2.2	Untertest 4 – Ähnlichkeiten	35
7.2.2.3	Untertest 5 – Matrizen.....	35
7.2.2.4	Gesamtwert über die Untertests 3, 4 und 5.....	36
7.3	Hypothesen zum Zusammenhang von Formale Folgerichtigkeit und Schulnoten.....	36
7.3.1	Hypothesen zum Zusammenhang von Formale Folgerichtigkeit und der Deutschnote im Semesterzeugnis der 2. und 3. Klassen.....	37
7.3.2	Hypothesen zum Zusammenhang von Formale Folgerichtigkeit und der Gesamtnote im Semesterzeugnis der ersten Klasse.....	37

7.3.3	Hypothesen zum Zusammenhang von Formale Folgerichtigkeit und der Mathematiknote im Semesterzeugnis der 2. und 3. Klassen.....	38
8	Methode.....	39
8.1	Beschreibung der Untersuchungsinstrumente	39
8.2	Untersuchungsdurchführung	40
9	Ergebnisdarstellung und Interpretation.....	41
9.1	Deskriptive Statistik	41
9.1.1	Verteilung des Alters.....	41
9.1.2	Verteilung des Geschlechts.....	43
9.1.3	Verteilung der Noten	43
9.1.4	Verteilung der Testkennwerte der Reasoning-Verfahren	47
9.2	T-Test für abhängige Stichproben.....	49
9.3	Korrelationsanalyse.....	51
9.3.1	Korrelationen zwischen Formale Folgerichtigkeit und Reasoning Verfahren.....	52
9.3.1.1	Korrelationen zwischen Formale Folgerichtigkeit und PSB.....	52
9.3.1.2	Korrelationen zwischen Formale Folgerichtigkeit und CFT 1.....	53
9.3.2	Korrelationen zwischen Formale Folgerichtigkeit und Schulnoten.....	54
9.3.2.1	Korrelationen zwischen Formale Folgerichtigkeit und der Gesamtnote im Semesterzeugnis der ersten Klasse.....	54
9.3.2.2	Korrelationen zwischen Formale Folgerichtigkeit und der Deutschnote im Semesterzeugnis der 2. und 3. Klassen.....	55
9.3.2.3	Korrelationen zwischen Formale Folgerichtigkeit und der Mathematiknote im Semesterzeugnis der 2. und 3. Klassen.....	55
10	Zusammenfassung der Ergebnisse.....	56
11	Diskussion.....	57

12	Literaturverzeichnis.....	61
	III. ANHANG.....	64
13	Tabellenverzeichnis.....	64
14	Abbildungsverzeichnis.....	65
15	Lebenslauf.....	93

I. THEORETISCHER TEIL

1 Einleitung

Das Adaptive Intelligenzdiagnostikum 2 (AID 2, Version 2.2, Kubinger, 2009) ist im deutschen Sprachraum eine beliebte Intelligenztestbatterie für Kinder und Jugendliche im Alter von sechs bis 15 Jahren, um kognitive Stärken und Schwächen abzuklären.

Die Aktualisierung des AID 2 (Kubinger & Holocher-Ertl, in Vorb.) im Rahmen eines großen Forschungsprojektes in den Jahren 2009 bis 2011 hat zur Entwicklung eines neuen Individualverfahrens geführt, das anhand anschaulich - figuralen Aufgabenmaterials die Erfassung von logischem Schlussfolgern ermöglicht. Um das neue Reasoning – Verfahren *Formale Folgerichtigkeit* (Hagenmüller, 2011) empirisch dahingehend zu überprüfen, ob es tatsächlich das Konstrukt des logisch – schlussfolgernden Denkens erfasst, wird in der vorliegenden Arbeit nach einer theoretischen Auseinandersetzung mit dem Thema eine konvergente Validierung durchgeführt, zu deren Zweck Korrelationen zwischen Testwerten von *Formale Folgerichtigkeit* und Testwerten anderer psychologisch – diagnostischer Verfahren, die ebenfalls das Konstrukt des logischen Schlussfolgerns erfassen, sowie Schulnoten in Deutsch und Mathematik berechnet werden. Wenn die Zusammenhänge der Testwerte des neuen Verfahrens mit Testwerten anderer etablierter Tests zur Messung von figuralem Reasoning in einem respektablen Ausmaß ausfallen, kann dem Test zugesprochen werden, dass er tatsächlich das intendierte Konstrukt misst.

Die vorliegende Studie wurde in Kooperation mit Ecker (2011) durchgeführt, der sich im Rahmen seiner Diplomarbeit mit etwaigen Auswirkungen von Multiple - Choice und freiem Antwortformat auf das Antwortverhalten von Kindern zwischen sechs und neun Jahren bei Reasoningaufgaben beschäftigt hat. Zu diesem Zwecke entwickelte er Multiple – Choice Testversionen einerseits für den Test *Formale Folgerichtigkeit*, der von Hagenmüller (2011) in der Originalversion als Individualverfahren mit freiem Antwortformat konstruiert wurde und andererseits für den Untertest drei des *Prüfsystems für Schul- und Bildungsberatung* (Horn, 1969) zur Erfassung von

„Denkfähigkeit“, welche in der vorliegenden Validierungsstudie ebenfalls Anwendung finden.

2 Logisch-schlussfolgerndes Denken als Konstrukt

Das Konstrukt des logisch-schlussfolgernden Denkens ist in der psychologischen Fachliteratur sehr vielfältig definiert, wobei man auf eine Begriffswolke aus *logischem, schlussfolgerndem, formalem, induktivem, deduktivem* etc. Denken, Schlussfolgern, Schließen oder Reasoning stößt, die im Folgenden aufzuklären versucht wird.

2.1 Der Begriff des schlussfolgernden Denkens

Kubinger (2006, S. 198) leitet seine Definition aus dem Angebot einschlägiger Verfahren zur Messung von schlussfolgernden Denkens wie folgt ab: „Reasoning ist die Fähigkeit, Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge erkennen und zweckentsprechend verwerten zu können“.

Nach Beckmann und Guthke (1999, S. 6) kann „Denken (und synonym Schlussfolgern) – als zentrale intellektuelle Fähigkeit im menschlichen kognitiven Repertoire – (...) als interpretierende und ordnungstiftende Verarbeitung von Information charakterisiert werden“.

Thurstone & Thurstone (1941; zitiert nach Stemmler et al., 2011, S. 149) definieren den Primärfaktor „Inductive Reasoning“ als „schlussfolgerndes Denken im Sinne des Auffindens einer allgemeinen Regel in einer vorgegebenen Abfolge von Zahlen oder Symbolen und Anwendung derselben bei der Vorhersage des nächstfolgenden Elementes.“

Als Gemeinsames der vorgestellten Definitionen kann das Aufgliedern von logischem Schlussfolgern in zwei Komponenten herausgestrichen werden – einerseits in eine analysierende, eine Regelmäßigkeit erkennende, als Induktion oder induktives Denken

bezeichnete und andererseits in eine Anwendungskomponente, worunter deduktives Denken zu verstehen ist.

2.2 Induktion und Deduktion als wesentliche kognitive Komponenten des logischen Schlussfolgerns

Nach Waldmann und Weinert (1990, S. 25f) bezeichnet *induktives Denken* „Prozesse, die ein Individuum dazu befähigen, aus spezifischen Erfahrungen allgemeine Regeln, Muster, Konzepte oder Gesetzmäßigkeiten abzuleiten und auf neue Ereignisse anzuwenden“. In den Worten von Beckmann und Guthke (1999, S. 4) sind induktive Schlüsse dadurch charakterisiert, „dass der Prozess des Schließens von der Kenntnis einzelner Fakten oder Teilfakten zur Kenntnis einer allgemeinen Regel verläuft, in der die einzelnen Fakten oder Teilfakten zusammengefasst werden“. Und Klauer (2001, S. 165) versteht unter induktivem Denken zwar ebenfalls ein „Entdecken von Regelmäßigkeiten oder Gesetzmäßigkeiten“, jedoch beschränkt er das induktive Denken „auf das hier und jetzt Gegebene (...) ohne dass eine Verallgemeinerung auf alle Fälle notwendig dazu gehört“. Seine Definition orientiert sich an den psychischen Prozessen, die kennzeichnend für induktives Denken sind, wobei er den Fokus auf der Feststellung von Gleichheit und Verschiedenheit von Merkmalen und Relationen bei verbalem, bildhaftem, geometrisch-figuralem, numerischen oder einem sonstigen Material legt.

Da Einzelbeobachtungen den Ausgangspunkt für die Wissenserweiterung beim induktiven Schließen bilden, sind induktive Schlussfolgerungen stets mit einer gewissen Unsicherheit verbunden (vgl. Beckmann & Guthke, 1999; Klauer, 2011).

Das *deduktive Denken* stellt einen systematischen Denkprozess dar, der von bestimmten Prämissen unter Verwendung logischer Regeln zu bestimmten Schlussfolgerungen führt, wobei die logische Gültigkeit des Schlusses im Vordergrund steht, wodurch deduktive Schlussfolgerungen im Gegensatz zu induktiven zwar sicher werden, jedoch keine neuen Erkenntnisse bringen (vgl. Waldmann & Weinert, 1990; Beckmann & Guthke, 1999; Meiser & Klauer, 2001).

Eine Zusammenführung der Begriffe findet sich bei Beckmann und Guthke (1999, S. 27), die die erfolgreiche Anwendung der induktiv gewonnenen Regel auf das nächste Item oder das einzusetzende Element als eine Deduktionsleistung bezeichnen und betonen, dass „eine strenge Trennung von induktivem und deduktivem Denken (...) artifiziell ist“.

3 Logisches Schlussfolgern und der theoretische Bezug zum Intelligenzbegriff

Es besteht weitgehend Einigkeit darüber, *Intelligenz* als multidimensionales Konstrukt aufzufassen, welches sich aus verschiedenen Teilfähigkeiten zusammensetzt. Die Unterschiede der Modelle und Theorien der Intelligenz ergeben sich durch die jeweilige besondere Anordnung oder Struktur dieser Teilfähigkeiten (vgl. Preckel & Vock, 2010).

Im folgenden Abschnitt werden psychometrische Intelligenzmodelle vorgestellt, die ihren Fokus auf die Beschreibung einzelner unterscheidbarer Fähigkeiten, von denen hohe Generalität und zeitliche Stabilität angenommen werden, legen. Diese faktorenanalytisch gewonnenen Strukturtheorien der Intelligenz unterscheiden sich vor allem durch Anzahl der angenommenen Fähigkeiten und deren Beziehung zueinander (vgl. Süß, 2003).

3.1 Spearman's Zwei-Faktoren-Theorie

Spearman geht in seiner Theorie von einer allgemeinen Intelligenz („General Intelligence“, „g-Faktor“, „Generalfaktor“) und für jede Aufgabe spezifischen Intelligenzfaktoren („s-factors“) aus, die von der allgemeinen Intelligenz beeinflusst werden (Spearman, 1904, 1914; Süß, 2003; Sternberg, 2009; Lohaus, Maass & Vierhaus, 2010).

Spearman berichtete von hohen Korrelationen ($r = .68$ bis $.84$) zwischen Tests zum induktiven Denken und dem g-Faktor, und später bezeichnete er das Erkennen von Relationen und das Finden einer Regel neben „spatial discrimination“ als die zentrale

Anforderung von Tests mit hoher g-Faktorsättigung (Spearman, 1927, 1946; zitiert nach Beckmann & Guthke, 1999). Faktorenanalytische Untersuchungen der Raven-Matrizen tests, die ursprünglich als Verfahren mit hoher g-Faktorsättigung konstruiert wurden, zeigten neben dem Generalfaktor Beteiligungen von induktiven Denken, räumlichem Vorstellungsvermögen, Wahrnehmungsgenauigkeit und Gedächtnis (Conrad, 1983; Beckmann & Guthke, 1999).

3.2 Thurstones Primary Mental Abilities

Thurstone (1938, zitiert nach Süß, 2003) entwickelte ein Modell mit sieben unabhängigen Fähigkeiten, die er „Primary Mental Abilities“ titulierte. Nach empirischen Überprüfungen stellte sich jedoch heraus, dass diese sieben Faktoren nicht voneinander unabhängig sind, was zu einer Modifizierung des Modells führte, wonach von einem hierarchischen Modell ausgegangen wurde, das einen Faktor Allgemeine Intelligenz („g“) und auf der zweiten Ebene sieben Einzelfähigkeiten – verbales Verständnis („Verbal Comprehension“), Wortflüssigkeit bzw. Leichtigkeit der Wortfindung („Verbal Fluency“), Rechenfähigkeit („Number“), räumliches Vorstellungsvermögen („Spatial Visualization“), Merkfähigkeit bzw. Kurzzeitgedächtnis („Memory“), Wahrnehmungs- und Auffassungsgeschwindigkeit („Perceptual Speed“) und Regelerkennen bzw. schlussfolgerndes Denken („Inductive Reasoning“) - umfasst (vgl. Süß, 2003; Sternberg, 2009; Lohaus, Maass & Vierhaus, 2010).

Der Faktor „Inductive Reasoning“, der durch das Erkennen einer Regel im Testmaterial, wie beispielsweise in Figurenfolgen, geprüft wird, wurde später in „Reasoning“ umbenannt und umfasst neben Induktion nun auch Deduktion (Thurstone, 1938, 1941; zitiert nach Beckmann & Guthke, 1999).

3.3 Cattell & Horns Theorie der fluiden und kristallinen Intelligenz

Horn und Cattell (1966; Süß, 2003; Kubinger, 2006; Lohaus, Maass & Vierhaus, 2010) gingen in ihrer ursprünglichen Theorie von zwei Faktoren aus, der „fluiden“ (fluid general intelligence, gf) und der „kristallinen Intelligenz“ (auch „kristallisierte Intelligenz“, Crystallized general intelligence, gc), wobei die „fluide Intelligenz“ die generelle

Denkfähigkeit, Verarbeitungsgeschwindigkeit, schlussfolgerndes und spontanes Denken umfasst und unter anderem mit Testaufgaben zum induktiven und deduktiven Denken wie beispielsweise anhand der von Cattell entwickelten Culture-Fair Tests erhoben wird und unter „kristalliner Intelligenz“ Faktenwissen über die Welt, Wortschatz, Rechenfähigkeiten und andere wissensorientierte Informationen subsumiert werden.

Die fluide Intelligenz wird als weitgehend angeboren betrachtet und kaum durch äußere Einflüsse bestimmt, wohingegen die kristalline Intelligenz stark bildungs- und kulturabhängig ist (vgl. Lohaus, Maass & Vierhaus, 2010). Im Laufe der Entwicklung wächst die fluide Intelligenz schnell an, erreicht früh ihren Gipfel und nimmt danach ab, die kristalline Intelligenz wächst hingegen eher langsam und bleibt dann langfristig stabil (vgl. Schweizer, 2006b). Cattell (1987, zitiert nach Süß, 2003) beschrieb in der Investmenttheorie das Verhältnis von fluider und kristalliner Intelligenz so, dass die kristalline Intelligenz das Produkt einer Investition der fluiden Intelligenz in geeignete Lerngelegenheiten ist.

3.4 Guilfords Würfelmodell

Guilford (1967, zitiert nach Guilford & Hoepfner, 1971; Süß, 2003; Conrad, 1983; Schweizer, 2006) spricht hingegen nicht von einem generellen Intelligenzfaktor sondern nimmt in seinem „Structure of Intellect“ - Modell (SOI-Modell, siehe Abbildung 1) mehrere Facetten von Intelligenzleistungen an, nämlich die auslösenden *Inhalte*, die im Organismus ablaufenden *Operationen*, und die resultierenden *Produkte*, durch deren Kombination 120 voneinander unabhängige Intelligenzfaktoren unterschieden werden können, von denen etwa 100 als nachgewiesen gelten. Die Operationen das Konstrukt Reasoning betreffend finden sich nach Guilford & Hoepfner (1971) in den Kategorien „cognition“ („Erkenntnisvermögen“) und „convergent production“ wieder, wobei unter ersterem schnelles Entdecken, Bewusstheit, Wiedererkennen von Informationen, Verständnis oder Begreifen und unter zweiterem die Entwicklung logischer Schlussfolgerungen aus gegebenen Informationen mit Betonung auf dem Finden der einzigen bzw. besten Lösung (vgl. Stemmler et al., 2011).

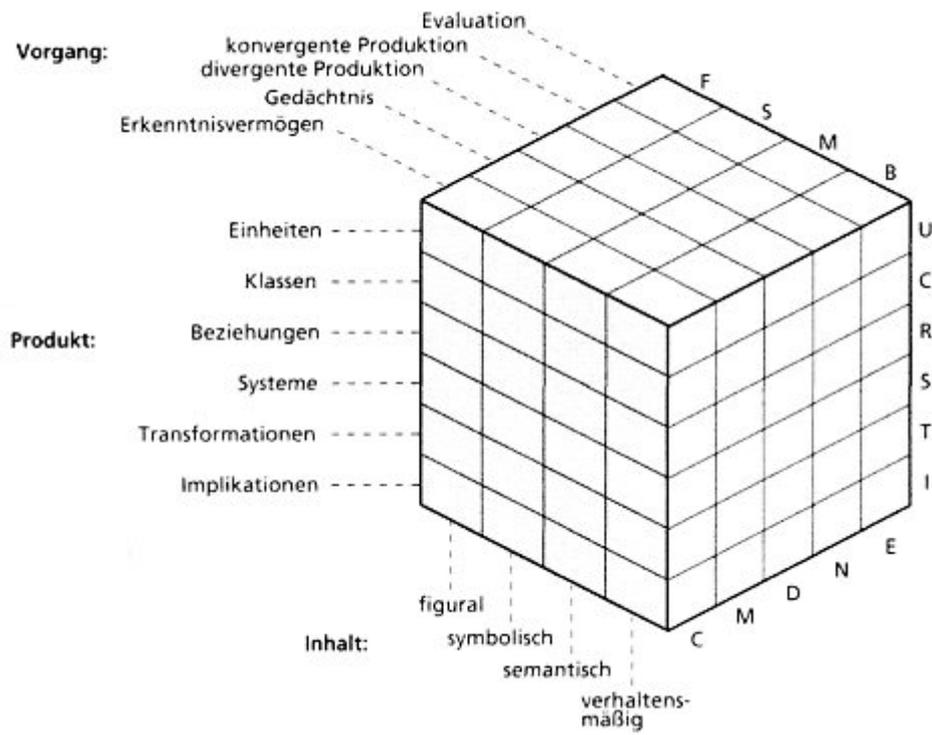


Abbildung 1: Guilfords Würfelmodell (Stemmler et al., 2011, S. 158)

3.5 Jägers Berliner Intelligenzstrukturmodell

Das Intelligenzstrukturmodell (siehe Abbildung 2) wurde von Jäger (1982, zitiert nach Süß, 2003) mit der Absicht entwickelt, die konkurrierenden Strukturmodelle der Intelligenz zu einem Gesamtmodell zu integrieren, wobei eine der Kernannahmen ist, dass an jeder Intelligenzleistung alle intellektuellen Fähigkeiten mit unterschiedlichem Gewicht beteiligt sind.

Die „Allgemeine Intelligenz“ setzt sich aus sieben darunter liegenden Fähigkeitskonstrukten zusammen, die zwei Facetten „Operationen“ („figural-bildhaft“, „verbal“, „numerisch“) und „Inhalten“ („Bearbeitungsgeschwindigkeit“, „Merkfähigkeit“, „Einfallsreichtum“, „Verarbeitungskapazität“) zugeordnet werden. Die

daraus resultierenden 12 Zellen können als spezifischere Fähigkeiten interpretiert werden (Süß, 2003). Schlussfolgerndes Denken wird durch den Operationsfaktor „Verarbeitungsgeschwindigkeit“ repräsentiert und kann anhand von drei Inhaltsebenen erfasst werden (Beckmann & Guthke, 1999).

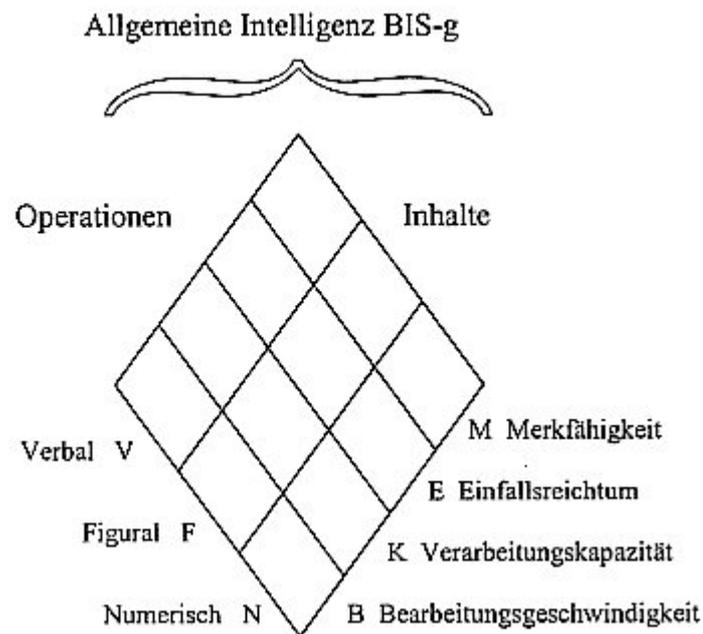


Abbildung 2: Jägers Intelligenzstrukturmodell (Preckel, 2003, S. 15)

3.6 Carrolls Three Stratum Theory

Carroll (1993, zitiert nach Süß, 2003; Sternberg, 2009; Lohaus, Maass & Vierhaus, 2010) unterscheidet in seiner „drei Schichten Theorie“ (siehe Abbildung 3) auf Ebene drei den Faktor „Allgemeine Intelligenz“, auf zweiter Ebene „Fluid Intelligence“, Crystallized Intelligence“, „General Memory and Learning“, „Broad Visual Perception“, „Broad Auditory Perception“, „Broad Retrieval Ability“, „Broad Cognitive Speediness“ und „Processing Speed“ und auf Schicht eins 68 Teilkonstrukte, unter denen auch induktives Denken, allgemeines schlussfolgerndes Denken und flexibles Schlussfolgern zu finden sind (Beckmann & Guthke, 1999). Nach Schweizer (2006b) erfährt dieses Modell bzw. seine Fortentwicklung zum CHC-Modell (Cattell, Horn, Carroll-Modell, McGrew, 1997; zitiert nach Schweizer, 2006b) momentan weltweit die größte Akzeptanz.

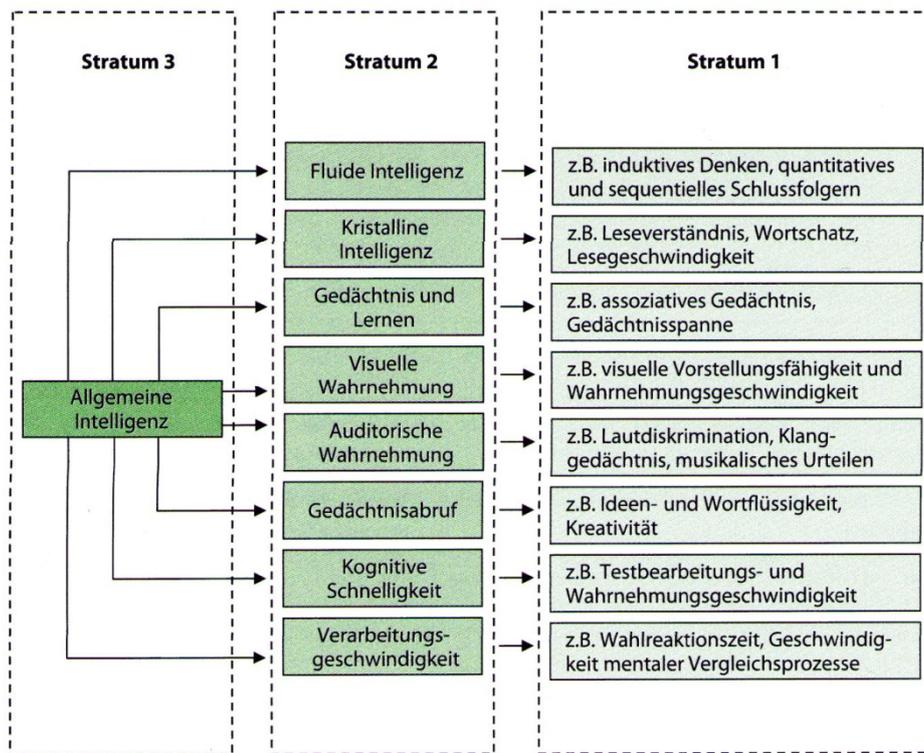


Abbildung 3: Carrolls Three Stratum Theory (Lohaus, Maass & Vierhaus, 2010, S. 123)

4 Messung von logisch-schlussfolgerndem Denken bei Kindern und Jugendlichen anhand figural-anschaulicher psychologisch-diagnostischer Verfahren

Psychologisch-diagnostische Verfahren zur Erhebung schlussfolgernden Denkens im figuralen Bereich (figurale Reasoning - Tests) enthalten meist Klassifikationsaufgaben, Reihenfortsetzungsaufgaben, Analogieaufgaben oder Matrizenaufgaben (Beckmann & Guthke, 1999; Preckel, 2003). Dieses Kapitel ist der beispielhaften Darstellung aktueller und klassischer Vertreter von Tests des interessierenden Konstrukts, die in der psychologisch-diagnostischen Begutachtung von Kindern und Jugendlichen Anwendung finden bzw. in der vorliegenden Studie eingesetzt wurden, gewidmet.

4.1 Matrizentests

Die Matrix als Aufgabentyp wurde erstmals von Spearman (1904, zitiert nach Preckel, 2003) in jener Absicht vorgestellt, einen starken Indikator für den g-Faktor zu erhalten, was von Jensen (1998, zitiert nach Preckel, 2003) in faktorenanalytischen Untersuchungen für die Matrizentests von Raven mit einer Ladung von .80 auf dem g-Faktor bestätigt werden konnte. Weiters wurden hohe Korrelationen mit anderen Maßen für die allgemeine Intelligenz wie auch für die fluide Intelligenz gefunden (z.B. Johanson et al., 2004, zitiert nach Schweizer, 2006b).

Figurale Matrizenaufgaben im engeren Sinne sind in einer Zeilen mal Spalten Anordnung präsentiert und bestehen aus einer sich aus der Kombination der Zeilen und Spalten ergebenden Menge an Zellen, in denen figurale Elemente in einer logischen Abfolge dargeboten werden. Bei diesen Aufgaben wird nun in den meisten Fällen die letzte Zelle frei gelassen, und der/die zu Testende muss beim Multiple-Choice Antwortformat die fehlende Figur aus einer gewissen Menge an Lösungsmöglichkeiten herausfinden (Preckel, 2003).

4.1.1 Matrizentests nach Raven

Die berühmten Matrizentests von Raven (1999, zitiert nach Rindermann & Kwiatkowski, 2010), namentlich *SPM* (*Standard Progressive Matrices*), *APM* (*Advanced Progressive Matrices*) und *CPM* (*Coloured Progressive Matrices*) sind mitunter die weltweit am häufigsten eingesetzten Intelligenztests. Sie erlauben eine Erhebung von schlussfolgerndem Denken anhand sprach- und zahlenfreier Aufgaben, die induktives Denken, die Fähigkeit zur Beziehungsstiftung, das Identifizieren abstrakter Beziehungen, das Erkennen von Regeln und in Analogien Denken erfordern (Rindermann & Kwiatkowski, 2010). In der Folge werden die drei Versionen der Progressiven Matrizentests nach Raven vorgestellt.

Die *Standard Progressive Matrices* (*SPM*, ursprünglich von Raven, 1936; in der deutschen Version von Heller, Kratzmeier & Lengfelder, 1998a; zitiert nach Preckel & Vock, 2010)

können ab einem Alter von sechs Jahren eingesetzt werden und gelten als mittelschwere Version der Raven's Progressive Matrices. Exemplarisch für die Progressiven Matrizen sei das Item C8 in Abbildung 4 dargestellt.

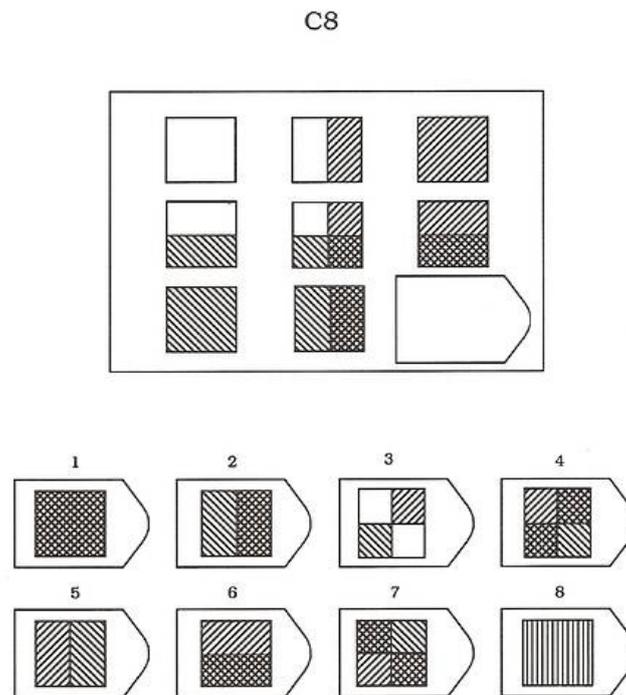


Abbildung 4: Matrizenaufgabe aus dem SPM (Werani, 2009, S. 7)

Die *Advanced Progressive Matrices* (APM, ursprünglich von Raven, 1962b; in der deutschen Version von Heller, Kratzmeier & Lengfelder, 1998b; zitiert nach Preckel & Vock, 2010) sind bei Jugendlichen ab 12 Jahren einsetzbar und umfassen die schwierigsten Matrizenaufgaben der Ravenfamilie, was sie zum Einsatz bei überdurchschnittlich befähigten Personen prädestiniert (Preckel, 2003).

Die leichteren *Coloured Progressive Matrices* (CPM, ursprünglich von Raven, 1962a; in der deutschen Version von Bulheller & Häcker, 2002; zitiert nach Preckel & Vock, 2010) sind für den Altersbereich von 3;9 bis sechs oder sieben Jahren vorgesehen und geistig retardierte Personen (Rindermann & Kwiatkowski, 2010).

Die Matrizentests von Raven können als Gruppen- oder Einzelverfahren eingesetzt werden und verfügen über ein Multiple-Choice Antwortformat. Allen Raventests können

gute Reliabilitätswerte ausgestellt werden (zwischen $r = .76$ und $.91$). Die Validität der Raven Tests betreffend kann den *Standard* und *Advanced Progressive Matrices* Kriteriumsvalidität durch Korrelationen mit Tests und Schulnoten, besonders der Mathematiknote, zugesprochen werden, wohingegen die Konstruktvalidität der *Coloured Progressive Matrices* als fraglich angesehen werden muss, da drei Faktoren gefunden wurden, jedoch nur die Berechnung eines Gesamtwertes vorgesehen ist (Preckel & Vock, 2010).

4.2 Grundintelligenztests nach Cattell

In Abschnitt 3.1.3 wurde bereits auf das Intelligenzmodell von Horn und Cattell (1966) eingegangen, das bekanntermaßen auf zwei Intelligenzfaktoren – der *fluiden* und der *kristallinen* Intelligenz – aufbaut. Zur Erhebung der fluiden Intelligenz, die nach Cattell für die Fähigkeit steht, neue Probleme und Situationen ohne Vorwissen erfolgreich bewältigen zu können, entwickelte dieser die sogenannten Grundintelligenztests – im Original als *Culture Fair Intelligence Scales* bekannt. Zur erfolgreichen Bearbeitung der Testaufgaben ist das Herstellen von Beziehungen, Regelerkennen, das schnelle Wahrnehmen von Merkmalen und das Bewältigen von Komplexität gefordert – Eigenschaften, die von kulturellen Erfahrungen und sozialen Randbedingungen weitgehend unabhängig sind. Insofern kann die Grundintelligenz als unverfälschtes Potenzial, das ein Kind kennzeichnet, verstanden werden (Schweizer, 2006b). Es folgt eine Erläuterung der Grundintelligenztests *CFT 1* und *CFT 20 - R* nach Cattell.

4.2.1 Grundintelligenztest Skala 1

Der *CFT 1* (Cattell, Weiß & Osterland, 1997) richtet sich an Kinder im Alter von fünf bis neun Jahren und gliedert sich in fünf Untertests mit figuralem Aufgabenmaterial (*Substitutionen, Labyrinth, Klassifikationen, Ähnlichkeiten* und *Matrizen*), von denen die ersten beiden die Wahrnehmungsgeschwindigkeit und die Untertests drei bis fünf das beziehungsstiftende Denken sowie das Erkennen von Gesetzmäßigkeiten – also Merkmale der fluiden Intelligenz oder Grundintelligenz erfassen. Die Aufgaben des *CFT 1* sind sehr ansprechend gestaltet und dürften auf Kinder durchaus motivierend wirken. In

der vorliegenden Studie kamen die Untertests drei bis fünf, die zusammen einen Testwert für das schlussfolgernde Denken liefern, zur Anwendung und werden nun unter Bezugnahme auf Cattells Intelligenzmodell vorgestellt.

Subtest 3 – Klassifikationen

Bei diesem Subtest muss das Kind eine Figur von fünf vorgegebenen Zeichnungen abgrenzen (siehe Abbildung 5). Nach Cattell und Mitarbeiter (1997) überprüft der Subtest drei das Klassifizieren von Merkmalen und das Herstellen von Beziehungen bei figuralanschaulichen Problemstellungen unterschiedlichen Komplexitätsgrades. Bezogen auf das zugrunde liegende Intelligenzmodell ist der Subtest *Klassifikationen* ein Vertreter des „General-Fluid-Ability-Faktors“ der sprachfreien Allgemeinintelligenz.

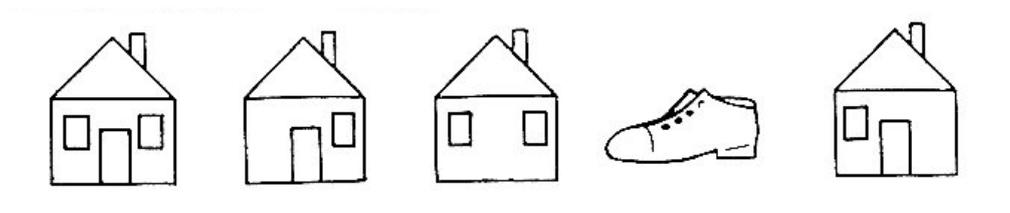


Abbildung 5: Beispielitem aus dem Subtest 3 *Klassifikationen* des CFT 1 (Cattell et al., 1997)

Subtest 4 – Ähnlichkeiten

Hier soll eine vorgegebene Zeichnung genau erfasst werden und unter fünf weiteren Zeichnungen, von denen vier detail- bzw. merkmalsverändert sind, herausgefunden werden (siehe Abbildung 6). Anhand des Subtests *Ähnlichkeiten* wird erfasst, bis zu welchem Komplexitätsgrad figurale Vorgaben wiedererkannt werden können, wenn diese Figuren im Kontext mit ähnlichen, merkmalsveränderten Figuren stehen. Cattell und Mitarbeiter (1997) gehen von mittleren Anteilen am „General-Fluid-Ability-Faktor“ aus.

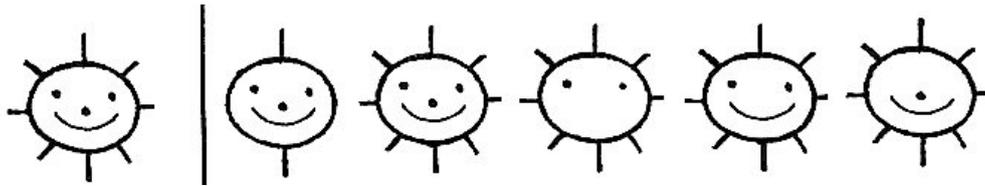


Abbildung 6: Beispielitem aus dem Subtest 4 *Ähnlichkeiten* des CFT 1 (Cattell et al., 1997)

Subtest 5 – Matrizen

Bei den Matrizenaufgaben im Subtest 5 soll von fünf figuralen Antwortmöglichkeiten diejenige ausgewählt werden, die eine zwei mal zwei - Matrix vervollständigt (siehe Abbildung 7). Anhand dieses Untertests wird jene Fähigkeit erfasst, Regeln und Zusammenhänge bei figuralen Problemstellungen zu erkennen, die nach Cattell, Weiß und Osterland (1997) hohe Anteile am „General-Fluid-Ability-Faktor“ hat.

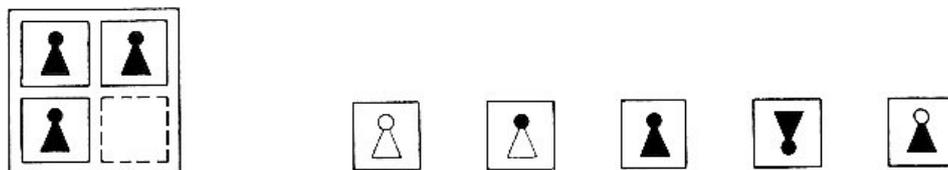


Abbildung 7: Beispielitem aus dem Subtest 5 *Matrizen* des CFT 1 (Cattell et al., 1997)

Die anhand der Subtests drei, vier und fünf erhobenen Faktoren können nach Cattell und Mitarbeiter (1997) als wesentliche Merkmale der Grundintelligenz - auch als allgemeine sprachfreie intellektuelle Leistungsfähigkeit umschrieben - angesehen werden, wobei die Anteile an der fluiden Intelligenz dominieren.

In der ersten Normierungsstudie von 1976 erreichte der Summenwert der Subtests drei bis fünf eine Split-half-Reliabilität zwischen .90 und .92 bezogen auf Kinder der 2. und 3. Klasse. Die Split-half-Reliabilität von .88 der revidierten Fassung wurde nur anhand einer kleinen und ausgelesenen Stichprobe gewonnen, wodurch Preckel und Vock (2010; Cattell et al., 1997) die Übertragbarkeit auf die Gesamtpopulation in Frage stellen.

Die faktorielle Validität des *CFT 1* kann als gegeben betrachtet werden und belegt die Annahme zweier separater Fähigkeiten. Die Kriteriumsvalidität wurde anhand hoher Zusammenhänge der Untertests drei bis fünf des *CFT 1* mit dem *HAWIK*-Gesamt-IQ ($r = .63$) und geringeren Zusammenhängen mit dem Verbalteil des *HAWIK* ($r = .46$) sowie anhand von Korrelationen mit Schulnoten, nämlich in Bezug auf die Deutschnote im Ausmaß von $r = .34$ bis $r = .43$ und der Mathematiknote zwischen $r = .42$ und $r = .48$, nach Meinung der Autoren belegt, wobei jeweils die Noten der Erstklässler am höchsten mit den Untertests drei bis fünf korrelieren (Cattell et al., 1997). Auf die im Rahmen der hier durchgeführten Studie gewonnenen Ergebnisse zu Reliabilität und Validität wird im empirischen Teil der Arbeit Bezug genommen.

4.2.2 Grundintelligenzskala 2

Der *CFT 20-R* (Weiß, 2008) kann bei Kindern ab achteinhalb Jahren eingesetzt werden und erfasst neben der fluiden Intelligenz anhand der Zusatzmodule *Wortschatztest* und *Zahlenfolgentest* auch Anteile der kristallinen Intelligenz. Der Test besteht wie der *CFT 1* aus figuralem Aufgabenmaterial und gliedert sich in vier Untertests (*Reihenfortsetzen, Klassifikationen, Matrizen und topologische Schlussfolgerungen*).

Die Reliabilität der *Grundintelligenzskala 2* ist nach Spearman-Brown mit Werten zwischen .91 und .96 als hoch einzuschätzen. Die Validität betreffend ergaben faktorielle Analysen drei Faktoren: Reasoning, Schlussfolgerndes Denken und Beziehungsstiftendes Denken. Korrelationen mit anderen Intelligenztests liegen zwischen .60 und .75 und mit der Mathematiknote bei .49 (Preckel & Vock, 2010).

Die *Grundintelligenztests* können als Einzel- oder Gruppenverfahren vorgegeben werden und verfügen über ein Multiple-Choice Antwortformat.

4.3 Der Untertest 3 des Prüfsystems für Schul- und Bildungsberatung

Das *Prüfsystem für Schul- und Bildungsberatung* ist für Personen zwischen neun und 20 Jahren konzipiert und wurde 1969 aus zehn Untertests des *Leistungsprüfsystems* (Horn,

1962) von Horn speziell für schulische Zwecke entwickelt. Den theoretischen Hintergrund bilden Thurstones Primärfaktoren, wobei der Faktor *Reasoning* neben dem Untertest vier anhand des im Rahmen der vorliegenden Untersuchung in einer adaptierten Version eingesetzten Untertests drei, der aus 40 figuralen Reihen besteht, für deren Bearbeitung fünf Minuten Zeit zur Verfügung stehen, erfasst werden soll. Wie in Abbildung 8 ersichtlich muss die Testperson hier kein fehlendes Symbol ergänzen sondern dasjenige durchstreichen, das nicht in die Reihe passt (Horn, 1969).

Zu den Gütekriterien werden im Manual des *PSB* keine genauen Angaben gemacht bzw. werden Werte aus dem *LPS* auf das *PSB* übertragen, obwohl einige der Untertests überarbeitet wurden, wie auch der für die vorliegende Studie adaptierte Untertest drei. Neben Validitätsdaten im Zusammenhang mit *Formale Folgerichtigkeit* (siehe Abschnitt 9.2.1.1) werden im Anhang auch Werte zur Reliabilität, die anhand der zugrundeliegenden Stichprobe gewonnen wurden, angeführt.

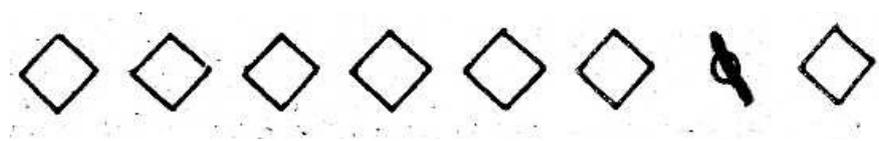


Abbildung 8: Originalitem des Untertests drei aus dem *PSB* (Horn, 1969)

Ecker (2011) adaptierte im Zuge seiner Diplomarbeit das *PSB* von Horn (1969) in eine Multiple-Choice und eine Freie-Antwort-Version, die auch in der vorliegenden Arbeit Anwendung finden. In Abbildung 9 ist das oben angeführte Original-Item des *PSB* in der adaptierten Version mit freiem Antwortformat und in Abbildung 10 mit Multiple-Choice-Antwortformat dargestellt.

Beispielaufgabe:

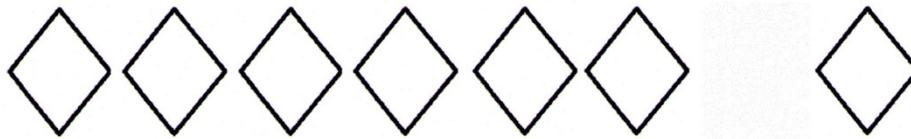


Abbildung 9: Beispielaufgabe aus der adaptierten Testversion des Untertest 3 aus dem PSB im freien Antwortformat (Ecker, 2011, S. 148)

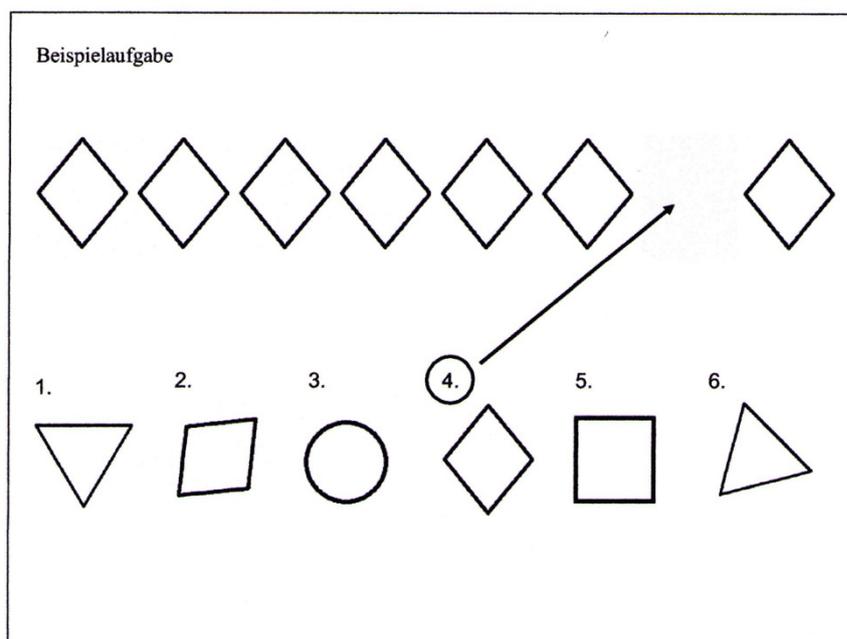


Abbildung 10: Beispielaufgabe aus der adaptierten Testversion des Untertest 3 aus dem PSB im Multiple-Choice-Antwortformat (Ecker, 2011, S. 156)

4.4 Der Untertest Formale Folgerichtigkeit

Der Test *Formale Folgerichtigkeit* (Hagenmüller, 2011) wurde in der Absicht entwickelt, die Intelligenztestbatterie AID 2 (Kubinger, 2009) um einen *Rasch*-Modell-konformen Subtest zur Erhebung von logisch-schlussfolgerndem Denken zu erweitern. Das Verfahren wurde als Einzeltest konzipiert und ist bei Kindern und Jugendlichen zwischen sechs und 15 Jahren einsetzbar, wobei das Testmaterial in zwei Testformen vorliegt und mit

zunehmendem Alter der Kinder und Jugendlichen schwierigere Aufgaben enthält. Das Kind bearbeitet ein Testheft mit 20 Items, bestehend aus Reihen von mehreren geometrischen Figuren in den Farben Gelb und Grün, wobei die letzte Figur fehlt, wie in Abbildung 11 zu sehen.



Abbildung 11: Beispielitem aus dem Testheft für 6 bis 7 Jahre aus dem Test *Formale Folgerichtigkeit* von Hagenmüller (2011, S. 35)

Das Kind soll aus 16 aus Holz gefertigten und gelb und grün lackierten geometrischen Steinen diejenige Figur heraussuchen, die die Reihe logisch ergänzt. In Abbildung 12 sind alle 16 Figuren, aus denen die richtige Antwort herausgesucht werden soll, in der von der Testautorin standardisierten Anordnung dargestellt. Sämtliche Items wurden auch aus diesen 16 geometrischen Figuren konstruiert.

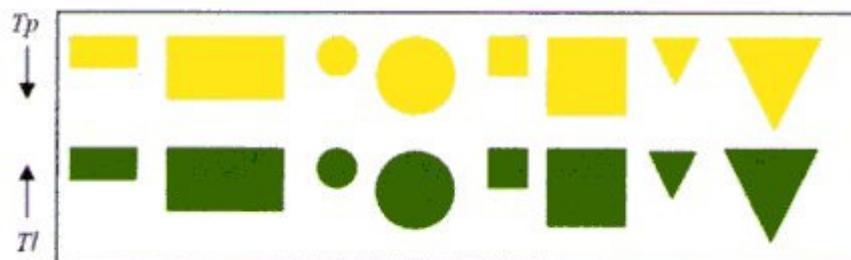


Abbildung 12: standardisierte Anordnung der Lösungssteine (Hagenmüller, 2011, S. 42)

In der vorliegenden Studie kam der Test *Formale Folgerichtigkeit* nicht nur in der soeben vorgestellten Originalversion von Hagenmüller (2011) zum Einsatz, sondern auch in einer von Ecker (2011) adaptierten Version. Der Test wurde dahingehend verändert, dass er als Papier - Bleistift Test mit einem Multiple - Choice Antwortformat in der Gruppe durchgeführt werden kann, wobei das Testmaterial aus Gründen der Ökonomie in Graustufen gehalten wurde (siehe Abbildung 13).

Die auf Grundlage der erhobenen Daten gewonnenen Ergebnisse die konvergente Validität betreffend werden in Abschnitt 9.2 vorgestellt, Aussagen zur Reliabilität finden sich im Anhang.

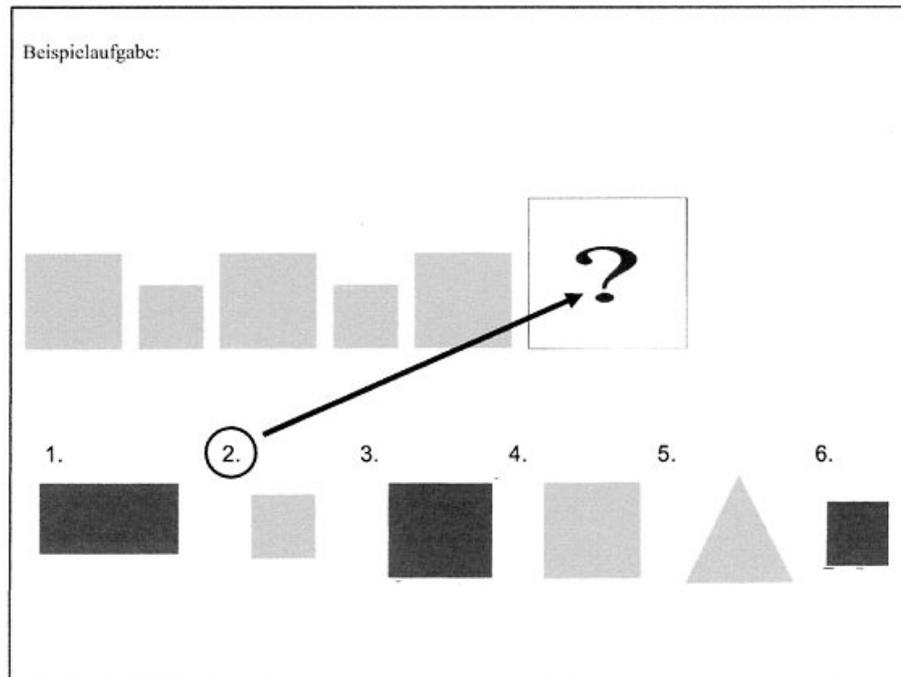


Abbildung 13: Beispielaufgabe der adaptierten Testversion des Untertest *Formale Folgerichtigkeit* (Ecker, 2011)

5 Das Gütekriterium Validität

Das wichtigste und gleichzeitig das empirisch am schwierigsten zu überprüfende Gütekriterium eines psychologisch-diagnostischen Verfahrens ist seine Validität (vgl. Kubinger, 2006; Rauchfleisch, 2005). Nach Lienert und Raatz (1998, S. 10) gibt „die Validität oder Gültigkeit eines Tests den Grad der Genauigkeit an, mit dem dieser Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen oder vorhersagen soll, tatsächlich mißt [sic] oder vorhersagt.“ Von Kubinger (2006, S. 50) auf den Punkt gebracht, ist ein Test dann valide, wenn „er tatsächlich jenes

Persönlichkeitsmerkmal misst, welches er zu messen behauptet.“ Beispielsweise gibt die Validität eines bestimmten Reasoning-Verfahrens darüber Auskunft, ob und in welchem Ausmaß dieses das Konstrukt Reasoning zu messen vermag.

Die *American Psychological Association* schlug 1954 (S. 13) in ihren „Technical Recommendations for Psychological Tests and Diagnostic Techniques“ vier Arten der Validität vor, „These four aspects of validity may be named content validity, predictive validity, concurrent validity and construct validity“. Fast 50 Jahre später führt Rauchfleisch (2005) in analoger Weise zur APA „die logische Validität (content validity), die Übereinstimmungsvalidität (concurrent validity), die Vorhersagevalidität (predictive validity) und die Konstruktvalidität (construct validity)“ an und spricht dabei von einer Einigung im Rahmen der internationalen Diskussion der Möglichkeiten von Validitätsbestimmungen.

Lienert und Raatz (1998) subsumieren die Übereinstimmungsvalidität und die Vorhersagevalidität unter die „kriterienbezogene Validität“, was auch schon 1955 (S. 281) Cronbach und Meehl vorschlugen, die sie als „*criterion-oriented validation procedures*“ bezeichneten. Ihnen gleich tut es Kubinger (2003) und spricht in analoger Weise von „inhaltlicher Gültigkeit“, „Konstruktvalidität“ und „Kriteriumsvalidität“, auf deren Konzeption und damit verbundenen Vorgehensweise, wie auch auf die der vorliegenden Untersuchung zugrundeliegenden konvergente Validität wird im Folgenden näher eingegangen.

5.1 Inhaltliche Gültigkeit

Von inhaltlicher Gültigkeit spricht man dann, wenn ein „Test bzw. seine Elemente (...) das zu erfassende Persönlichkeitsmerkmal oder die in Frage stehende Verhaltensweise repräsentieren“ (Lienert & Raatz, 1998, S. 10). Die Entscheidung darüber, ob ein Test inhaltlich valide ist, wird von „Experten“ getroffen, die im sogenannten „Expertenrating“ jedes einzelne Testitem dahingehend prüfen, „ob es in Bezug auf die gegebene operationale Definition dessen, was der Test messen soll, passt“ (Kubinger, 2006, S. 51). Als Beispiel für das Vorliegen von inhaltlicher Validität können Schulleistungstests

angeführt werden, bei denen LehrerInnen Testfragen auswählen, die das Stoffgebiet im jeweiligen Unterrichtsfach am besten repräsentieren. In jenem Fall spricht man von „logischer Validität“ (Kubinger, 2006; Lienert & Raatz, 1998). Hingegen ist die Rede von „trivialer Validität“, sobald es beispielsweise bei einem Einstellungsgespräch oder Probearbeiten darum geht, exemplarische Aufgaben aus dem späteren Tätigkeitsfeld auszuführen (Kubinger, 2006; Lienert & Raatz, 1998). Jedenfalls entscheiden in beiden Fällen „Experten“ durch die Auswahl der Aufgaben über das Vorliegen von inhaltlicher Gültigkeit.

5.2 Konstruktvalidität

Dieser Validierungsansatz betont das theoretische Konstrukt, das anhand eines psychologisch-diagnostischen Verfahrens erhoben werden soll. Folglich ist einem Test Konstruktvalidität zuzusprechen, wenn er gewisse theoretische bzw. theoriegeleitete Vorstellungen in Bezug auf ein bestimmtes Konstrukt erfüllt (Kubinger, 2006).

Ein (psychologisches) Konstrukt ist „ein nicht unmittelbar fassbarer Begriff, der sich auf nicht direkt beobachtbare Entitäten oder Eigenschaften bezieht. (...) Konstrukte werden aus einem theoretischen Zusammenhang heraus sowie mit Hilfe von beobachtbaren Ereignissen erschlossen“ (Häcker & Stapf, 2009, S. 536).

Bekannte Konstrukte in der Psychologie sind beispielsweise Angst, Interessen und nicht zuletzt Intelligenz und als Teilkomponente von intelligentem Verhalten das Konstrukt des *schlussfolgernden Denkens*. Solche nicht direkt beobachtbaren Merkmale versucht man anhand von beobachtbaren Ereignissen - nämlich der von Testpersonen in verschiedensten psychologisch-diagnostischen Verfahren gezeigten Ergebnisse - zu erschließen.

Die Berechnungsmethoden der Konstruktvalidität werden im Folgenden umrissen, wobei die Einteilung in „Faktorenanalyse“, „multitrait - multimethod“ und „nicht-korrelative Ansätze“ an Kubinger (2003, 2006) angelehnt ist.

5.2.1 Faktorenanalyse

Beim faktorenanalytischen Ansatz wird im Sinne einer konvergenten Validität „geprüft, ob die fragliche Testbatterie tatsächlich (exklusiv) diejenigen Faktoren abbildet, welche konstrukttheoretisch postuliert werden“ (Kubinger, 2003, S. 199).

5.2.2 Die Mehr-Konstrukt-mehr-Methoden-Matrix

Der von Campbell und Fiske (1959) vorgestellte Ansatz einer sogenannten “multitrait - multimethod Matrix“ (MTMM) prüft neben der konvergenten auch die diskriminante Validität eines Tests. Anhand einer solchen Matrix kann auch die Dimensionalität der interessierenden Eigenschaften in Bezug auf unterschiedliche Messmethoden bzw. -bedingungen bestimmt werden (Kubinger, 2003).

In diesem Zusammenhang sei auf das Begriffspaar *konvergente* und *diskriminante* Validität eingegangen, wobei unter konvergenter Validität verstanden wird, dass ein Test mit einem konstrukt nahen Außenkriterium hoch korrelieren soll und unter diskriminanter Validität eine niedrige Korrelation mit konstruktfernen Kriterien (Rost, 2004).

5.2.3 Nicht-korrelative Ansätze

Zu den nicht-korrelativen Ansätzen der Konstruktvalidierung sind einerseits die *Analyse interindividueller Unterschiede* zwischen Gruppen, die das fragliche Konstrukt in unterschiedlicher Ausprägung repräsentieren („Extremgruppenvalidierung“) zu zählen und andererseits die *Analyse von intraindividuellen Unterschieden*, die sich beispielsweise bei mehrmaliger Vorgabe eines Reasoning-Tests nicht ergeben sollten. Ebenfalls nicht korrelativ erfolgt die Konstruktvalidierung über *Selbst- bzw. Fremdbeobachtung* während der Bearbeitung eines Tests, wobei hier durch Beobachtung des anhand der Methode des sogenannten „lauten Denkens“ zugänglich gemachten Entscheidungs- und Lösungsprozesses untersucht wird, inwieweit das intendierte Konstrukt erfasst wird (Kubinger, 2003, 2006).

5.3 Kriteriumsvalidität

Lienert & Raatz (1998) betonen, dass die kriterienbezogene Validität historisch und praktisch gesehen das bedeutsamste Validitätskonzept darstellt und auch heute noch eine dominierende Rolle spielt. Kubinger (2006) bestätigt ebenfalls, dass der Nachweis der Kriteriumsvalidität der inhaltlichen Gültigkeit und der Konstruktvalidität überlegen ist und führt als Grund die statistische Kennzahl, die das Ausmaß der Validität eines Tests absolut zu bestimmen erlaubt, an. Die Kriteriumsvalidität, die anhand der Korrelation eines Tests mit einem als relevant erachteten Außenkriterium berechnet wird, liefert eine statistische Kennzahl in Form des Validitätskoeffizienten (Kubinger, 2003). Dieser gilt immer nur für ein bestimmtes Validitätskriterium und für die Population, aus der die Validierungsstichprobe stammt. Verschiedene Untersuchungen ergeben verschiedene Validitätskoeffizienten (Lienert & Raatz, 1998).

Durch die Wahl des Validitätskriteriums ergeben sich zwei verschiedene Arten der Kriteriumsvalidität. Es wird zwischen der *Übereinstimmungsvalidität* (konkurrente Validität), bei der als Kriterium ein anderer Test herangezogen wird, und der *Vorhersagevalidität* (prognostische Validität), wo das Validitätskriterium ein zukünftiges Ereignis darstellt, unterschieden (Kubinger, 2003; Lienert & Raatz, 1998).

Lienert und Raatz (1998, S. 228) heben in ihren Ausführungen zur Validität hervor, dass der Nachweis der Kriteriumsvalidität „mindestens so unentbehrlich ist wie die Aufklärung der psychologischen Faktoren, die hinter einem Test stehen“ und weiter dass erst das Vorliegen eben dieser den Diagnostiker berechtigt, „seine Methoden als brauchbar und anderen Verfahren als überlegen zu bezeichnen“. Das Ausmaß der Korrelation zwischen dem zu validierenden Test und einem konstruktnahen Außenkriterium betreffend, ist zu sagen, dass selbst bei Erfassung desselben Konstrukts die Zusammenhänge meist nur moderat ausfallen, so liegen sie z.B. bei *APM* und *CFT* um $r = .50$ (Heller, Kratzmeier & Lengfelder, 1998b; zitiert nach Preckel & Vock, 2010).

5.4 Das vorliegende Validierungskonzept

Das in der vorliegenden Studie zur Anwendung kommende konvergente Validierungskonzept basiert auf einer Korrelationsanalyse der Testwerte von *Formale Folgerichtigkeit* mit freiem Antwortformat (Hagenmüller, 2011) sowie Multiple – Choice Antwortformat (Ecker, 2011), des adaptierten Untertests des *Prüfsystems für Schul- und Bildungsberatung* in freiem und Multiple – Choice Antwortformat (Horn, 1969; Ecker, 2011), der Untertests *Klassifikationen, Ähnlichkeiten und Matrizen* der *Grundintelligenztest-Skala 1 (CFT 1, Cattell, Weiß & Osterland, 1997)* und der Schulnoten im Semesterzeugnis in den Fächern Deutsch und Mathematik.

II. EMPIRISCHER TEIL

6 Ziel der Untersuchung

In der vorliegenden Untersuchung wird überprüft, ob der von Hagenmüller (2011) im Zuge ihrer Diplomarbeit neu konstruierte Untertest *Formale Folgerichtigkeit* der in Überarbeitung und Erweiterung stehenden Intelligenz-Testbatterie AID 2 (Kubinger, 2009) das Konstrukt des schlussfolgernden Denkens zu erheben vermag. Zu diesem Zweck findet das Konzept der konvergenten Validierung seine Anwendung, wobei Korrelationen zwischen den von den Testpersonen im Untertest *Formale Folgerichtigkeit* erreichten Testergebnisse und konstruktnahen Außenkriterien berechnet werden, die erwartungsgemäß positiv ausfallen sollen. Als Außenkriterien fungieren einerseits die Untertests *Klassifikationen, Ähnlichkeiten und Matrizen* des CFT 1 (Cattell et al., 1997) und der Untertest drei einer adaptierten Version des PSB (Original von Horn, 1969; in der adaptierten Version von Ecker, 2011) und die Schulnoten in Deutsch und Mathematik.

Gelingt über die Korrelationsanalyse der Nachweis, dass ein Zusammenhang von einem angemessenen Ausmaß zwischen dem zu untersuchenden Test *Formale Folgerichtigkeit*, den weiteren bearbeiteten Reasoning-Verfahren und den Schulnoten besteht, so darf

dem vorliegenden Test Übereinstimmungs- bzw. Kriteriumsvalidität zugesprochen werden, was in weiterer Folge als empirischer Beleg für seine Daseinsberechtigung verstanden werden kann.

Im Folgenden werden die Fragestellung und Hypothesen näher beleuchtet, die Untersuchungsmethode inklusive der verwendeten Verfahren vorgestellt sowie die Durchführung der Studie beschrieben. Zum Schluss werden die gewonnenen Ergebnisse der durchgeführten Studie dargelegt.

7 Fragestellungen und Hypothesen

7.1 Unterschiede zwischen freiem Antwortformat und Multiple-Choice Antwortformat

Bevor eine Reihe von Hypothesen zum Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* und anderen psychologisch-diagnostischen Verfahren zur Erhebung von logisch-schlussfolgerndem Denken vorgestellt werden, wird zunächst die Frage geklärt, ob sich in den erhobenen Daten generelle Unterschiede in den Testleistungen in Abhängigkeit vom Antwortformat zeigen. In diese Berechnungen werden einerseits Testkennwerte des Tests *Formale Folgerichtigkeit* und andererseits der adaptierten Version des *PSB*, sowohl im freien als auch Multiple-Choice Antwortformat, Einzug halten, womit sich folgende Unterschiedshypothesen ergeben:

H₀a: Es gibt keinen signifikanten Unterschied in den Testergebnissen zwischen der Bearbeitung des Untertests *Formale Folgerichtigkeit* im freien Antwortformat und im Multiple – Choice Antwortformat.

H₁a: Es gibt einen signifikanten Unterschied in den Testergebnissen zwischen der Bearbeitung des Untertests *Formale Folgerichtigkeit* im freien Antwortformat und im Multiple – Choice Antwortformat.

H₀ b: Es gibt keinen signifikanten Unterschied in den Testergebnissen zwischen der Bearbeitung des *PSB* im freien Antwortformat und im Multiple – Choice Antwortformat.

H₁ b: Es gibt einen signifikanten Unterschied in den Testergebnissen zwischen der Bearbeitung des *PSB* im freien Antwortformat und im Multiple – Choice Antwortformat.

Da aufgrund theoretischer Überlegungen, die der Durchführung der vorliegenden Studie vorangegangen sind, und einer entsprechenden Auswahl der Untersuchungsinstrumente positive Zusammenhänge zwischen dem Untertest *Formale Folgerichtigkeit*, den weiteren Reasoning-Tests und den Schulnoten erwartet werden, erfolgte die Formulierung der weiteren Hypothesen (H₁) und deren statistische Überprüfung einseitig.

7.2 Hypothesen zum Zusammenhang von *Formale Folgerichtigkeit* und Reasoning Verfahren

Die Hauptfragestellung der vorliegenden Arbeit bezieht sich darauf, ob der neu entwickelte Untertest *Formale Folgerichtigkeit* von Hagenmüller (2011) in der Lage ist, das Konstrukt Reasoning in ähnlichem Maße zu erfassen wie andere anerkannte ausreichend validierte Reasoning-Verfahren. Hierfür werden Korrelationen zwischen den erzielten Ergebnissen der teilnehmenden Kinder im Test *Formale Folgerichtigkeit*, den Untertests *Klassifikationen*, *Ähnlichkeiten* und *Matrizen* des *CFT 1* und dem Untertest drei des *PSB* berechnet.

7.2.1 Hypothesen zum Zusammenhang von *Formale Folgerichtigkeit* und *PSB*

Da sowohl *Formale Folgerichtigkeit* als auch der *PSB* jeweils mit MC-Antwortformat und offenem Antwortformat bearbeitet wurden, ergeben sich bei der Frage nach dem

Zusammenhang zwischen *Formaler Folgerichtigkeit* und dem *PSB* vier Hypothesen bzw. Korrelationen, die nachfolgend angeführt werden.

H₀ a: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem *PSB* im freien Antwortformat.

H₁ a: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem *PSB* im Multiple-Choice Antwortformat.

H₀ b: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem *PSB* im freien Antwortformat.

H₁ b: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem *PSB* im freien Antwortformat.

H₀ c: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem *PSB* im Multiple-Choice Antwortformat.

H₁ c: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem *PSB* im Multiple-Choice Antwortformat.

H₀ d: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem *PSB* im Multiple-Choice Antwortformat.

H₁ d: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem PSB im Multiple-Choice Antwortformat.

7.2.2 Hypothesen zum Zusammenhang von *Formale Folgerichtigkeit* und CFT 1

Im Zusammenhang mit den Untertests *Klassifikationen*, *Ähnlichkeiten* und *Matrizen* und dem Gesamtwert über diese drei Untertests des CFT 1 ergeben sich acht Hypothesen, wobei eine differenzierte Betrachtung von *Formaler Folgerichtigkeit* nach freiem und MC-Antwortformat erfolgt.

7.2.2.1 Untertest 3 – *Klassifikationen*

H₀ a: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem Untertest 3 *Klassifikationen* des CFT 1.

H₁ a: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem Untertest 3 *Klassifikationen* des CFT 1.

H₀ b: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem Untertest 3 *Klassifikationen* des CFT 1.

H₁ b: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem Untertest 3 *Klassifikationen* des CFT 1.

7.2.2.2 Untertest 4 - Ähnlichkeiten

H₀ a: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem Untertest 4 *Ähnlichkeiten* des CFT 1.

H₁ a: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem Untertest 4 *Ähnlichkeiten* des CFT 1.

H₀ b: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem Untertest 4 *Ähnlichkeiten* des CFT 1.

H₁ b: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem Untertest 4 *Ähnlichkeiten* des CFT 1.

7.2.2.3 Untertest 5 – Matrizen

H₀ a: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem Untertest 5 *Matrizen* des CFT 1.

H₁ a: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem Untertest 5 *Matrizen* des CFT 1.

H₀ b: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale*

Folgerichtigkeit mit Multiple-Choice Antwortformat und dem Untertest 5 Matrizen des CFT 1.

H₁ b: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem Untertest 5 Matrizen des CFT 1.

7.2.2.4 Gesamtwert über die Untertests *Klassifikationen, Ähnlichkeiten und Matrizen* des CFT 1

H₀ a: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem Gesamtwert über die Untertests *Klassifikationen, Ähnlichkeiten und Matrizen* des CFT 1.

H₁ a: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit freiem Antwortformat und dem Gesamtwert über die Untertests *Klassifikationen, Ähnlichkeiten und Matrizen* des CFT 1.

H₀ b: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem Gesamtwert über die Untertests *Klassifikationen, Ähnlichkeiten und Matrizen* des CFT 1.

H₁ b: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* mit Multiple-Choice Antwortformat und dem Gesamtwert über die Untertests *Klassifikationen, Ähnlichkeiten und Matrizen* des CFT 1.

7.3 Hypothesen zum Zusammenhang von *Formale Folgerichtigkeit* und Schulnoten

Als weitere Außenkriterien im Sinne einer Kriteriumsvalidierung (eventuell sogar einer prognostischen Validierung) wurden die Noten der Kinder im Semesterzeugnis, das wenige Tage nach Abschluss der Untersuchung ausgehändigt wurde, herangezogen. Da

das Zeugnis der ersten Klasse nur in einer Gesamtnote über alle Fächer besteht, wurde hier nur dieser Wert und in der zweiten und dritten Klasse die Semesternoten in Deutsch und Mathematik in die Berechnungen einbezogen. Die Sachunterrichtnoten wurden aus der Analyse ausgeschlossen, da sie über alle SchülerInnen hinweg „Sehr gut“ waren.

7.3.1 Hypothesen zum Zusammenhang von *Formale Folgerichtigkeit* und der Gesamtnote im Semesterzeugnis der ersten Klasse

H₀ a: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im freien Antwortformat und der Gesamtnote im Semesterzeugnis bei der ersten Klasse.

H₁ a: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im freien Antwortformat und der Gesamtnote im Semesterzeugnis bei der ersten Klasse.

H₀ b: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im Multiple-Choice Antwortformat und der Gesamtnote im Semesterzeugnis bei der ersten Klasse.

H₁ b: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im Multiple-Choice Antwortformat und der Gesamtnote im Semesterzeugnis bei der ersten Klasse.

7.3.2 Hypothesen zum Zusammenhang von *Formale Folgerichtigkeit* und der Deutschnote im Semesterzeugnis der 2. und 3. Klassen

H₀ a: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale*

Folgerichtigkeit im freien Antwortformat und der Deutschnote im Semesterzeugnis bei den 2. und 3. Klassen.

H₁ a: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im freien Antwortformat und der Deutschnote im Semesterzeugnis bei den 2. und 3. Klassen.

H₀ b: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im Multiple-Choice Antwortformat und der Deutschnote im Semesterzeugnis bei den 2. und 3. Klassen.

H₁ b: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im Multiple-Choice Antwortformat und der Deutschnote im Semesterzeugnis bei den 2. und 3. Klassen.

7.3.3 Hypothesen zum Zusammenhang von *Formale Folgerichtigkeit* und der Mathematiknote im Semesterzeugnis der 2. und 3. Klassen

H₀ a: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im freien Antwortformat und der Mathematiknote im Semesterzeugnis bei den 2. und 3. Klassen.

H₁ a: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im freien Antwortformat und der Mathematiknote im Semesterzeugnis bei den 2. und 3. Klassen.

H₀ b: Es besteht kein signifikanter Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im Multiple-Choice Antwortformat und der Mathematiknote im Semesterzeugnis bei den 2. und 3. Klassen.

H₁ b: Es besteht ein signifikanter positiver Zusammenhang zwischen dem Untertest *Formale Folgerichtigkeit* im Multiple-Choice Antwortformat und der Mathematiknote im Semesterzeugnis bei den 2. und 3. Klassen.

8 Methode

8.1 Beschreibung der Untersuchungsinstrumente

Für die folgende Untersuchung kommt der zu validierende Untertest *Formale Folgerichtigkeit*, der im Zuge der Neubearbeitung und Erweiterung des AID 2 (Kubinger, 2009) von Hagenmüller (2011) im Rahmen ihrer Diplomarbeit entwickelt wurde, zum Einsatz. Die Vorgabe erfolgt einerseits in der Originalversion von Hagenmüller mit offenem Antwortformat und andererseits in einer von Ecker (2011) im Zuge dessen Diplomarbeit entwickelten Multiple-Choice Version. Für weitere Details sei auf Abschnitt 4.3 verwiesen.

Bei den Verfahren, anhand derer die konvergente Validierung des Untertests *Formale Folgerichtigkeit* durchgeführt wird, handelt es sich um eine adaptierte Version des Untertests drei des *Prüfsystems für Schul- und Bildungsberatung* (Horn, 1969; Ecker, 2011), der gleichfalls sowohl mit freiem als auch mit Multiple-Choice Antwortformat vorgegeben wird (siehe Abschnitt 4.4) und schließlich um die Untertests *Klassifikationen, Ähnlichkeiten* und *Matrizen* aus der *Grundintelligenztest Skala CFT 1* von Cattell, Weiß und Osterland (1997), die in der Originalversion mit Multiple-Choice Antwortformat zur Bearbeitung vorgelegt werden. Eine genaue Testbeschreibung des CFT 1 findet sich in Abschnitt 4.2.1.

Als Außenkriterien im Sinne einer (prognostischen) Kriteriumsvalidität wurden weiters die Gesamtnoten der Kinder der ersten Klasse, sowie die Deutsch- und Mathematiknoten der Zweit- und DrittklässlerInnen im Semesterzeugnis erhoben, wie auch Alter und Geschlecht als demografische Variablen.

8.2 Untersuchungsdurchführung

Die im Rahmen der vorliegenden Studie durchgeführte Testung erfolgte an der Wiener Volksschule St. Ursula, wobei sich die Klassenvorstände einer ersten, einer zweiten und von zwei dritten Klassen zur Teilnahme bereiterklärten. Bevor mit der Datenerhebung begonnen wurde, wurde die Studie in einem Bericht dem Stadtschulrat für Wien vorgestellt, der die Genehmigung zur Durchführung erteilte (siehe Anhang). Den KlassenlehrerInnen und Eltern der Kinder wurde eine Beschreibung des Projekts ausgehändigt und von den teilnehmenden Kindern eine schriftliche Teilnahmebestätigung der Eltern eingehoben.

Die Durchführung des Projekts erfolgte in Kooperation mit Johann Ecker (2011), der im Rahmen seiner Diplomarbeit Unterschiede zwischen dem freien und dem Multiple-Choice Antwortformat bei Kindern beleuchtete und im Zuge dessen die Multiple-Choice Versionen des Untertests *Formale Folgerichtigkeit* und des Untertests drei des *PSB* entwickelt hatte, die auch im Rahmen der vorliegenden Untersuchung Anwendung fanden.

Im Hinblick auf etwaige Unterschiede zwischen den Antwortformaten wurden die Verfahren einmal im freien Antwortformat und einmal im Multiple-Choice Format vorgegeben, wobei ein Mindestzeitraum von zwei Wochen zwischen den Erhebungen liegen sollte, um eventuellen Lern- und Erinnerungseffekten Einhalt zu gebieten. So wurden die teilnehmenden Klassen in zwei Hälften geteilt, wobei eine Hälfte zuerst die Multiple-Choice Tests in einem Gruppensetting bearbeitete und die andere Hälfte zuerst die Tests im freien Antwortformat im Zuge von Einzeltestungen. Bei einem zweiten Termin bearbeiteten die Kinder, die beim ersten Termin mit den Multiple-Choice-Verfahren getestet wurden, die Tests mit freiem Antwortformat und umgekehrt. Der *CFT 1* wurde von den Kindern lediglich einmal in seiner Originalversion im Rahmen von Gruppentestungen im Klassenverband bearbeitet. Als TestleiterInnen fungierten sowohl bei den Einzel- als auch bei den Gruppentestungen zuvor geschulte und instruierte DiplomandInnen des Arbeitsbereichs Psychologische Diagnostik der Fakultät für Psychologie, die die Kinder von der Klasse abholten und nach der Testung wieder

zurückbrachten. Die Klassen wurden für die Gruppentestung wiederum in zwei Hälften geteilt und bearbeiteten in getrennten Räumlichkeiten parallel die Multiple-Choice Verfahren, was zu einem möglichst ungestörten Ablauf, einer besseren Handhabbarkeit der Kinder, einem ruhigeren Arbeitsklima und besserer Konzentration der Kinder beitragen sollte. Der benötigte Zeitrahmen betrug bei den Gruppentestungen ca. eine Schulstunde und den Einzeltestungen ca. 30 Minuten, was durchaus auf moderate Akzeptanz seitens der Kinder und der LehrerInnen stieß.

9 Ergebnisdarstellung und Interpretation

Die Analyse des Datenmaterials wurde anhand der Statistiksoftware SPSS für Windows in der Version 14.0 durchgeführt.

9.1 Deskriptive Statistik

9.1.1 Verteilung des Alters

An der vorliegenden Studie haben insgesamt 84 Schülerinnen und Schüler im Alter von 6; 1 bis 9; 9 Jahren teilgenommen, wobei die grafische Darstellung des Alters in Abbildung 14 eine linksschiefe Verteilung zeigt und daher anstatt des Mittelwerts der Median mit einem Dezimalwert von 8,29, was einem Alter von 8 Jahren und fast vier Monaten entspricht, Anwendung findet.

Tabelle 1: Deskriptive Statistik des Alters der teilnehmenden Kinder (n=84)

Alter der Kinder		
N	Gültig	84
	Fehlend	0
	Mittelwert	7,9833
	Median	8,2900
	Minimum	6,08
	Maximum	9,75

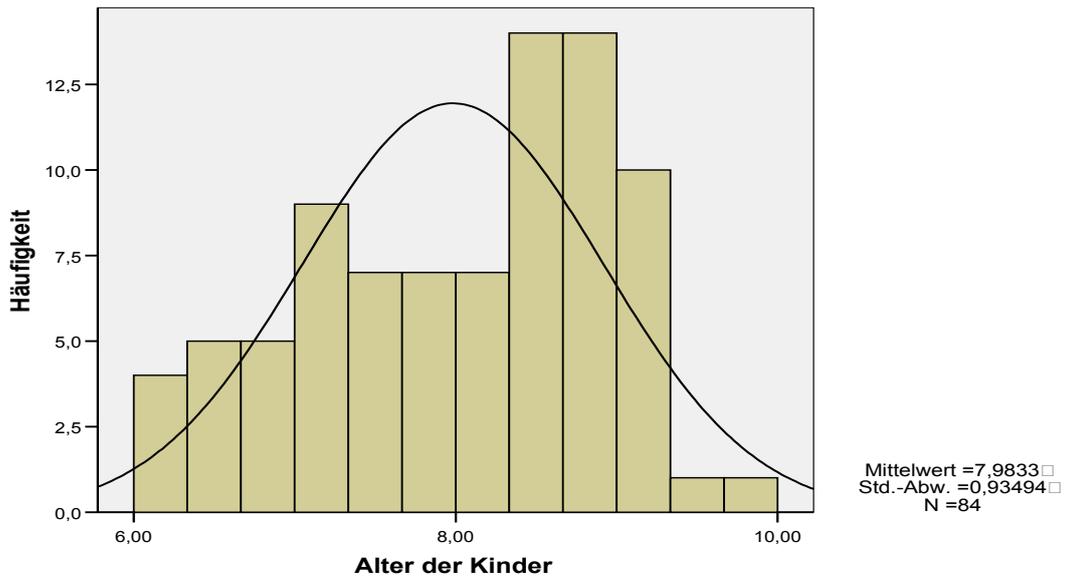


Abbildung 14: Verteilung des Alters der teilnehmenden Kinder (n=84)

Auch die statistische Überprüfung des Alters auf Normalverteilung anhand des Kolmogorov-Smirnov-Anpassungstest zeigt, dass das Alter in der vorliegenden Studie nicht normalverteilt ist (siehe Tabelle 1). Dieser Umstand lässt sich dadurch erklären, dass neben einer ersten Klasse und einer zweiten Klasse zwei dritte Klassen an der Untersuchung teilgenommen haben.

Tabelle 2: SPSS-Output: Überprüfung der Normalverteilung des Alters anhand des Kolmogorov-Smirnov-Anpassungstest

Kolmogorov-Smirnov-Anpassungstest

		Alter der Kinder
N		84
Parameter der Normalverteilung(a,b)	Mittelwert	7,9833
	Standardabweichung	,93494
Extremste Differenzen	Absolut	,156
	Positiv	,071
	Negativ	-,156
Kolmogorov-Smirnov-Z		1,429
Asymptotische Signifikanz (2-seitig)		,034

a Die zu testende Verteilung ist eine Normalverteilung. b Aus den Daten berechnet.

9.1.2 Verteilung des Geschlechts

In Bezug auf das Geschlecht zeigt die Stichprobe mit 46 zu 38 ein Übergewicht auf der männlichen Seite der UntersuchungsteilnehmerInnen. In Tabelle 3 ist die Anzahl der UntersuchungsteilnehmerInnen je Klasse getrennt nach Geschlecht wiedergegeben.

Tabelle 3: Verteilung des Geschlechts der VersuchsteilnehmerInnen ($n=84$)

Klasse	Geschlecht		Gesamt
	Weiblich	Männlich	
1a	10	13	23
2a	10	9	19
3a	10	12	22
3b	8	12	20
Gesamt	38	46	84

9.1.3 Verteilung der Noten

Neben Alter und Geschlecht der Kinder wurden auch die Schulnoten in den Fächern Deutsch und Mathematik bzw. im Falle der Erstklässler die Gesamtnote des Semesterzeugnisses erhoben. Wie den Tabellen vier und fünf und den Abbildungen 15 und 16 entnommen werden kann, ist die Verteilung der Noten besonders bei den Zweit- und Drittklässlern insofern sehr schief, dass die überwiegende Mehrheit der Kinder sowohl in Deutsch als auch in Mathematik ein Sehr gut im Semesterzeugnis hat und nur wenige übrige Kinder die Note Gut.

Tabelle 4: Verteilung der Deutschnote in der 2. und 3. Klasse (n=60)

Klasse	Deutschnote	
	Sehr gut	Gut
2a	16	3
3a	19	3
3b	17	2
Gesamt	52	8

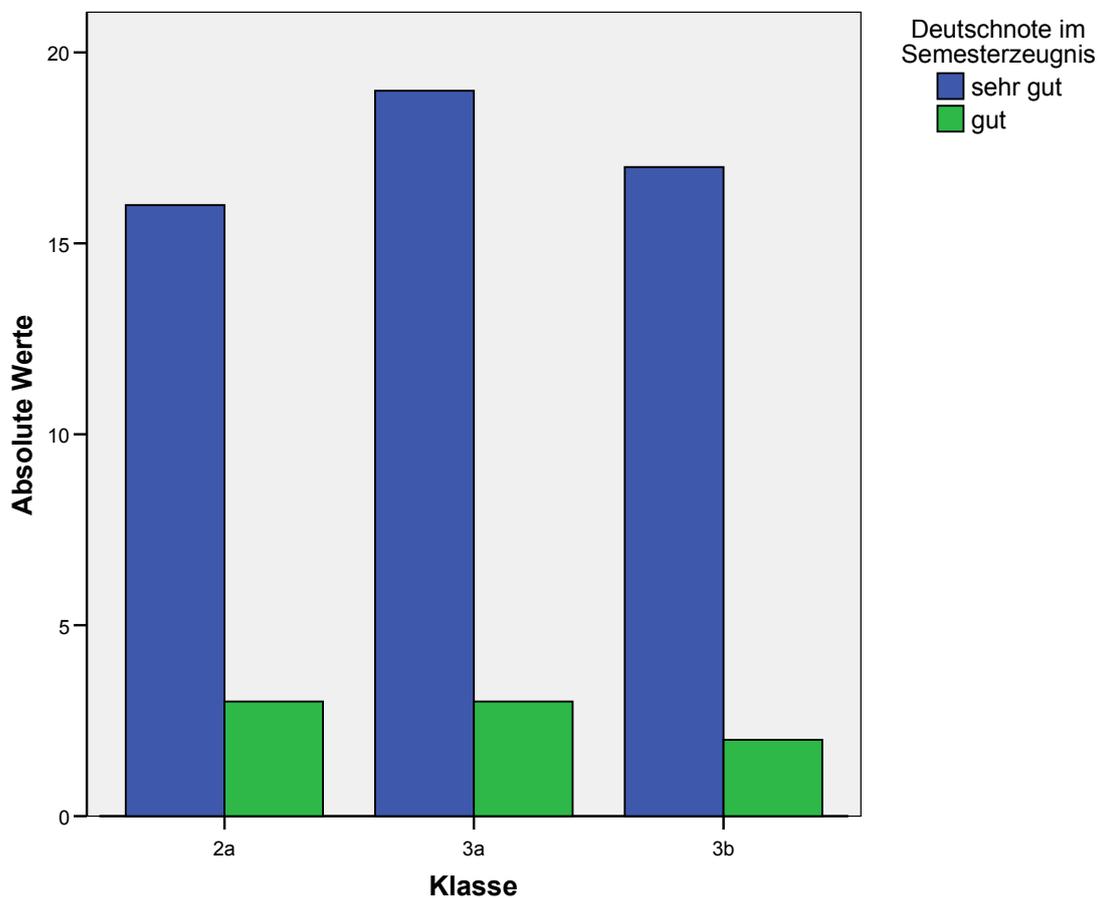


Abbildung 15: Deutschnoten der SchülerInnen der 2. und 3. Klasse (n=60)

Tabelle 5: Verteilung der Mathematiknote in der 2. und 3. Klasse (n=60)

Klasse	Mathematiknote	
	Sehr gut	Gut
2a	16	3
3a	17	5
3b	19	0
Gesamt	52	8

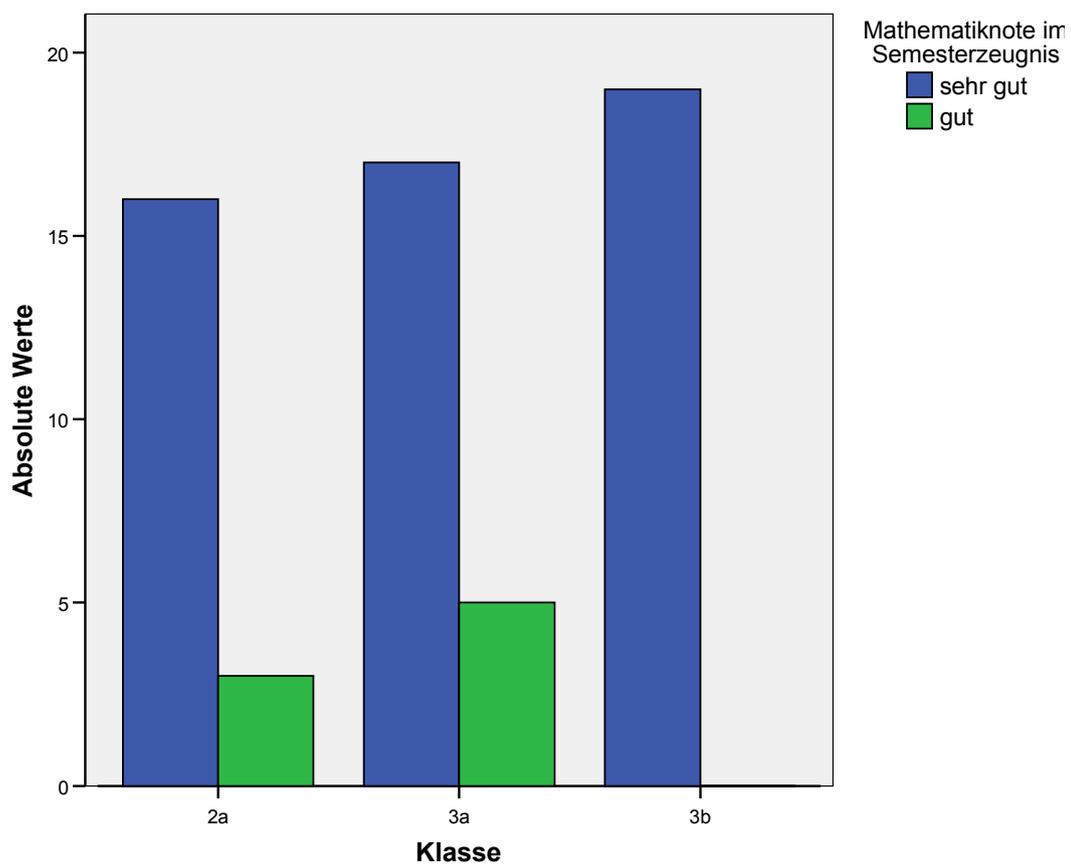


Abbildung 16: Mathematiknoten der SchülerInnen der 2. und 3. Klasse (n=60)

Die Verteilung der Gesamtnote der Kinder der ersten Klasse zeigt schon ein etwas differenzierteres Bild, wonach die Noten Sehr gut und Gut annähernd gleich häufig und einmal die Note Befriedigend vergeben wurden.

Tabelle 6: Verteilung der Gesamtnote in der 1. Klasse (n=24)

Klasse	Gesamtnote der 1. Klasse			Fehlend
	<i>Sehr gut</i>	<i>Gut</i>	<i>Befriedigend</i>	
1a	11	10	1	2

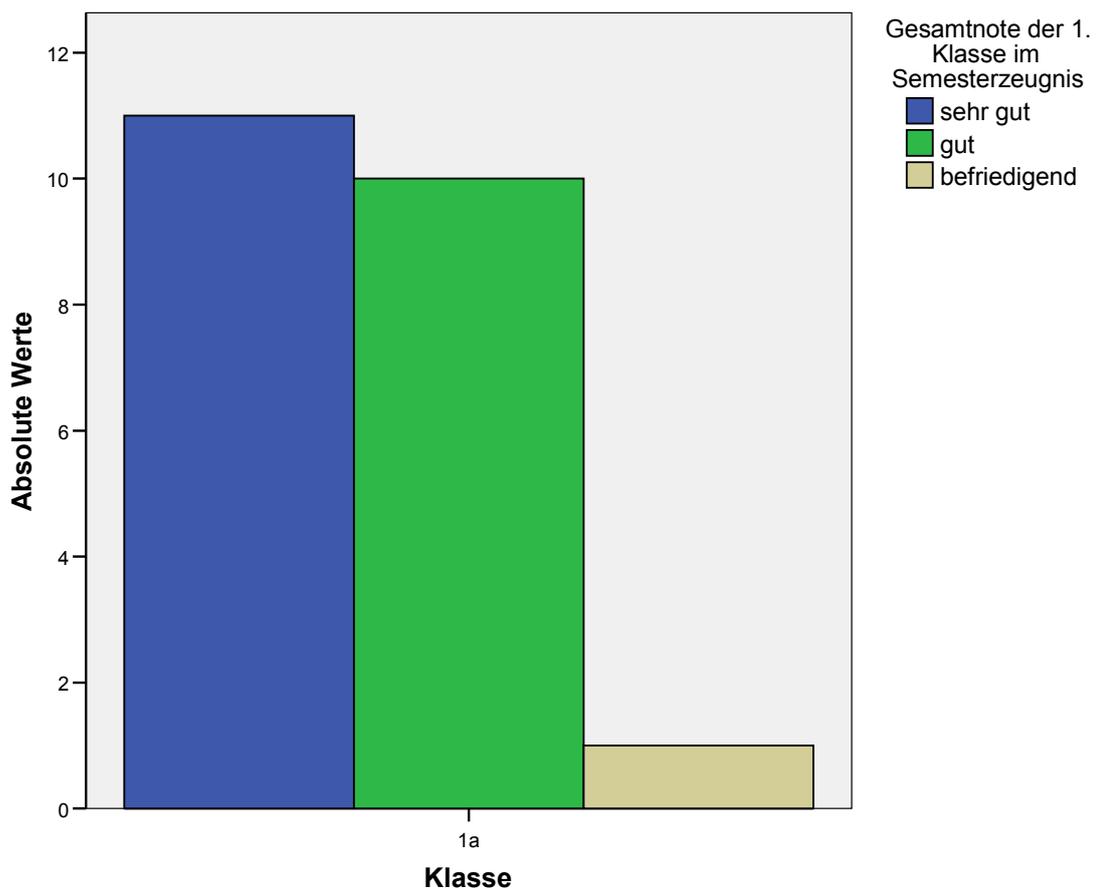


Abbildung 17: Gesamtnoten der SchülerInnen der 1. Klasse (n=24)

9.1.4 Verteilung der Testkennwerte der Reasoning-Verfahren

In Tabelle 7 sind sämtliche Mittelwerte und Standardabweichungen der erhaltenen Testkennwerte aufgelistet, die mangels normierter Testwerte fast ausschließlich in Rohwerten angeführt sind, mit Ausnahme des Testkennwertes über die *Untertests drei bis fünf* des CFT 1, der sowohl als Rohscore als auch als T-Wert angegeben ist.

Tabelle 7: SPSS-Output: Deskriptive Statistiken der Testkennwerte über alle Klassen

Deskriptive Statistiken

	N	Mittelwert	Standard abweichung	Minimum	Maximum
Summenscore CFT, Untertest 3	83	8,80	2,213	2	12
Summenscore CFT, Untertest 4	83	9,78	1,601	5	12
Summenscore CFT, Untertest 5	83	10,06	2,386	4	12
Summenscore CFT Untertest 3,4,5	83	28,64	4,728	12	35
T-Wert CFT 3_4_5	83	57,01	8,912	32	73
Summenscore FF frei	79	12,28	3,724	2	24
Summenscore FF MC	83	13,25	3,367	2	20
Summenscore PSB frei	79	23,61	4,614	13	32
Summenscore PSB MC	83	24,39	5,014	9	33

In Bezug auf die Verteilung der von der Gesamtstichprobe in den einzelnen Reasoning-Verfahren erreichten Ergebnisse konnte bei den Untertests vier (*Ähnlichkeiten*) und fünf (*Matrizen*) des CFT 1 keine Normalverteilung festgestellt werden (siehe K-S-Anpassungstest in Tabelle 7). Auf Rohscoreebene betrachtet zeigt sich bei diesen beiden Untertests ein gerundeter Mittelwert von 10 gelösten Items bei einer Anzahl von 12 zu bearbeitenden Items pro Untertest, womit bei diesen Untertests von einem Deckeneffekt

ausgegangen werden muss, sprich die Untertests sind für die UntersuchungsteilnehmerInnen zu einfach gewesen und differenzierten demnach wenig gut im oberen Segment der Stichprobe, was sich auch in der grafischen Darstellung der Datenverteilung widerspiegelt (siehe Anhang). Auf Klassenebene betrachtet, muss an dieser Stelle explizit erwähnt werden, dass die Instruktionen des *CFT 1* in den Untertests vier und fünf unterschiedliche Bearbeitungszeiten in der ersten Klasse und den zweiten und dritten Klassen vorsehen, so stehen für den Untertest 4 in der ersten Klasse vier Minuten und in der zweiten und dritten Klasse nur eineinhalb Minuten zur Verfügung, im Untertest fünf haben die Erstklässler ganze sechs Minuten Zeit, während den Zweit- und Drittklässlern 4,5 Minuten zugestanden werden. Die nach den einzelnen Klassen differenzierte Verteilung des Untertests vier (siehe Anhang) ergab, dass lediglich in der ersten Klasse keine Normalverteilung vorliegt, in den anderen drei Klassen jedoch schon, wobei die Vermutung nahe liegt, dass dieser Umstand durch die sehr unterschiedlichen Zeitlimits zustande gekommen ist. Die Verletzung der Normalverteilung des Untertest fünf des *CFT 1* ist auf Klassenebene auf die 3a-Klasse zurück zu führen, wie im Anhang ersichtlich. Insgesamt neigen alle in die Untersuchung eingegangenen Untertests des *CFT 1* zu einer linksschiefen Verteilung (siehe Anhang), wenngleich teilweise keine statistische Signifikanz erreicht wurde. Dem *Gesamtwert* über die Untertests drei bis fünf des *CFT 1* kann durch die hier mögliche Normierung durchaus Normalverteilung attestiert werden.

Tabelle 8: SPSS-Output: Testung auf Normalverteilung der Testkennwerte in den Reasoning – Verfahren

Kolmogorov-Smirnov-Anpassungstest

	Roh-score CFT, UT 3	Roh-score CFT, UT 4	Roh-score CFT, UT5	Roh-score CFT, UT 3,4,5	T-Werte CFT 3_4_5	Roh-score FF frei	Roh-score FF MC	Roh-score PSB frei	Roh-score PSB MC
N	83	83	83	83	83	79	83	79	83
Parameter der Normalverteilung(a,b)									
Mittelwert	8,80	9,78	10,06	28,64	57,01	12,28	13,25	23,61	24,39
Standardabweichung	2,213	1,601	2,386	4,728	8,912	3,724	3,367	4,614	5,014
Extremste Differenzen									
Absolut	,141	,229	,244	,145	,121	,090	,133	,150	,096
Positiv	,075	,127	,208	,089	,111	,074	,097	,071	,071
Negativ	-,141	-,229	-,244	-,145	-,121	-,090	-,133	-,150	-,096
Kolmogorov-Smirnov-Z	1,282	2,083	2,219	1,320	1,101	,804	1,209	1,335	,876
Asymptotische Signifikanz (2-seitig)	,075	,000	,000	,061	,177	,538	,108	,057	,427

- a Die zu testende Verteilung ist eine Normalverteilung.
- b Aus den Daten berechnet.

9.2. T-Test für abhängige Stichproben

Um eventuelle Unterschiede in den Testleistungen in Abhängigkeit vom Antwortformat erkennen zu können, wurden bei den Tests *Formale Folgerichtigkeit* und *PSB* T-Tests für abhängige Stichproben gerechnet.

Die zweiseitige Signifikanzprüfung ergibt mit einem Wert von $p = 0,001$ bzw. $p < 0,001$ bei einem Alpha-Niveau von 5 % einen signifikanten Unterschied in der Testleistung zwischen den beiden Antwortformaten sowohl bei *Formale Folgerichtigkeit* als auch *PSB* (siehe

Tabelle 10). Bei näherem Betrachten der Mittelwerte zeigt sich bei beiden Verfahren ein minimaler Vorsprung des Multiple-Choice-Antwortformates.

Tabelle 9: SPSS-Output: Statistik bei gepaarten Stichproben bei *Formale Folgerichtigkeit*

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standard abweichung	Standardfehler des Mittelwertes
Paaren 1	Summenscore FF frei	12,29	78	3,746	,424
	Summenscore FF MC	13,13	78	3,343	,379
Paaren 2	Summenscore PSB frei	23,60	78	4,644	,526
	Summenscore PSB MC	24,42	78	5,077	,575

Tabelle 10: SPSS-Output: T-Test bei gepaarten Stichproben bei *Formale Folgerichtigkeit*

Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren 1	Summenscore FF frei & Summenscore FF MC	78	,374	,001
Paaren 2	Summenscore PSB frei & Summenscore PSB MC	78	,657	,000

Tabelle 11: SPSS-Output: Gepaarte Differenzen des T-Tests für gepaarte Stichproben bei *Formale Folgerichtigkeit*

		Gepaarte Differenzen					T	Df
		Mittelwert	Standard abweichung	Standard fehler des Mittelwertes	95% Konfidenzintervall der Differenz			
					Untere	Obere		
Paare n 1	Summenscore FF frei - Summenscore FF MC	-,833	3,979	,450	-1,730	,064	-1,850	77
Paare n 2	Summenscore PSB frei - Summenscore PSB MC	-,821	4,048	,458	-1,733	,092	-1,790	77

9.3. Korrelationsanalyse

Der folgende Abschnitt ist der Darstellung der anhand von Korrelationsberechnungen erhaltenen Zusammenhänge zwischen den verschiedenen Reasoning-Verfahren gewidmet. Da zum Zeitpunkt der Auswertung der vorliegenden Untersuchung noch keine Normierungsdaten zum Untertest *Formale Folgerichtigkeit* vorlagen, wurde bei den Berechnungen der Korrelationen auf die parameterfreie Rangkorrelation nach Spearman zurückgegriffen. Sämtliche Korrelationen wurden einseitig getestet, da stets von einem positiven Zusammenhang auszugehen war. Zunächst folgt in den Tabellen 12 und 13 eine Übersicht über die erhaltenen Korrelationskoeffizienten, welche in weiterer Folge im Einzelnen erläutert werden.

Tabelle 12: Korrelationen zwischen Kennwerten von Formale Folgerichtigkeit im freien Antwortformat und Kennwerten weiterer Reasoning-Verfahren und Schulnoten

	N	Formale Folgerichtigkeit freie Antwort
<i>Klassifikationen - CFT 1</i>	78	.345**
<i>Ähnlichkeiten - CFT 1</i>	78	.337**
<i>Matrizen - CFT 1</i>	78	.320**
Summenwert CFT 1	78	.451**
PSB freie Antwort	79	.405**
PSB Multiple - Choice	78	.360**
Formale Folgerichtigkeit Multiple - Choice	78	.320**
Deutschnote	55	-.232*
Mathematiknote	55	-.237*
Gesamtnote 1. Klasse	22	-.330

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).

** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

Tabelle 13: Korrelationen zwischen Kennwerten von Formale Folgerichtigkeit im Multiple - Choice Antwortformat und Kennwerten weiterer Reasoning-Verfahren und Schulnoten

	N	Formale Folgerichtigkeit Multiple – Choice
Klassifikationen - CFT 1	82	.289**
Ähnlichkeiten - CFT 1	82	.164
Matrizen - CFT 1	82	.301**
Summenwert CFT 1	82	.298**
PSB freie Antwort	78	.334**
PSB Multiple - Choice	83	.587**
Formale Folgerichtigkeit freie Antwort	78	.320**
Deutschnote	60	-.047
Mathematiknote	60	-.009
Gesamtnote 1. Klasse	21	-.450*

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).

** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

9.3.1 Korrelationen zwischen *Formale Folgerichtigkeit* und Reasoning Verfahren

9.3.1.1 Korrelationen zwischen *Formale Folgerichtigkeit* und PSB

Sämtliche Korrelationen zwischen den Versionen des PSB und *Formale Folgerichtigkeit* sind auf dem 0,01 - Niveau signifikant. Der größte Zusammenhang zwischen *Formale Folgerichtigkeit* und PSB besteht bei der Multiple-Choice Version des PSB und der Multiple-Choice Version von *Formale Folgerichtigkeit* mit einem Korrelationskoeffizienten von $r = .587$. Die Originalversion von *Formale Folgerichtigkeit* im freien Antwortformat korreliert in einem Ausmaß von $r = .405$ mit dem PSB im freien Antwortformat. Die zwischen PSB und *Formale Folgerichtigkeit* im gleichen Antwortformat gefundenen Korrelationen sind von sehr passablem Ausmaß und stellen erste Befunde für eine vorliegende konvergente Validität dar.

Immerhin korreliert der PSB im MC-Format mit *Formale Folgerichtigkeit* im freien Antwortformat noch zu $r = .36$ und der Zusammenhang von PSB im freien AF und *Formale*

Folgerichtigkeit in der MC-Version beträgt $r = .334$, wobei das geringere Ausmaß der beiden letztgenannten Korrelationen die Vermutung nahelegt, dass in Abhängigkeit vom Antwortformat doch etwas leicht Anderes gemessen wird.

9.3.1.2 Korrelationen zwischen *Formale Folgerichtigkeit* und CFT 1

Da die Untertests drei bis fünf des CFT 1 über ein Multiple-Choice Antwortformat verfügen, könnten hier Korrelationen von höherem Ausmaß zu der Multiple-Choice Version von *Formale Folgerichtigkeit* erwartet werden. Sämtliche Korrelationen zwischen den Subtests des CFT 1 und *Formale Folgerichtigkeit* sind auf dem 0,01-Niveau signifikant.

Untertest 3 – Klassifikationen

Der Untertest *Klassifikationen* korreliert entgegen der oben geäußerten Überlegung die Ähnlichkeit zum Multiple-Choice Antwortformat betreffend mit *Formale Folgerichtigkeit* im freien Antwortformat ($r = .345$) höher als im Multiple-Choice Antwortformat ($r = .289$).

Untertest 4 – Ähnlichkeiten

Beim Untertest *Ähnlichkeiten* zeigt sich anhand der gefundenen Korrelationen nur mehr ein Zusammenhang von $r = .164$ mit der Multiple-Choice Version von *Formale Folgerichtigkeit*, wohingegen mit *Formale Folgerichtigkeit* im freien Antwortformat eine Korrelation im Ausmaß von $r = .337$ besteht.

Untertest 5 – Matrizen

Beim Untertest *Matrizen* zeigt sich bei beiden Antwortformaten von *Formale Folgerichtigkeit* eine ähnlich starke Korrelation, einerseits von $r = .32$ im freien und andererseits von $r = .301$ im Multiple-Choice Antwortformat.

Gesamtwert über die Untertests 3, 4 und 5

Die zu einem *Gesamtwert* zusammengefassten Untertests drei bis fünf des *CFT 1* korrelieren mit *Formale Folgerichtigkeit* im freien Antwortformat in einem durchaus beachtlichen Ausmaß von $r = .451$ und im Multiple-Choice Format zu $r = .298$.

Über sämtliche angewendete Untertests des *CFT 1* konnten statistisch signifikante positive Zusammenhänge mit *Formale Folgerichtigkeit* beobachtet werden, wobei sich die größten Korrelationen zwischen den *CFT* - Daten und *Formale Folgerichtigkeit* im freien Antwortformat zeigten.

9.3.2 Korrelationen zwischen *Formale Folgerichtigkeit* und Schulnoten

Zwischen *Formale Folgerichtigkeit* und Schulnoten bestehen nur teilweise signifikante Korrelationen, die auf einem Signifikanzniveau von 0,05 gelten. Da für die erste Klasse nur eine Gesamtnote in die Berechnungen einging und von der zweiten und dritten Klasse Mathematik- und Deutschnote zur Verfügung standen, erfolgt die Darstellung der Ergebnisse getrennt nach der Gesamtnote der Erstklässler, der Deutsch- und der Mathematiknote der Zweit- und Drittklässler.

9.3.2.1 Korrelationen zwischen *Formale Folgerichtigkeit* und der Gesamtnote im Semesterzeugnis der ersten Klasse

Die Korrelation der Gesamtnote der Erstklässler im Semesterzeugnis mit dem Test *Formale Folgerichtigkeit* erreicht im Multiple-Choice Antwortformat einen sehr akzeptablen Wert von $r = -.45$. Für das freie Antwortformat konnte hingegen keine signifikante Korrelation gefunden werden.

9.3.2.2 Korrelationen zwischen *Formale Folgerichtigkeit* und der Deutschnote im Semesterzeugnis der 2. und 3. Klassen

Bei den Noten der Zweit- und Drittklässler stellt sich ein konträres Bild zu den oben berichteten Zusammenhängen dar. Es zeigt sich nun ein signifikanter Zusammenhang von immerhin $r = - .232$ zwischen der Deutschnote und *Formale Folgerichtigkeit* in der Version freie Antwort. Zwischen der Multiple-Choice Version von *Formale Folgerichtigkeit* und der Deutschnote konnte hingegen kein signifikanter Zusammenhang festgestellt werden.

9.3.2.3 Korrelationen zwischen *Formale Folgerichtigkeit* und der Mathematiknote im Semesterzeugnis der 2. und 3. Klassen

Ein ähnliches Bild zeigt sich in Bezug auf die Mathematiknote, wobei wieder ein signifikanter Zusammenhang mit *Formale Folgerichtigkeit* im freien Antwortformat ($r = - .237$), aber kein signifikanter Zusammenhang zu *Formaler Folgerichtigkeit* mit Multiple-Choice Antwortformat gefunden wurde.

Im Sinne einer konvergenten Validierung des Reasoning-Verfahrens *Formale Folgerichtigkeit* in Bezug auf Schulnoten wäre zu erwarten gewesen, dass die Korrelation zwischen Test und Mathematiknote höher ausfallen würde als die Korrelation mit der Deutschnote. Jedoch muss an dieser Stelle darauf hingewiesen werden, dass die Streuung der Noten der zweiten und dritten Klasse nicht sehr breit ist und von 60 Kindern in Deutsch und Mathematik jeweils 52 ein Sehr gut im Semesterzeugnis hatten. So gesehen scheint die Gesamtnote der ersten Klasse mit einer Verteilung von 11 Sehr gut, 10 Gut und einem Befriedigend eine geeignetere Variable zur Berechnung einer Korrelation zu sein, die mit $r = - .45$ sogar sehr ansprechend ausfällt, jedoch ist sie nicht nach Schulfächern differenziert im Sinne eines einheitlichen Fähigkeitskonstrukts interpretierbar, sondern nur dahingehend, dass ein gewisser Zusammenhang zwischen dem Test *Formale Folgerichtigkeit* in der Multiple-Choice Version mit einem globalen

schulischen Leistungswert besteht, der für den Test im freien Antwortformat nach statistischer Überprüfung jedoch nicht gegeben ist.

10 Zusammenfassung der Ergebnisse

In Bezug auf die Untertests drei bis fünf der Grundintelligenztest – Skala 1

In Bezug auf den CFT 1 lässt sich der größte Zusammenhang zwischen *Formale Folgerichtigkeit* im freien Antwortformat und dem Gesamtwert der Untertests drei bis fünf des CFT 1 ($r = ,451$), die nach Angabe der Testautoren zusammen einen Testwert für schlussfolgerndes Denken darstellen, finden.

Die Korrelationen der Untertests des CFT 1 sowie des Gesamtwertes für schlussfolgerndes Denken mit *Formale Folgerichtigkeit* im freien Antwortformat ($r = .32$ bis $r = .451$) sind durchwegs größer als die Korrelationen mit der Multiple-Choice Version ($r = .164$ bis $r = .301$) von *Formale Folgerichtigkeit*.

In Bezug auf den Untertest drei des Prüfsystems für Schul- und Bildungsberatung

Der größte Zusammenhang im Rahmen der Korrelationsanalyse lässt sich zwischen *Formale Folgerichtigkeit* im Multiple-Choice Antwortformat und dem PSB im Multiple-Choice Format ($r = .587$) beobachten. Korreliert man die freie Antwortversionen von PSB und *Formale Folgerichtigkeit*, ergibt sich ein Zusammenhang von $r = .405$.

In Bezug auf die Schulnoten

Die größte Korrelation im Bereich der Schulnoten ($r = - .45$) findet sich zwischen der Gesamtnote der 1. Klasse und dem Test *Formale Folgerichtigkeit* im Multiple-Choice Antwortformat, wohingegen der Zusammenhang zu *Formale Folgerichtigkeit* im freien Antwortformat nicht signifikant ausfällt. Bei den Noten in *Deutsch* und *Mathematik* der Zweit- und DrittklässlerInnen zeigt sich in Bezug auf das Antwortformat ein konträres

Ergebnis, nämlich eine signifikante Korrelation zwischen *Formale Folgerichtigkeit* im freien Antwortformat und der *Deutschnote* ($r = - .232$) sowie mit der *Mathematiknote* ($r = - .237$); die Korrelationen der *Deutsch-* und *Mathematiknote* mit der Multiple-Choice Version von *Formale Folgerichtigkeit* fallen nicht signifikant aus.

Aufgrund der gefundenen positiven Zusammenhänge zwischen *Formale Folgerichtigkeit* und den eingesetzten Reasoning-Verfahren kann dem Test *Formale Folgerichtigkeit* konvergente Kriteriumsvalidität zugesprochen werden. Die Korrelationen mit den Schulnoten lassen keine differenzierten Schlussfolgerungen zu.

11 Diskussion

Durchführung von Gruppentestungen im Volksschulalter

Durch die Dauer der Testung und Aufgabenfülle kam es nach anfänglich guter Motivation besonders bei der Gruppentestung bei manchen Kindern zu Ermüdung, Frustration, Motivationsabfall, was dazu führte, dass schwierigere Aufgaben teilweise ausgelassen wurden oder scheinbar irgendeine Lösung angekreuzt wurde, in der Hoffnung dadurch schneller fertig zu werden. Manchen Kindern fiel es besonders schwer, sich nur auf ihre Testbögen zu konzentrieren und still zu arbeiten, wodurch es teilweise zu Störungen und Ablenkungen der anderen Kinder kam. Besonders eindrücklich zu beobachten war, dass sich durch die Bank die Buben schwerer taten, ruhig sitzend und konzentriert zu arbeiten.

Da die Kinder in der Gruppentestung durch Unterschiede im Arbeitsstil, der Motivation und Konzentration ein sehr heterogenes Arbeitstempo an den Tag legten und bei *PSB* und *Formale Folgerichtigkeit* für die vorliegende Testung keine Zeitbeschränkungen eingeführt wurden, waren einige Kinder früher fertig als andere, was im ersteren Fall zu Langeweile und langen Erholungspausen zwischen den Tests, und im anderen Fall zu Überforderung und Frustration mit wenig Regenerationsphasen führte. Der *CFT 1* verfügt zwar über Zeitbeschränkungen, jedoch sind diese mit Ausnahme des Untertests vier bei den Zweit- und Drittklässlern im Ausmaß von 1,5 Minuten sehr großzügig (vier bis sechs

Minuten), sodass auch hier viele Kinder vor Ablauf der Zeit fertig waren, was hier manchmal zu einem Zurückblättern zu bereits abgeschlossenen Untertests oder ein Vorblättern zum nächsten Untertest des *CFT 1* und dortigen Ausbessern oder Vorarbeiten verführte. Bei *Formale Folgerichtigkeit* existiert im freien Antwortformat eine Zeitbeschränkung von 30 Sekunden bis eineinhalb Minuten pro Item, für die MC-Version in der Gruppe ist jedoch keine Zeitbeschränkung vorgesehen. Für den *PSB* in der Originalversion, der ebenfalls als Gruppentest durchgeführt werden kann, haben die Testautoren ein Zeitlimit von fünf Minuten für die Bearbeitung von 40 figuralen Reihen vorgegeben, im Gegensatz dazu brauchten manche Kinder bis zu 15 Minuten für die Bearbeitung der 35 Items der adaptierten Multiple-Choice Version des *PSB*. Zusammenfassend scheinen für Gruppentestungen im Volksschulalter ein angemessenes Zeitlimit für die Bearbeitung solcherlei Aufgaben und fixe Pausen zum Zwecke ausreichender Erholung unumgänglich, was leider in der vorliegenden Studie nicht ausreichend berücksichtigt wurde.

Multiple-Choice Antwortdesign im Volksschulalter

Das Multiple-Choice Antwortformat war für die meisten Kinder eine neue und ungewohnte Art, Aufgaben zu bearbeiten, sodass es hier teilweise zu Verständnisschwierigkeiten kam. Ein Kennenlernen und Üben von Multiple-Choice Aufgaben vor einer solchen Testung wäre für UntersuchungsleiterInnen und -teilnehmerInnen mit Sicherheit sehr günstig, um Verständnisschwierigkeiten und Überforderung besonders bei den Kleinsten zu vermeiden.

Zeichenfähigkeit im Volksschulalter

Bei einigen Items der adaptierten Version des *PSB* im freien Antwortformat, taten sich die Kinder schwer, die geometrischen Figuren richtig zu zeichnen, was teilweise zu Überforderung und Frustration angesichts der vermeintlich fehlenden Zeichenbegabung führte. Somit wäre es bei Kindern im Volksschulalter hilfreich, vor der eigentlichen Testung eine gewisse Vertrautheit mit neuen Figuren zu schaffen oder nur solche Formen auszuwählen, die den Kindern vertraut sind, um Nachteile ihrerseits zu vermeiden.

Volksschulnoten als Außenkriterium

Es stellt sich die Frage, inwieweit die Schulnoten der Kinder im Rahmen der vorliegenden Studie als Außenkriterium im Zuge der Kriteriumsvalidierung geeignet sind, wenn man bedenkt, dass die gesamte Stichprobe im Umfang von 84 Kindern die Note Sehr gut im Fach Sachunterricht hat und von den 60 Kindern der zweiten und dritten Klassen nur jeweils acht Kinder entweder ein Gut in Mathematik oder in Deutsch hatten und der Rest ein Sehr gut. Nichtsdestotrotz ließ sich ein positiver Zusammenhang der Schulnoten mit dem Test *Formale Folgerichtigkeit* finden.

Conclusio

Die Kinder der vorliegenden Stichprobe hatten teilweise Probleme mit dem Multiple-Choice Antwortformat, dem Einzeichnen von Lösungen, der Fülle von zu bearbeitenden Aufgaben und nicht zuletzt mit den an sie gestellten Herausforderungen durch die Gruppenbedingung.

Für die Übertragbarkeit der Ergebnisse die konvergente Validität des Untertests *Formale Folgerichtigkeit* betreffend auf den gesamten Altersbereich scheinen aus den oben erwähnten Gründen der speziellen Stichprobe von sehr jungen Kindern noch differenzierte Analysen des Altersbereiches von 10 bis 15 Jahren getrennt nach Hauptschule und Gymnasium notwendig.

12 Literaturverzeichnis

- Abel, J. (1988). Untersuchung zur Reliabilität, Validität und Parallelität des "Prüfsystem für Schul- und Bildungsberatung (PSB). *Diagnostica* 34(4), 351-363.
- American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin. Supplement*, 51(2, Part 2), 1-38.
- Beckmann, J. F. & Guthke, J. (1999). *Psychodiagnostik des schlussfolgernden Denkens. Handbuch zur adaptiven computergestützten Intelligenz-Lerntestbatterie für schlussfolgerndes Denken (ACIL)*. Göttingen: Hogrefe.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Cattell, R. B., Weiß, R. H. & Osterland, J. (1997). *Grundintelligenztest Skala 1 (CFT 1)*. (5., revidierte Auflage). Göttingen: Hogrefe.
- Conrad, W. (1983). *Intelligenzdiagnostik*. In K.J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik (Enzyklopädie der Psychologie, Themenbereich B, Serie II, Band 2. S. 104-201)*. Göttingen: Hogrefe.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52(4), 281-302.
- Ecker, J. (2011). *Erfassung wesentlicher kognitiver Operationen bei der Lösung von Aufgaben zum logisch-schlussfolgernden Denken im Kindesalter unter Berücksichtigung der Antwortformate Multiple-Choice und freies Antwortformat*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Guilford, J. P. & Hoepfner, R. (1971). *The analysis of Intelligence*. New York: McGraw-Hill.
- Häcker, H. & Stapf, K. H. (2009). *Dorsch Psychologisches Wörterbuch (15., überarbeitete und erweiterte Aufl.)*. Bern: Hans Huber.
- Hagenmüller, B. (2011). *Entwicklung des Untertests „Formale Folgerichtigkeit“ zur Erfassung von Reasoning in der Intelligenz-Testbatterie AID 3*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology* 57(5), 253-270.

- Horn, W., Lukesch, H., Kormann, A. & Mayrhofer, S. (2002). *Prüfsystem für Schul- und Bildungsberatung für 4. bis 6. Klassen – revidierte Fassung. (PSB-R-4-6)*. Göttingen: Hogrefe.
- Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung (P-S-B)*. Göttingen: Hogrefe.
- Klauer, K. J. (2001). *Training des induktiven Denkens*. In K. J. Klauer (Hrsg.), *Handbuch Kognitives Training* (2. überarbeitete und erweiterte Aufl., S. 165-209). Göttingen: Hogrefe.
- Kubinger, K. D. (2003). *Gütekriterien*. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 195-204). Berlin: Beltz.
- Kubinger, K. D. (2006). *Psychologische Diagnostik*. Göttingen: Hogrefe.
- Kubinger, K. D. (2009). *Adaptives Intelligenz Diagnostikum – Version 2.2 (AID 2) samt AID 2 - Türkisch*. Göttingen: Beltz.
- Kubinger, K. D. & Holocher-Ertl, S. (in Vorb.). *Adaptives Intelligenzdiagnostikum – Version 3*.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz.
- Lohaus, A., Maass, A., Vierhaus, M. (2010). *Entwicklungspsychologie des Kindes- und Jugendalters*. Berlin, Heidelberg: Springer-Verlag.
- Meiser, T. & Klauer, K. C. (2001). *Training des deduktiven Denkens*. In K. J. Klauer (Hrsg.), *Handbuch Kognitives Training* (2. überarbeitete und erweiterte Aufl., S. 211-234). Göttingen: Hogrefe.
- Preckel, F. (2003). *Diagnostik intellektueller Hochbegabung. Testentwicklung zur Erfassung der fluiden Intelligenz*. Göttingen: Hogrefe.
- Preckel, F. & Vock, M. (2010). *Intelligenzdiagnostik*. In E. Walther, F. Preckel & S. Mecklenbräuer (Hrsg.), *Befragung von Kindern und Jugendlichen. Grundlagen, Methoden und Anwendungsfelder* (S. 99-131). Göttingen: Hogrefe.
- Rauchfleisch, U. (2005). *Testpsychologie*. (4. Aufl.). Göttingen: Vandenhoeck & Ruprecht.
- Rindermann, H. & Kwiatkowski, V. (2010). *Intelligenzdiagnostik im Kindesalter*. In C. Quaiser-Pohl & H. Rindermann (Hrsg.), *Entwicklungsdiagnostik* (S. 102-132). München: Reinhardt.

- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. (2., vollständig überarbeitete und erweiterte Auflage). Bern: Hans Huber.
- Schweizer, K. (2006a). *Intelligenz*. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 2-15). Heidelberg: Springer.
- Schweizer, K. (2006b). *Intelligenzdiagnostik*. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 70-83). Heidelberg: Springer.
- Spearman, C. (1904). "General intelligence", objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-293.
- Spearman, C. (1914). The theory of two factors. *Psychological Review*, 21(2), 101-115.
- SPSS Inc. (2005). SPSS für Windows, Version 14.0.
- Stemmler, G., Hagemann, D., Amelang, M. & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung* (7. vollständig überarbeitete Aufl.). Stuttgart: Kohlhammer.
- Sternberg, R. J. (2009). *Cognitive Psychology*. (5. Aufl.). Belmont: Cengage.
- Süß, H.-M. (2003). *Intelligenztheorien*. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 217-224). Berlin: Beltz.
- Waldmann, M. & Weinert, F. E. (1990). *Intelligenz und Denken. Perspektiven der Hochbegabtenforschung*. Göttingen: Hogrefe.
- Werani, A. (2009). Die Rolle der Qualität inneren Sprechens beim Problemlösen. *Journal für Psychologie*, 17(3), 1-22, Zugriff am 27.10.2011, verfügbar unter <http://www.journal-fuer-psychologie.de/jfp-3-2009-05.html>.
- Weiß, R. H. (2008). *Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest – Revision (WS/ZF-R)*. Göttingen: Hogrefe.

III ANHANG

13 Tabellenverzeichnis

Tabelle 1: Deskriptive Statistik des Alters der teilnehmenden Kinder (n=84).....	41
Tabelle 2: SPSS-Output: Überprüfung der Normalverteilung des Alters anhand des Kolmogorov-Smirnov-Anpassungstest.....	42
Tabelle 3: Verteilung des Geschlechts der VersuchsteilnehmerInnen (n=84).....	43
Tabelle 4: Verteilung der Deutschnote in der 2. und 3. Klasse (n=60).....	44
Tabelle 5: Verteilung der Mathematiknote in der 2. und 3. Klasse (n=60).....	45
Tabelle 6: Verteilung der Gesamtnote in der 1. Klasse (n=24).....	46
Tabelle 7: SPSS-Output: Deskriptive Statistiken der Testkennwerte über alle Klassen.....	47
Tabelle 8: SPSS-Output: Testung auf Normalverteilung der Testkennwerte in den Reasoning – Verfahren.....	49
Tabelle 9: SPSS-Output: Statistik bei gepaarten Stichproben bei <i>Formale Folgerichtigkeit</i>	50
Tabelle 10: SPSS-Output: T-Test bei gepaarten Stichproben bei <i>Formale Folgerichtigkeit</i>	50
Tabelle 11: SPSS-Output: Gepaarte Differenzen des T-Tests für gepaarte Stichproben bei <i>Formale Folgerichtigkeit</i>	50
Tabelle 12: Korrelationen zwischen Kennwerten von <i>Formale Folgerichtigkeit</i> im freien Antwortformat und Kennwerten weiterer Reasoning-Verfahren und Schulnoten.....	51
Tabelle 13: Korrelationen zwischen Kennwerten von <i>Formale Folgerichtigkeit</i> im Multiple - Choice Antwortformat und Kennwerten weiterer Reasoning-Verfahren und Schulnoten.....	52

14 Abbildungsverzeichnis

Abbildung 1: Guilfords Würfelmodell	13
Abbildung 2: Jägers Intelligenzstrukturmodell	14
Abbildung 3: Carrolls Three Stratum Theory	15
Abbildung 4: Matrizenaufgabe aus dem SPM	17
Abbildung 5: Beispielitem aus dem Subtest 3 Klassifikationen des CFT 1	19
Abbildung 6: Beispielitem aus dem Subtest 4 Ähnlichkeiten des CFT 1	20
Abbildung 7: Beispielitem aus dem Subtest 5 Matrizen des CFT 1	20
Abbildung 8: Originalitem des Untertests drei aus dem PSB	22
Abbildung 9: Beispielaufgabe aus der adaptierten Testversion des Untertest 3 aus dem PSB im freien Antwortformat	23
Abbildung 10: Beispielaufgabe aus der adaptierten Testversion des Untertest 3 aus dem PSB im Multiple-Choice-Antwortformat	23
Abbildung 11: Beispielitem aus dem Testheft für 6 bis 7 Jahre aus dem Test Formale Folgerichtigkeit	24
Abbildung 12: standardisierte Anordnung der Lösungssteine von Formale Folgerichtigkeit.....	24
Abbildung 13: Beispielaufgabe der adaptierten Testversion des Untertest Formale Folgerichtigkeit	25
Abbildung 14: Verteilung des Alters der teilnehmenden Kinder.....	42
Abbildung 15: Deutschnoten der SchülerInnen der 2. und 3. Klasse (n=60).....	44
Abbildung 16: Mathematiknoten der SchülerInnen der 2. und 3. Klasse (n=60).....	45
Abbildung 17: Gesamtnoten der SchülerInnen der 1. Klasse (n=24).....	46

Antrag beim Stadtschulrat



universität
wien



**Test- und
Beratungsstelle**

Fakultät für Psychologie

Institut für Entwicklungspsychologie und
Psychologische Diagnostik
Liebiggasse 5
1010 Wien

T +43 (1) 4277-478 51

An Herrn Mag. Klaus Rick
Präsidium des Stadtschulrates für Wien
Wipplingerstraße 28
1010 Wien

Wien, am 20.10.2010

Betrifft: Studie zur Erfassung wesentlicher kognitiver Operationen bei der Lösung von Aufgaben zum logisch-schlussfolgernden Denken im Kindesalter (6 - 9Jahre)

Viele heute etablierte psychologisch-diagnostische Verfahren zur Erfassung der intellektuellen Leistungsfähigkeit und des logischen-schlussfolgernden Denkens sind mit dem multiple-choice Antwortformat gestaltet. Dieses ist dadurch gekennzeichnet, dass die getesteten Kinder aus einer Reihe von verfügbaren Antwort-Alternativen jene auswählen, die sie für richtig halten. Dem gegenüber stehen psychologisch-diagnostische Instrumente, bei denen ohne eine Auswahl von Antwortalternativen die Lösung der Aufgabe von den Kindern gefunden werden muss – das sogenannte freie Antwortformat.

In Hinblick auf eine gültige Interpretation der Testwerte stellt sich hier die Frage, inwieweit diese zwei unterschiedlichen Antwortformate zu gleichen Ergebnissen

kommen bzw. sich dahingehend unterscheiden. Bisherige Studien, die allerdings nur mit Erwachsenen durchgeführt wurden, deuten darauf hin, dass die beiden Antwortformate zu unterschiedlichen Ergebnissen führen, da dabei jeweils unterschiedliche kognitive Operationen wesentlich zur Lösung beitragen. Gerade im Bereich von Kindern und Jugendlichen ist diese Frage von hoher Relevanz, weil aufgrund der häufig fehlenden Kontrollstrategien und der beobachteten Impulsivität im Verhalten während der Testung, Auswirkung auf die erzielten Ergebnisse, je nach verwendetem Antwortformat, zu erwarten sind.

Wir ersuchen nun Schulen und Kinderbetreuungseinrichtungen in Wien und Niederösterreich um deren Unterstützung bzw. Mitwirkung an unserem Projekt. Die Schüler(innen)-Testung wird von Wien aus koordiniert (Projektleitung: Dr. Stefana Holocher-Ertl; Projektkoordination/Ansprechpartner: Carina Heuschober & Johann Ecker). Wir bemühen uns, den Aufwand für die jeweilige Schule so gering wie möglich zu halten.

Die teilnehmenden Schulen leisten einen wesentlichen Beitrag zur Etablierung einer fundierten Leistungsdiagnostik. Damit sollen die Schüler(innen) in Zukunft optimal hinsichtlich verschiedener leistungsbezogener Aspekte beraten werden können, was wiederum den Schulen zugute kommt.

Es besteht die Möglichkeit – sollte das gewünscht sein –, den Schüler(innen) bzw. deren Eltern als Dankeschön einen kurzen schriftlichen Ergebnisbericht über kognitive Stärken und Schwächen des Kindes auszuhändigen.

Mit der Bitte um Unterstützung und Genehmigung dieser Studie und freundlichen Grüßen!

Johann Ecker

Diplomand an der Fakultät für Psychologie
Universität Wien

Carina Heuschober

Diplomandin an der Fakultät für Psychologie
Universität Wien

Anlagen:

Ansprechpartner

Untersuchungsziel und –ablauf

Informationen zu den eingesetzten Leistungstests

Stichprobenzusammensetzung

Elternbrief

Lehrer(innen)brief

Ansprechpartner:

Projektleitung: Projektleiterin ist Universitätsassistentin, Dr. Stefana Holocher-Ertl.

Projektkoordination:

Als **Ansprechpartner** stehen Ihnen Johann Ecker und Carina Heuschober zur Verfügung.
Bitte entnehmen Sie nachfolgend die Kontaktdaten:

Johann Ecker

Diplomand an der Fakultät für Psychologie
Universität Wien

E-Mail: a0404089@unet.univie.ac.at

Tel.: +43 (0) 699 1922 5646

Carina Heuschober

Diplomandin an der Fakultät für Psychologie
Universität Wien

E-Mail: carina.heuschober@st.ursula-wien.at

Tel.: +43 (0) 660 7399225

Untersuchungsablauf:

Jede/r an der Studie teilnehmende Schüler/in wird zweimal getestet. Die Hälfte der Kinder zuerst mit Testaufgaben zum logisch-schlussfolgernden Denken mit multiple-choice Antwortformat, die andere Hälfte zuerst mit Testaufgaben zum logisch-schlussfolgernden Denken mit freiem Antwortformat. Beim zweiten Testtermin werden dann diejenigen, die zuerst Aufgaben mit freiem Antwortformat bearbeiteten, mit den

multiple-choice Aufgaben getestet, und umgekehrt. Die Testaufgaben mit freiem Antwortformat werden dabei in Einzeltestungen vorgegeben, die Aufgaben mit multiple-choice Antwortformat werden im Rahmen einer Gruppentestung durchgeführt.

Die Hälfte der teilnehmenden Kinder werden dabei mit den Verfahren **PSB** (Horn, 2003), *logische Blöcke* aus dem **AID 3** (Kubinger, 2009) und drei Untertests aus dem **CFT 1** (Cattell, Weiß und Osterland, 1997) getestet. Bei der anderen Hälfte der Teilnehmer(innen) kommt statt des **CFT 1** ein **Matrizentest** zum Einsatz.

Bei den **Einzeltestungen** wird jedes Kind von einer/einem Testleiter/in einzeln von der Klasse abgeholt und in einen von der Schule zur Verfügung gestellten Raum (freier Klassenraum, Bibliothek, Arzttraum etc.) gebracht, in dem die Testungen stattfinden. Der Raum muss abgesehen von zwei Stühlen und einem Tisch über keine weitere Ausstattung verfügen.

Sowohl die Einzel- als auch Gruppentestung finden während der Unterrichtszeit des Kindes statt und dauern **jeweils ca. 1 Schulstunde**, wobei diese angegebene Zeitspanne ein Durchschnittswert ist und je nach individueller Arbeitsweise der Kinder schwanken kann.

Natürlich werden die gewonnenen Daten im Sinne des **Datenschutzes** ausschließlich für wissenschaftliche Zwecke verwendet, die Testleiter(innen) sind angewiesen, die Ergebnisse der Kinder noch vor der Testung zu **anonymisieren**.

Es werden nur Kinder getestet, deren **Eltern** ihr **Einverständnis** gegeben haben. Der Elternbrief ist in Deutsch abgefasst und enthält wichtige Informationen an die Erziehungsberechtigten. Bei der Terminvereinbarung legen wir großen Wert darauf, den Ablauf des regulären Unterrichts so wenig wie möglich zu unterbrechen. Wir werden versuchen – im Rahmen der Möglichkeiten – Rücksicht auf die Wünsche der einzelnen Schulen zu nehmen.

Testleiter(innen) sind Diplomanden(innen) des Arbeitsbereichs Psychologische Diagnostik der Fakultät für Psychologie.

Informationen zu den eingesetzten Verfahren

Um etwaige Unterschiede in den Ergebnissen zu erfassen, kommen bei dieser Studie folgende Verfahren zum Einsatz:

- Das *Prüfsystem für Schul- und Bildungsberatung (PSB; Horn, 2003)* ist ein etabliertes Verfahren zur Erfassung der kognitiven Leistungsfähigkeit von Kindern und Jugendlichen. Aus diesem Verfahren wird ein Untertest zur Anwendung kommen, der speziell logisches Schlussfolgern misst.
- Untertest „*Logische Blöcke*“ aus der *aktualisierten und erweiterten Intelligenztestbatterie AID 3 (Adaptives Intelligenz Diagnostikum: in der aktuellen Fassung AID 2, Version 2.2, Kubinger, 2009)*. Dieses Verfahren ist ein Untertest zur Erfassung des logisch-schlussfolgernden Denkens anhand visuellen Stimulusmaterials.
- Ein *Matrizentest* zur Erfassung des logischen Schlussfolgerns bei Kindern und Jugendlichen, der aktuell an der Fakultät für Psychologie an der Universität Wien entwickelt wurde, und sich an einschlägigen derzeit am Markt befindlichen Matrizentests für Kinder und Jugendliche orientiert.
- Die Untertests „*Klassifikationen*“, „*Ähnlichkeiten*“ und „*Matrizen*“ des **CFT 1** (*Grundintelligenztest Skala 1, Cattell, Weiß und Osterland, 1997*). Der Untertest „*Klassifikationen*“ erfasst das beziehungsstiftende Denken bei figuralem Material. Anhand des Untertests „*Ähnlichkeiten*“ wird erhoben, bis zu welchem Komplexitätsgrad figurale Vorgaben wieder erkannt werden können, wenn diese Figuren im Kontext mit ähnlichen, jedoch merkmalsveränderten Figuren stehen. Beim letzten zum Einsatz kommenden Subtest „*Matrizen*“ wird die Fähigkeit

erfasst, Regeln und Zusammenhänge bei figuralen Problemstellungen zu erkennen.

Teilnehmende Schulen:

Die Direktion der Volksschule St. Ursula hat ihr prinzipielles, mündliches Einverständnis zur Durchführung der Studie gegeben:

Volksschule St. Ursula
Franz Asenbauer Gasse 49
A - 1230 Wien

Zu weiteren Volksschulen wurde bereits Kontakt aufgenommen.

Stichprobenzusammensetzung:

In dieser Studie sollen insgesamt 120 Kinder im Alter von sechs bis neun Jahren getestet werden.

Genehmigung des Stadtschulrates



Carina Heuschober
Große Mohrengasse 21/21
1020 Wien

Ihr Zeichen, Ihre Nachricht von -----	Unser Zeichen/GZ 100.015/0067-kanz1/2010	Sachbearbeiter: Dipl. Päd. Elisabeth Kugler Elisabeth.Kugler@ssr-wien.gv.at	Tel: 52525 DW: 77125 Fax: 9977125	Datum 29.11.2010
---	---	---	---	---------------------

Sehr geehrte Frau Heuschober!

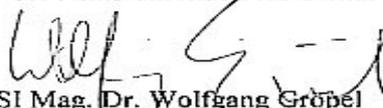
Der Stadtschulrat für Wien erteilt Ihnen die Genehmigung, im Rahmen Ihrer Diplomarbeit eine Erhebung mit dem Arbeitstitel „Studie zur Erfassung wesentlicher kognitiver Operationen bei der Lösung von Aufgaben zum logisch-schlussfolgerndem Denken im Kindesalter (6-9 Jahre)“ an den genannten Wiener Schulen, die bis längstens Ende Juni 2011 abgeschlossen sein muss, durchzuführen.

Die Ergebnisse unterliegen der Anonymität und dürfen nur für das Forschungsprojekt Verwendung finden. Außerdem möchte ich Sie daran erinnern, dass das Einvernehmen mit der betroffenen Direktion herzustellen ist und eine Mitarbeit der Lehrer/innen freiwillig sein muss.

Die Bewilligung der Untersuchung ist an die Bedingung geknüpft, dass eine Zusammenfassung der Arbeit dem Stadtschulrat für Wien, mit Anführung obiger Geschäftszahl, zugesendet wird.

Die schriftliche Einverständniserklärung der Eltern der zu untersuchenden Kinder muss vorliegen.

Mit freundlichen Grüßen
Für die Amtsführende Präsidentin


LSI Mag. Dr. Wolfgang Gröpel
Abteilungsleiter APS

Nachrichtlich an:
Brigitte Buschek

VS 23 St. Ursula, Franz Asenbauer Gasse 49

Die **Direktionen** werden gebeten, an den betroffenen Schulen die Mitglieder des Schulforums bzw. die KlassenelternvertreterInnen von der Durchführung der Erhebung zu informieren.

Elternbrief



Johann Ecker
Diplomand an der Fakultät für Psychologie

Universität Wien
E-Mail: a0404089@unet.univie.ac.at

Tel.: +43 (0) 699 1922 5646

Carina Heuschober
Diplomandin an der Fakultät für
Psychologie

Universität Wien
E-Mail: carina.heuschober@st.ursula-wien.at

Tel.: +43 (0) 660 7399225

Wien, Oktober 2010

Liebe Eltern!

Viele Kinder werden im Laufe ihrer schulischen Karriere aus den unterschiedlichsten Gründen mit psychologischen Tests untersucht. Im Rahmen schulpsychologischer Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Schullaufbahnberatungen, Verhaltensprobleme im schulischen Kontext, Aufmerksamkeitsprobleme etc. kommt dabei meist u.a. ein Verfahren zur Erfassung des logisch-schlussfolgernden Denkens zum Einsatz. Diese Verfahren sind zumeist mit einem *multiple-choice* Antwortformat gestaltet, das sich dadurch auszeichnet, dass die Kinder aus einer Reihe von verfügbaren Antwort-Alternativen jene auswählen, die sie für richtig halten. Dem gegenüber stehen psychologisch-diagnostische Instrumente, bei denen ohne eine Auswahl von Antwortalternativen, die Lösung der Aufgabe von den Kindern gefunden werden muss – das sogenannte freie Antwortformat.

Im Rahmen unserer Diplomarbeit an der Fakultät für Psychologie, Universität Wien (betreut von Dr. Stefana Holocher-Ertl), führen wir nun in den nächsten Wochen mit Einverständnis der Direktion und der Lehrkräfte in der Klasse Ihres Kindes eine Testung mit einigen Aufgaben aus Tests zur Erfassung des logisch-schlussfolgernden Denkens durch.

Ziel dieser Untersuchung ist es, herauszufinden, ob Leistungsunterschiede in Abhängigkeit von den Antwortformaten *multiple-choice* und *freie Antwort* bestehen. Forschungen dieser Art sind wichtig, um den Standard psychologischer Diagnostik laufend zu verbessern und dem neuesten Stand der Wissenschaft anpassen zu können.

Wir wenden uns daher mit der Bitte an Sie, Ihr Kind an diesem Projekt teilnehmen zu lassen, vorausgesetzt natürlich, Ihr Kind ist einverstanden. Die Untersuchung findet zu zwei Zeitpunkten während der Schulzeit statt und dauert ca. jeweils eine Schulstunde. Erfahrungsgemäß macht den Kindern die Mitarbeit an den Aufgaben viel Spaß.

Für Rückfragen stehen wir Ihnen gerne telefonisch oder per E-Mail zur Verfügung:

Carina Heuschober

E-Mail: carina.heuschober@st.ursula-wien.at

Tel.: +43 (0) 650 417 19 17

Johann Ecker

E-Mail: a0404089@unet.univie.ac.at

Tel.: +43 (0) 699 1922 5646

Die gewonnen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke genutzt. Sämtliche Ergebnisse der Schüler(innen) werden (noch während der Testung) anonymisiert.

Wir bitten Sie, mit Ihrer Unterschrift auf dem beiliegenden Formular, Ihr Einverständnis zur Teilnahme Ihres Kindes an der oben beschriebenen Untersuchung zu erteilen.

Mit freundlichen Grüßen und herzlichen Dank im Voraus!

Carina Heuschober & Johann Ecker

.....bitte hier abtrennen.....

Ich erkläre mich mit der Teilnahme meiner Tochter/meines Sohnes _____, an der *Testung für die Diplomarbeit von Carina Heuschober und Johann Ecker* einverstanden.

.....

Datum, Unterschrift

Johann Ecker
Diplomand an der Fakultät für Psychologie

Universität Wien
E-Mail: a0404089@unet.univie.ac.at

Tel.: +43 (0) 699 1922 5646

Carina Heuschober
Diplomandin an der Fakultät für
Psychologie

Universität Wien
E-Mail: carina.heuschober@st.ursula-
wien.at

Tel.: +43 (0) 660 7399225

Wien, Oktober 2010

Liebe Lehrende!

Viele Kinder werden im Laufe ihrer schulischen Karriere aus den unterschiedlichsten Gründen mit psychologischen Tests untersucht. Im Rahmen schulpsychologischer Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Schullaufbahnberatungen, Verhaltensprobleme im schulischen Kontext, Aufmerksamkeitsprobleme etc. kommt dabei meist u.a. ein Verfahren zur Erfassung des logisch- schlussfolgernden Denkens zum Einsatz. Diese Verfahren sind zumeist mit einem multiple-choice Antwortformat gestaltet, das sich dadurch auszeichnet, dass die Kinder aus einer Reihe von verfügbaren Antwort-Alternativen jene auswählen, die sie für richtig halten. Dem gegenüber stehen psychologisch-diagnostische Instrumente, bei denen ohne eine Auswahl von Antwortalternativen, die Lösung der Aufgabe von den Kindern gefunden werden muss – das sogenannte freie Antwortformat.

Im Rahmen unserer Diplomarbeit an der Fakultät für Psychologie, Universität Wien (betreut von Dr. Stefana Holocher-Ertl), führen wir nun in den nächsten Wochen mit Einverständnis der Direktion in Ihrer Schule eine Testung mit einigen Aufgaben aus Verfahren zur Erfassung des logischen Schlussfolgerns (PSB, Horn, W., 2003; AID-3 LB, Kubinger, 2009; Matrizentest in Erprobung; CFT 1, Cattell, Weiß und Osterland, 1997) durch.

Ziel dieser Untersuchung ist es, herauszufinden, ob Leistungsunterschiede je nach vorgegebenem Antwortformat (multiple-choice vs. freies Antwortformat) bestehen.

Forschungen dieser Art sind wichtig, um den Standard psychologischer Diagnostik laufend zu verbessern und dem neuesten Stand der Wissenschaft anpassen zu können.

Wir wenden uns daher mit der Bitte an die Eltern Ihrer Schüler(innen), diese an der Untersuchung teilnehmen zu lassen, vorausgesetzt natürlich, dass die Kinder damit einverstanden sind. Die Untersuchung findet zu zwei Zeitpunkten während der Schulzeit statt und dauert ca. jeweils eine Schulstunde. Erfahrungsgemäß macht den Kindern die Mitarbeit an den Aufgaben viel Spaß.

Wir möchten Sie bitten, die Elternbriefe an Ihre Schüler(innen) zu übergeben und die Rückmeldungen der Eltern einzusammeln. Wenn sich Eltern Ihrer Schüler(innen) mit der Teilnahme an der Untersuchung einverstanden erklären, werden wir uns bei der Terminvereinbarung sehr bemühen, den Ablauf Ihres Unterrichts so wenig wie möglich zu stören.

Für eventuelle Rückfragen erreichen Sie uns gerne telefonisch oder per E-Mail:

Johann Ecker

E-Mail: a0404089@unet.univie.ac.at

Tel.: +43 (0) 699 1922 5646

Carina Heuschober

E-Mail: carina.heuschober@st.ursula-wien.at

Tel.: +43 (0) 650 417 19 17

Die gewonnen Daten werden im Sinne des Datenschutzes ausschließlich für wissenschaftliche Zwecke genutzt. Sämtliche Ergebnisse der Schüler(innen) werden (noch während der Testung) anonymisiert.

Mit freundlichen Grüßen und herzlichen Dank im Voraus!

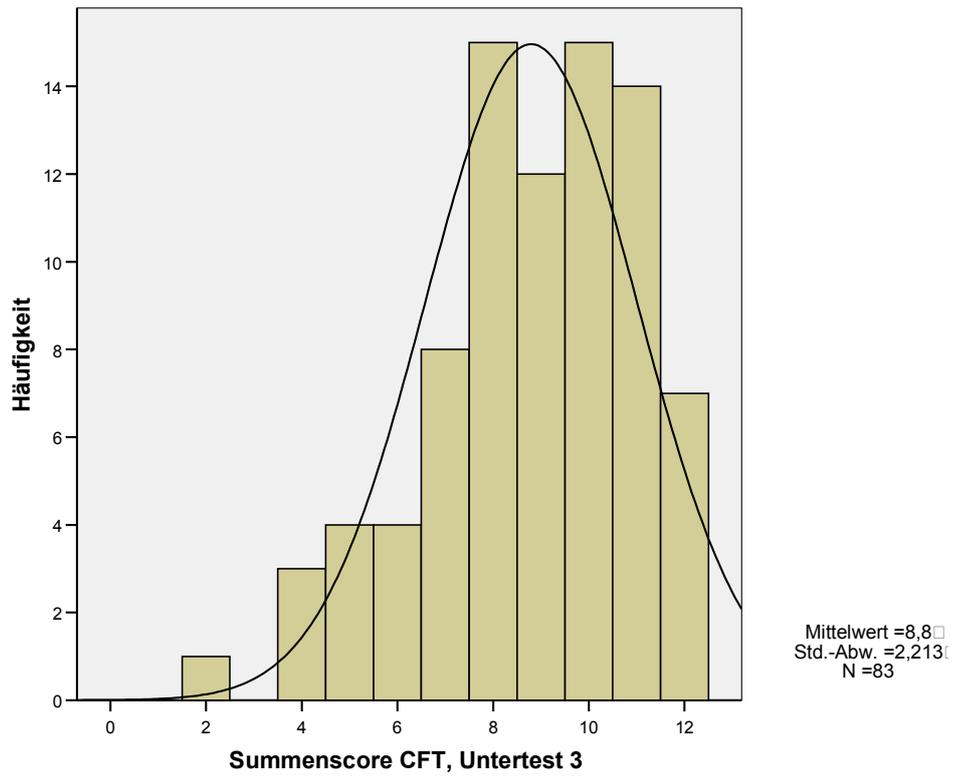
Carina Heuschober & Johann Ecker

Alter der Kinder

Alter der Kinder

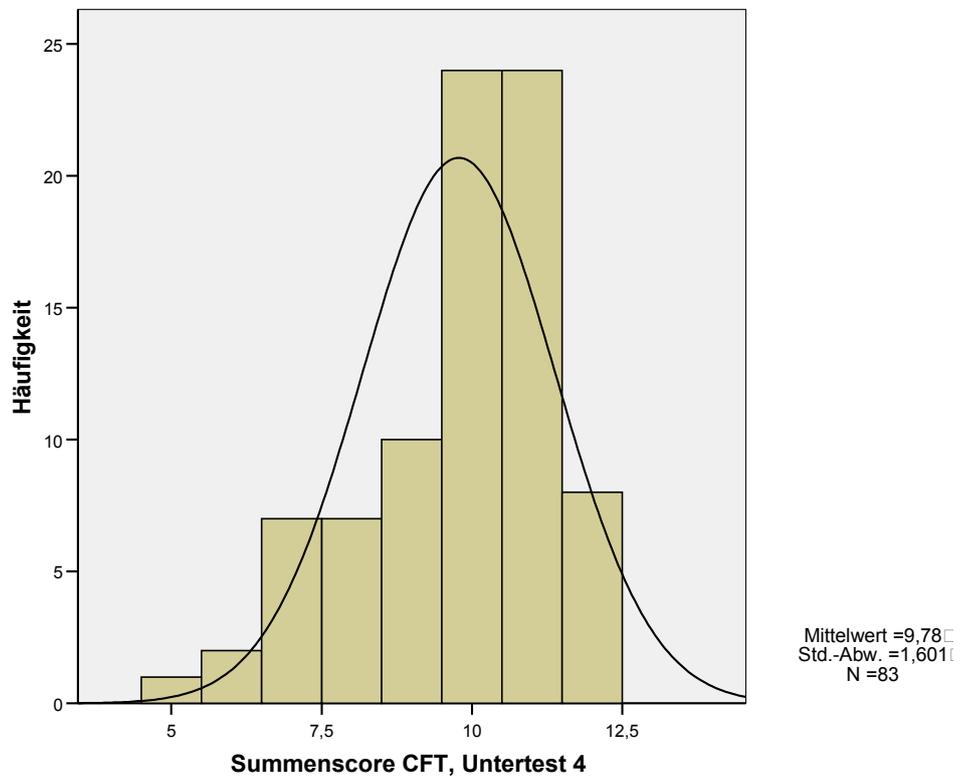
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	6,08	4	4,8	4,8	4,8
	6,42	1	1,2	1,2	6,0
	6,50	2	2,4	2,4	8,3
	6,58	2	2,4	2,4	10,7
	6,67	1	1,2	1,2	11,9
	6,75	3	3,6	3,6	15,5
	6,92	1	1,2	1,2	16,7
	7,08	6	7,1	7,1	23,8
	7,17	2	2,4	2,4	26,2
	7,25	1	1,2	1,2	27,4
	7,42	1	1,2	1,2	28,6
	7,50	3	3,6	3,6	32,1
	7,58	3	3,6	3,6	35,7
	7,67	3	3,6	3,6	39,3
	7,75	3	3,6	3,6	42,9
	7,92	1	1,2	1,2	44,0
	8,00	1	1,2	1,2	45,2
	8,08	3	3,6	3,6	48,8
	8,25	1	1,2	1,2	50,0
	8,33	2	2,4	2,4	52,4
	8,42	8	9,5	9,5	61,9
	8,50	4	4,8	4,8	66,7
	8,58	2	2,4	2,4	69,0
	8,67	5	6,0	6,0	75,0
	8,75	6	7,1	7,1	82,1
	8,92	3	3,6	3,6	85,7
	9,00	3	3,6	3,6	89,3
	9,08	1	1,2	1,2	90,5
	9,17	3	3,6	3,6	94,0
	9,25	2	2,4	2,4	96,4
	9,33	1	1,2	1,2	97,6
	9,50	1	1,2	1,2	98,8
	9,75	1	1,2	1,2	100,0
	Gesamt	84	100,0	100,0	

Rohscoreverteilung des Untertests drei des CFT

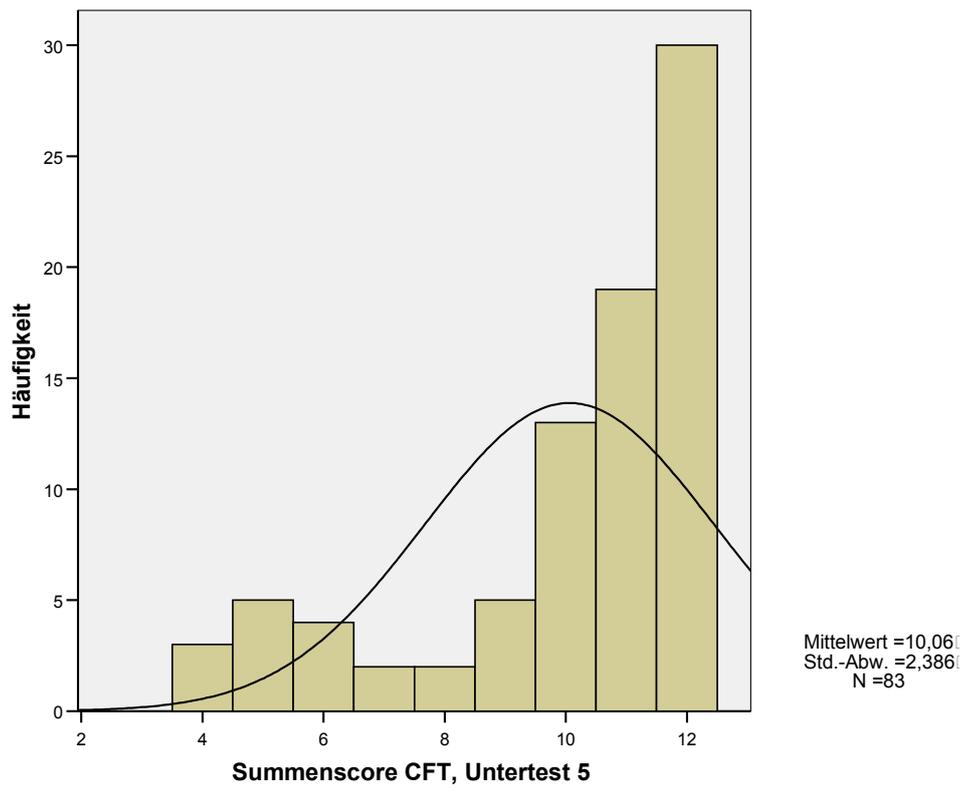


1

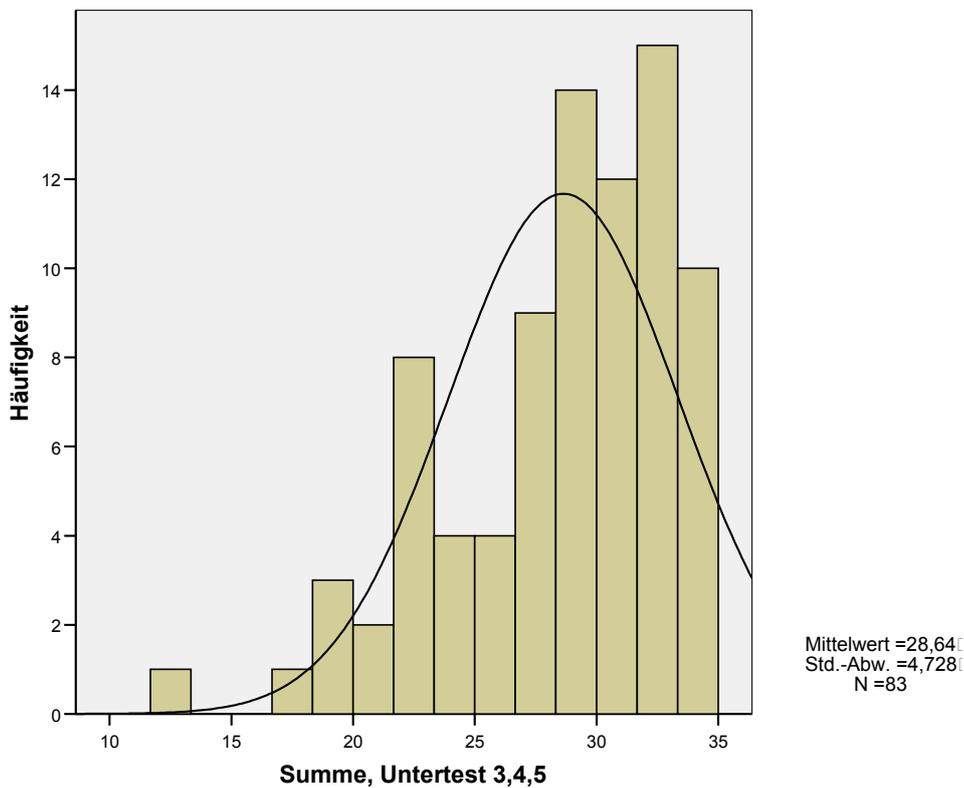
Rohscoreverteilung des Untertests vier des CFT 1



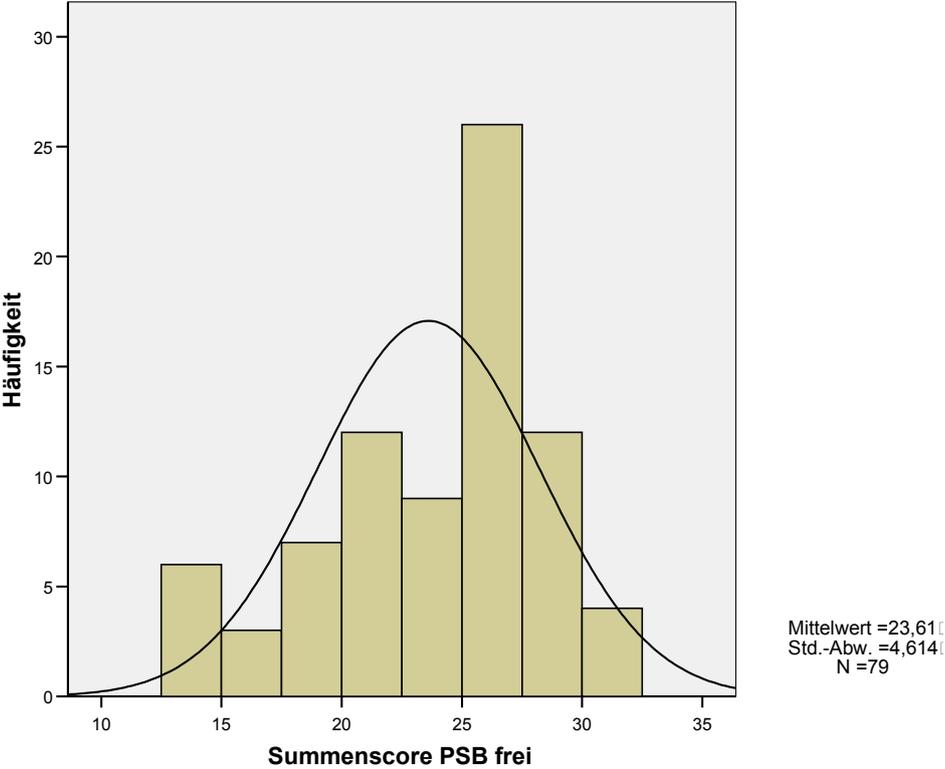
Rohscoreverteilung des Untertests fünf des CFT 1



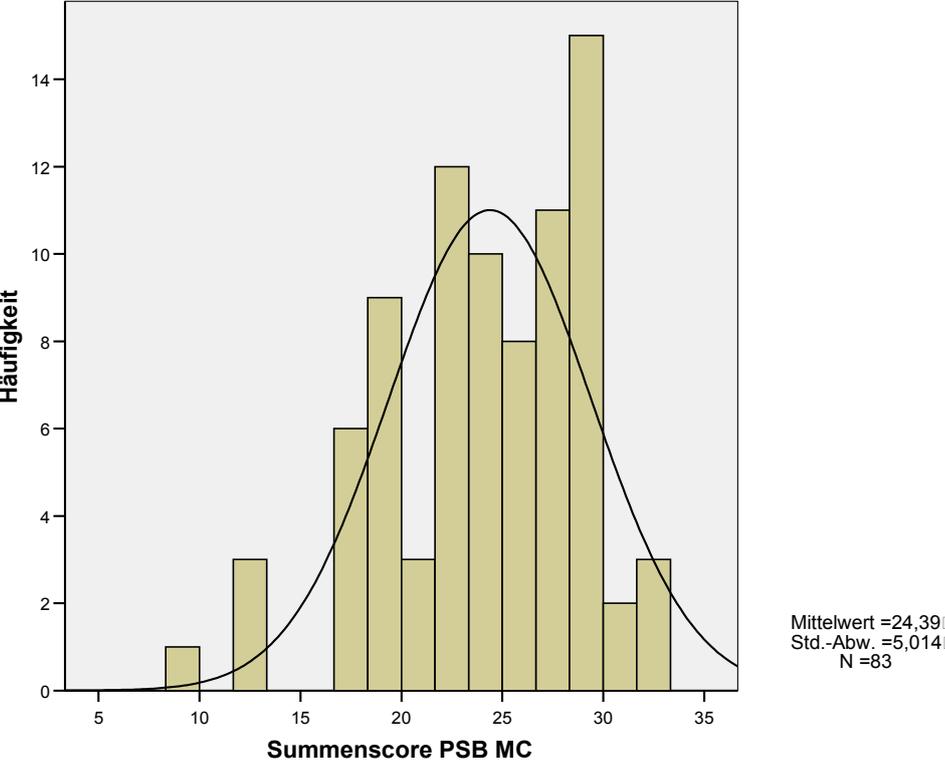
Rohscoreverteilung des Summenwertes über die Untertest drei bis vier des CFT 1



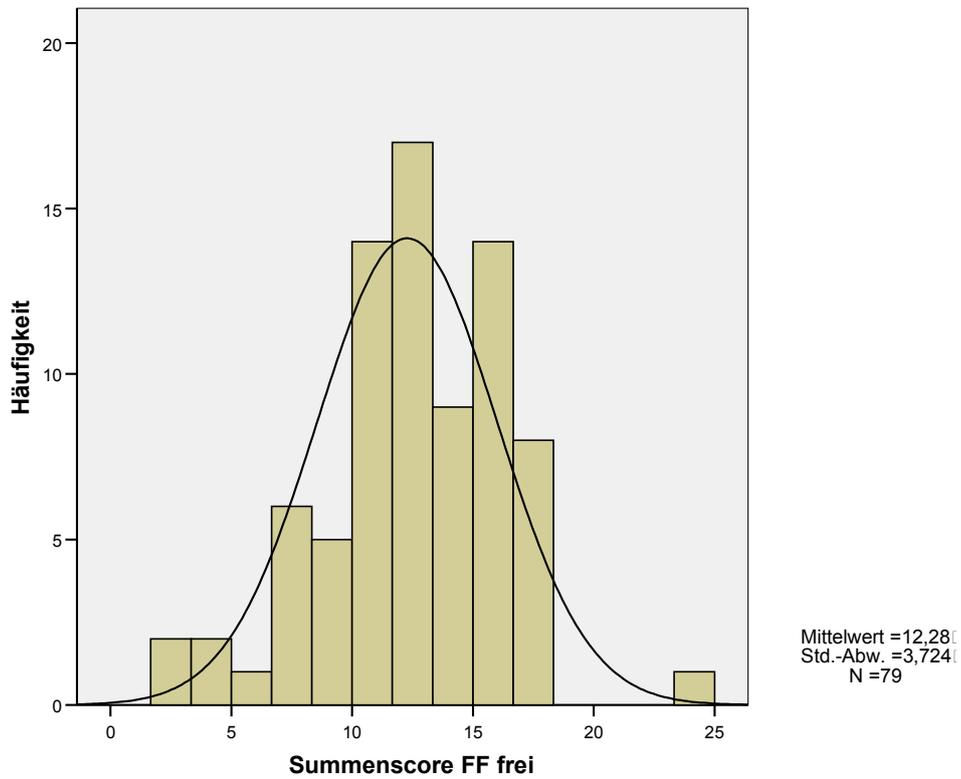
Rohscoreverteilung des Untertests drei des PSB im freien Antwortformat



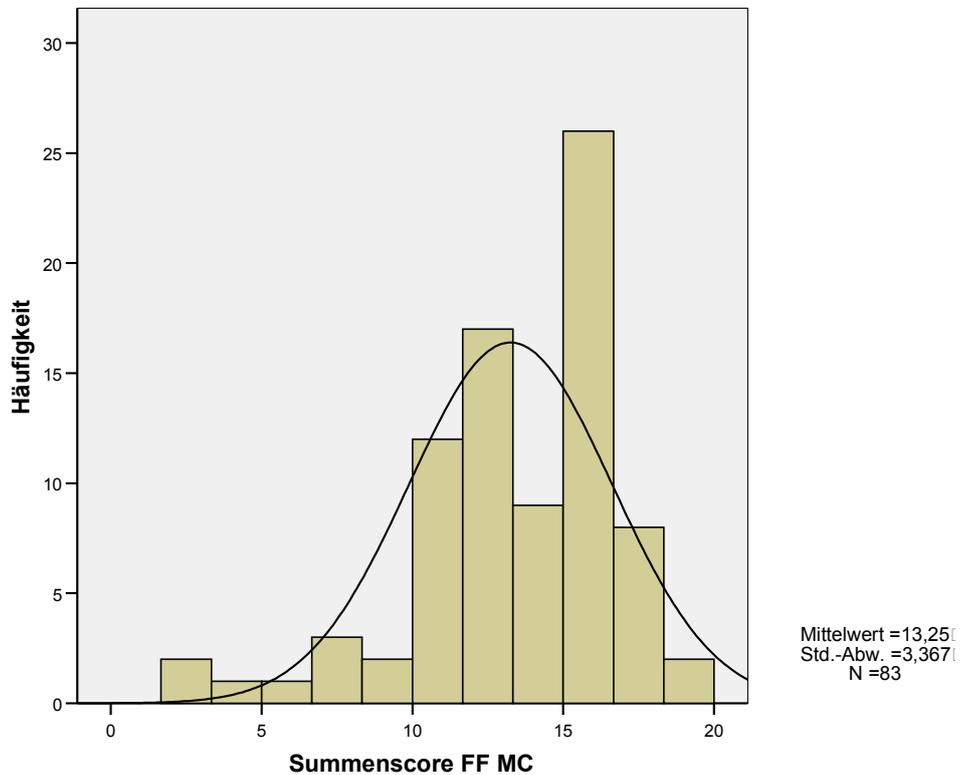
Rohscoreverteilung des Untertests drei des PSB im Multiple-Choice Antwortformat



Rohscoreverteilung von Formale Folgerichtigkeit im freien Antwortformat



Rohscoreverteilung von Formale Folgerichtigkeit im Multiple-Choice Antwortformat



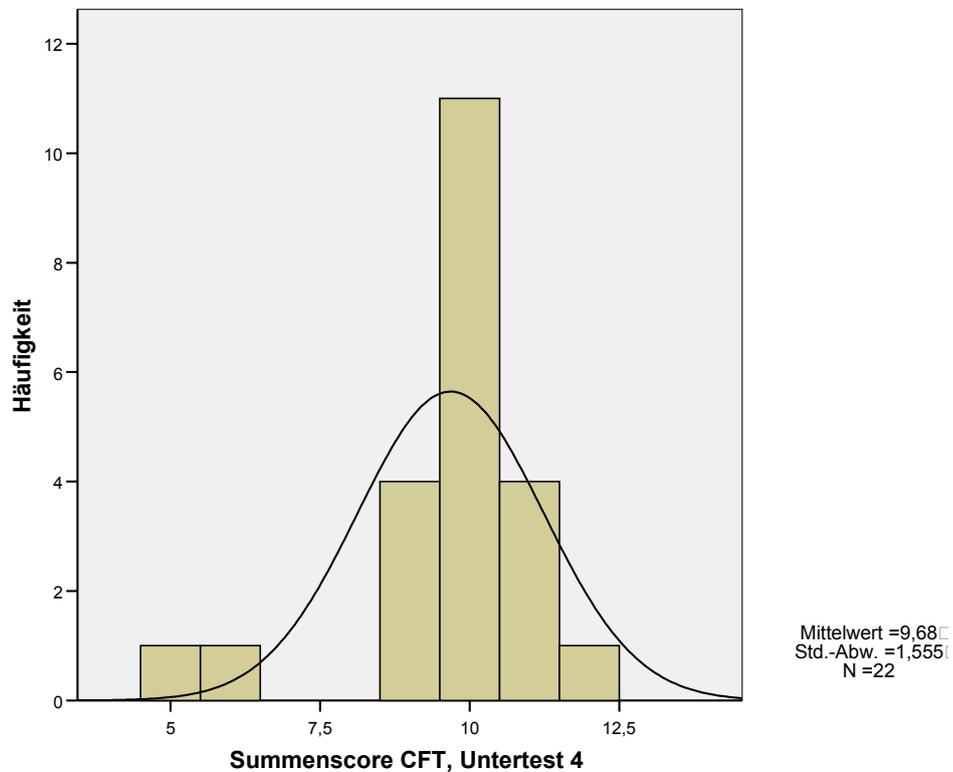
Rohscoreverteilung des Untertests 4 des CFT 1 – 1a-Klasse

Kolmogorov-Smirnov-Anpassungstest

		Summenscor e CFT, Untertest 4
N		22
Parameter der Normalverteilung(a,b)	Mittelwert	9,68
	Standardabweichung	1,555
Extremste Differenzen	Absolut	,308
	Positiv	,192
	Negativ	-,308
Kolmogorov-Smirnov-Z		1,446
Asymptotische Signifikanz (2-seitig)		,030

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.



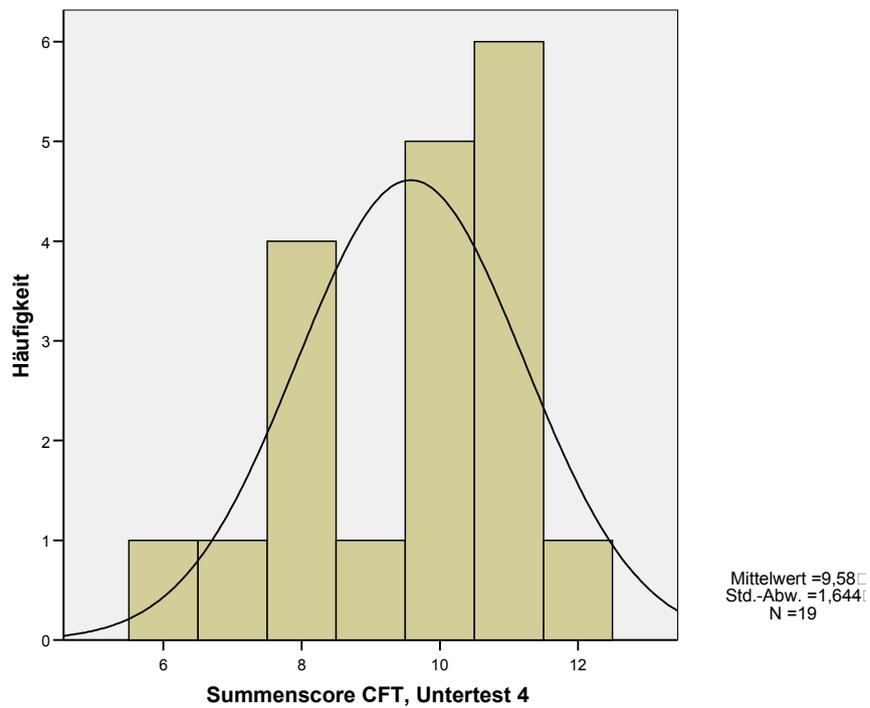
Rohscoreverteilung des Untertests 4 des CFT 1 - 2a-Klasse

Kolmogorov-Smirnov-Anpassungstest

		Summenscor e CFT, Untertest 4
N		19
Mittelwert		9,58
Parameter der Normalverteilung(a,b)	Standardabweichung	1,644
Extremste Differenzen	Absolut	,233
	Positiv	,147
	Negativ	-,233
Kolmogorov-Smirnov-Z		1,014
Asymptotische Signifikanz (2-seitig)		,255

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.



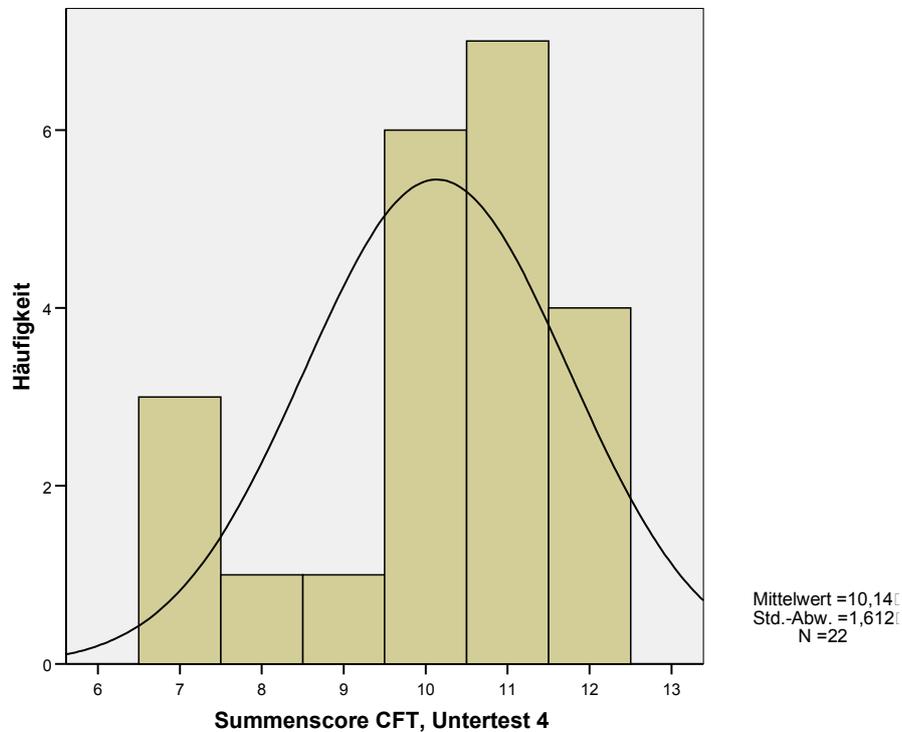
Rohscoreverteilung des Untertests 4 des CFT 1 – 3a-Klasse

Kolmogorov-Smirnov-Anpassungstest

		Summenscor e CFT, Untertest 4
N		22
Parameter der Normalverteilung(a,b)	Mittelwert	10,14
	Standardabweichung	1,612
Extremste Differenzen	Absolut	,239
	Positiv	,124
	Negativ	-,239
Kolmogorov-Smirnov-Z		1,121
Asymptotische Signifikanz (2-seitig)		,162

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.



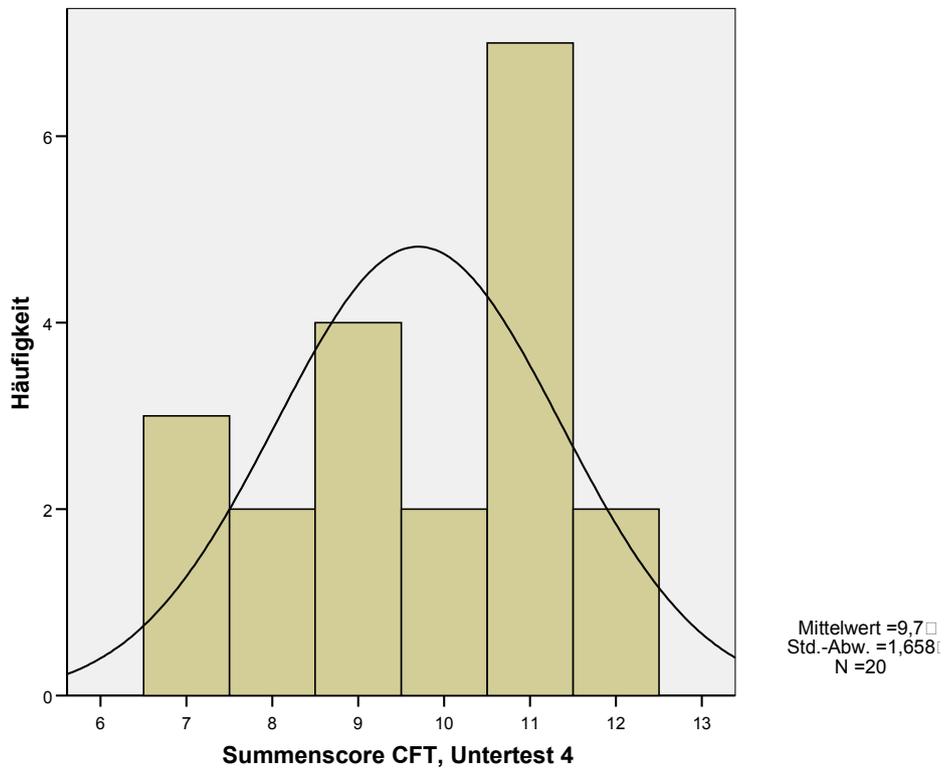
Rohscoreverteilung des Untertests 4 des CFT 1 – 3b-Klasse

Kolmogorov-Smirnov-Anpassungstest

		Summenscor e CFT, Untertest 4
N		20
Mittelwert		9,70
Parameter der Normalverteilung(a,b)	Standardabweichung	1,658
Extremste Differenzen	Absolut	,234
	Positiv	,116
	Negativ	-,234
Kolmogorov-Smirnov-Z		1,045
Asymptotische Signifikanz (2-seitig)		,225

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

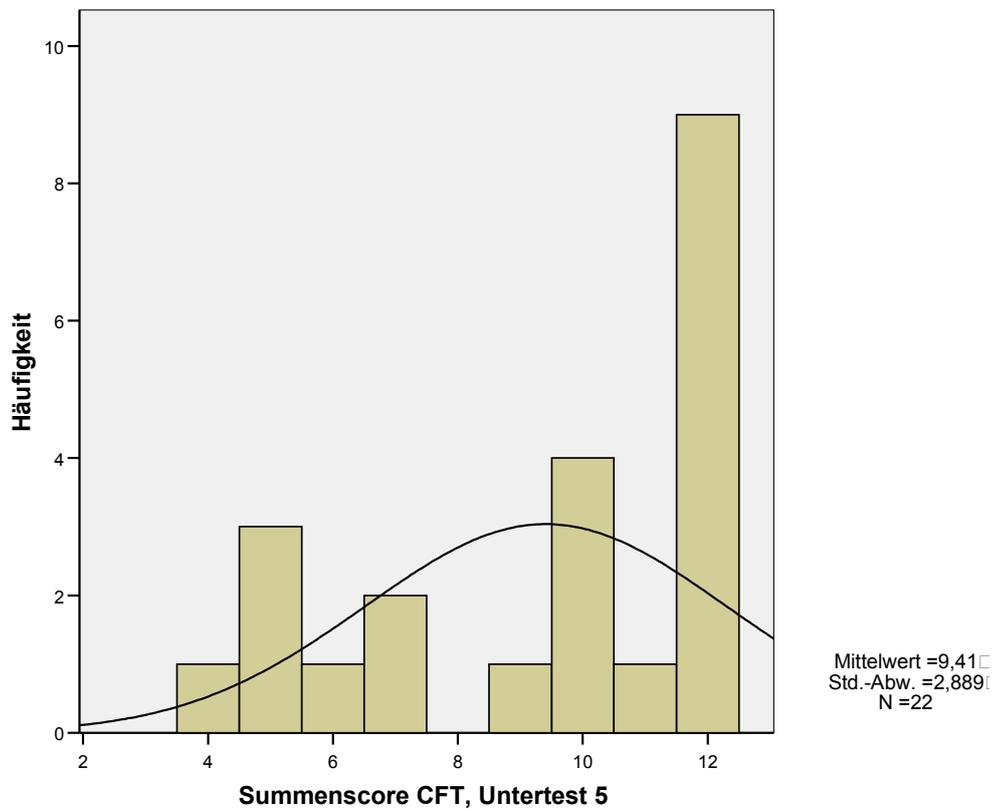


Rohscoreverteilung des Untertests 5 des CFT 1 – 1a-Klasse

Kolmogorov-Smirnov-Anpassungstest

		Summenscor e CFT, Untertest 5
N		22
Mittelwert		9,41
Parameter der Normalverteilung(a,b)	Standardabweichung	2,889
Extremste Differenzen	Absolut	,224
	Positiv	,185
	Negativ	-,224
Kolmogorov-Smirnov-Z		1,051
Asymptotische Signifikanz (2-seitig)		,219

a Die zu testende Verteilung ist eine Normalverteilung.
b Aus den Daten berechnet.



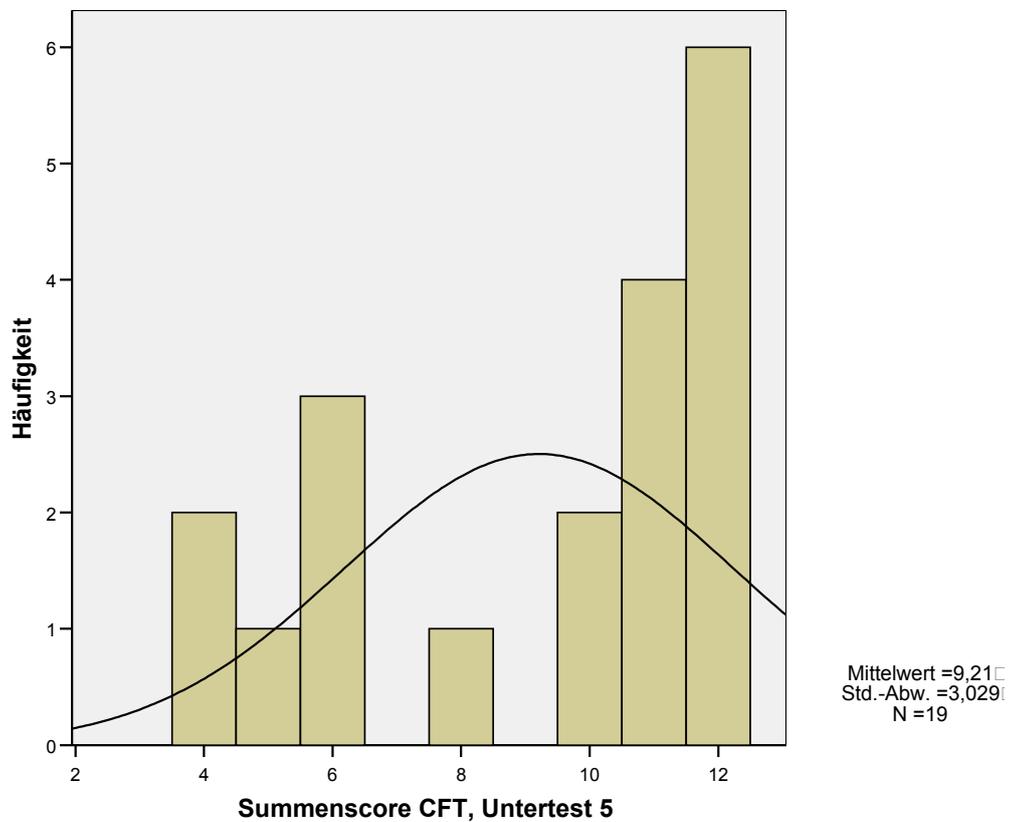
Rohscoreverteilung des Untertests 5 des CFT 1 – 2a-Klasse

Kolmogorov-Smirnov-Anpassungstest

		Summenscor e CFT, Untertest 5
N		19
Mittelwert		9,21
Parameter der Normalverteilung(a,b)	Standardabweichung	3,029
Extremste Differenzen	Absolut	,249
	Positiv	,179
	Negativ	-,249
Kolmogorov-Smirnov-Z		1,085
Asymptotische Signifikanz (2-seitig)		,190

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

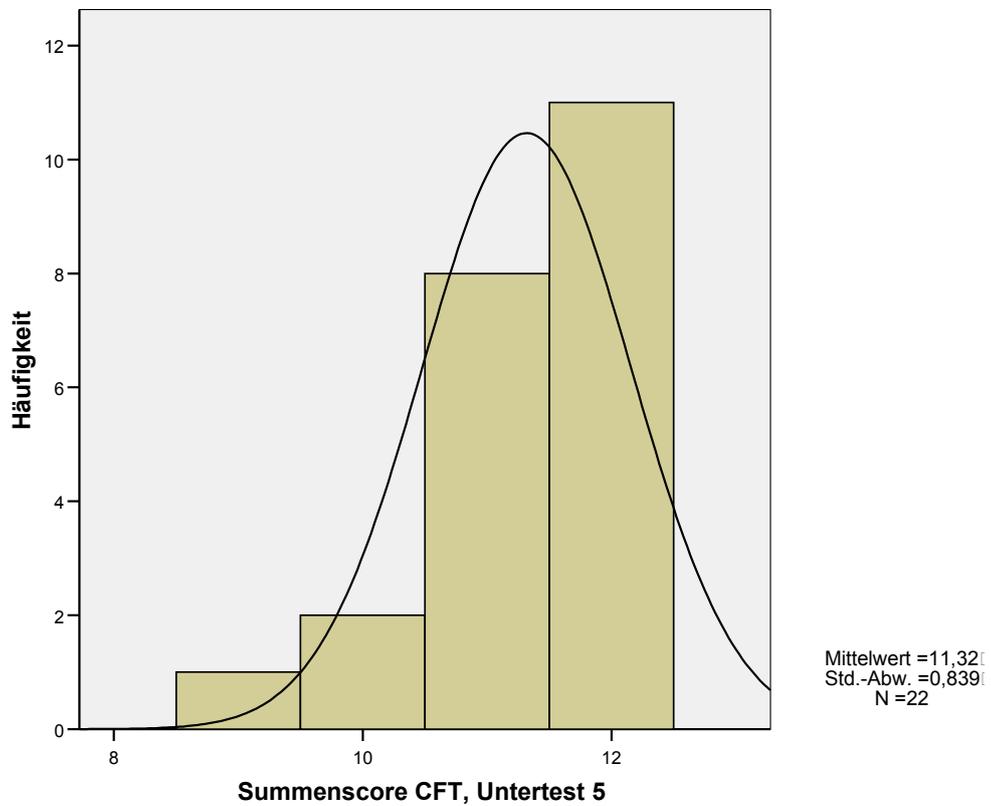


Rohscoreverteilung des Untertests 5 des CFT 1 – 3a-Klasse

Kolmogorov-Smirnov-Anpassungstest

		Summenscor e CFT, Untertest 5
N		22
Parameter der Normalverteilung(a,b)	Mittelwert	11,32
	Standardabweichung	,839
Extremste Differenzen	Absolut	,292
	Positiv	,208
	Negativ	-,292
Kolmogorov-Smirnov-Z		1,369
Asymptotische Signifikanz (2-seitig)		,047

a Die zu testende Verteilung ist eine Normalverteilung.
b Aus den Daten berechnet.



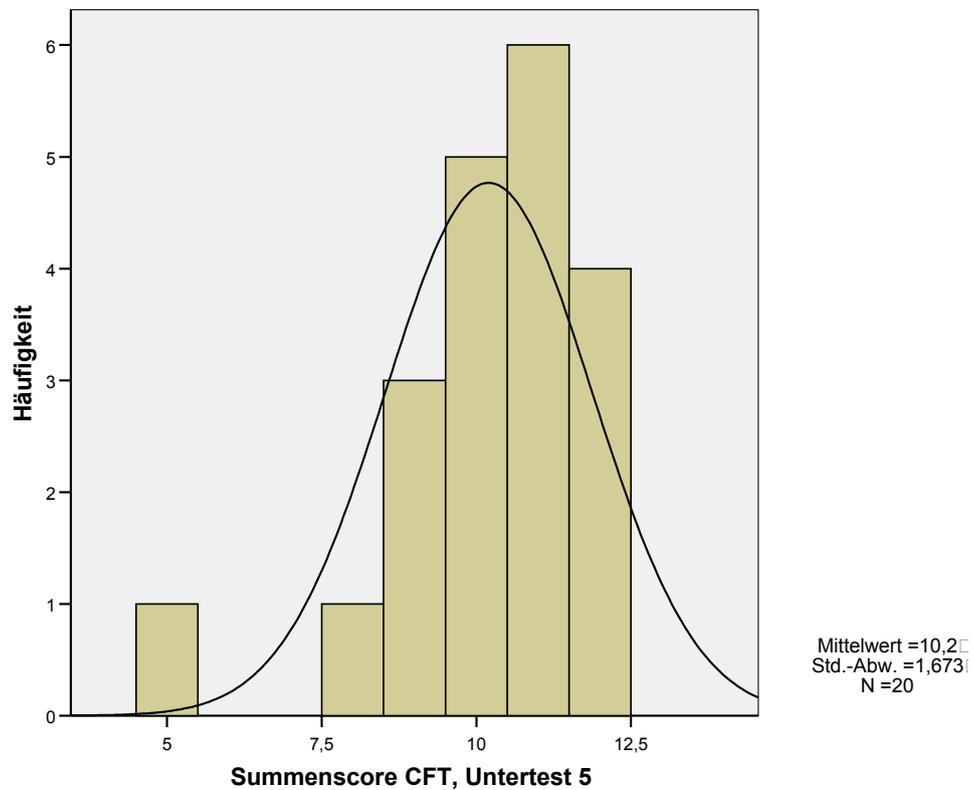
Rohscoreverteilung des Untertests 5 des CFT 1 – 3b-Klasse

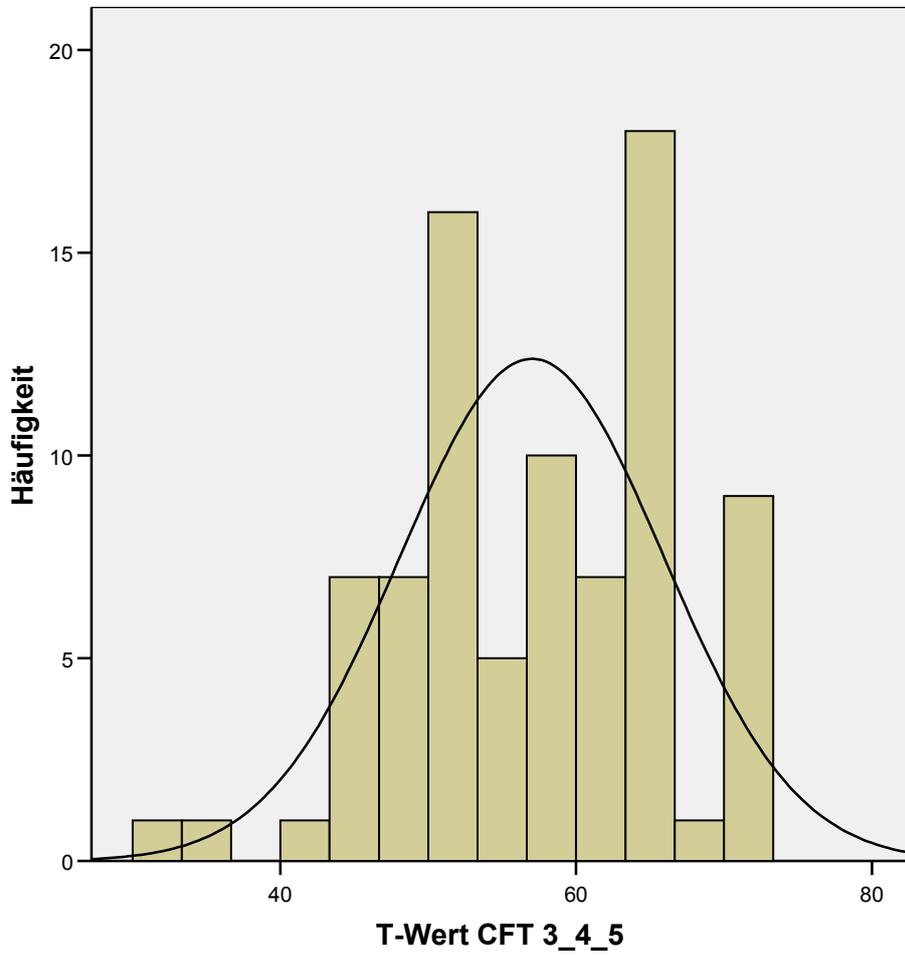
Kolmogorov-Smirnov-Anpassungstest

		Summenscor e CFT, Untertest 5
N		20
Parameter der Normalverteilung(a,b)	Mittelwert	10,20
	Standardabweichung	1,673
Extremste Differenzen	Absolut	,202
	Positiv	,141
	Negativ	-,202
Kolmogorov-Smirnov-Z		,905
Asymptotische Signifikanz (2-seitig)		,385

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.





Mittelwert =57,01
Std.-Abw. =8,912
N =83

Interkorrelationen aller Verfahren und Noten

Korrelationen

	Deutschnote	Mathe-note	Gesamt-note	CFT UT 3	CFT UT 4	CFT UT 5	CFT UT 3,4,5	FF frei	FF MC	PSB frei	PSB MC
Spearman-Rho											
Deutschnote im Semesterzeugnis	1,000	,135	.	-,084	-,137	,009	-,058	-,232(*)	-,047	-,170	-,117
		,153	.	,263	,149	,473	,329	,044	,360	,108	,187
	60	60	0	60	60	60	60	55	60	55	60
Mathematiknote im Semesterzeugnis	,135	1,000	.	-,264(*)	-,054	-,095	-,199	-,237(*)	-,009	-,085	,055
	,153	.	.	,021	,342	,234	,064	,041	,474	,268	,337
	60	60	0	60	60	60	60	55	60	55	60
Gesamtnote der 1. Klasse im Semesterzeugnis	.	.	1,000	,000	-,133	-,299	-,197	-,330	-,450(*)	,130	-,429(*)
	.	.	.	,499	,277	,088	,190	,067	,020	,283	,026
	0	0	22	22	22	22	22	22	21	22	21
Summenscore CFT, Untertest 3	-,084	-,264(*)	,000	1,000	,288(**)	,342(**)	,750(**)	,345(**)	,289(**)	,562(**)	,519(**)
	,263	,021	,499	.	,004	,001	,000	,001	,004	,000	,000
	60	60	22	83	83	83	83	78	82	78	82
Summenscore CFT, Untertest 4	-,137	-,054	-,133	,288(**)	1,000	,361(**)	,715(**)	,337(**)	,164	,321(**)	,362(**)
	,149	,342	,277	,004	.	,000	,000	,001	,071	,002	,000
	60	60	22	83	83	83	83	78	82	78	82
Summenscore CFT, Untertest 5	,009	-,095	-,299	,342(**)	,361(**)	1,000	,730(**)	,320(**)	,301(**)	,269(**)	,409(**)
	,473	,234	,088	,001	,000	.	,000	,002	,003	,009	,000
	60	60	22	83	83	83	83	78	82	78	82
Summenscore CFT, Untertest 3,4,5	-,058	-,199	-,197	,750(**)	,715(**)	,730(**)	1,000	,451(**)	,298(**)	,508(**)	,562(**)
	,329	,064	,190	,000	,000	,000	.	,000	,003	,000	,000
	60	60	22	83	83	83	83	78	82	78	82
Summenscore FF frei	-,232(*)	-,237(*)	-,330	,345(**)	,337(**)	,320(**)	,451(**)	1,000	,320(**)	,405(**)	,360(**)
	,044	,041	,067	,001	,001	,002	,000	.	,002	,000	,001
	55	55	22	78	78	78	78	79	78	79	78
Summenscore FF MC	-,047	-,009	-,450(*)	,289(**)	,164	,301(**)	,298(**)	,320(**)	1,000	,334(**)	,567(**)
	,360	,474	,020	,004	,071	,003	,003	,002	.	,001	,000
	60	60	21	82	82	82	82	78	83	78	83
Summenscore PSB frei	-,170	-,085	,130	,562(**)	,321(**)	,269(**)	,508(**)	,405(**)	,334(**)	1,000	,629(**)
	,108	,268	,283	,000	,002	,009	,000	,000	,001	.	,000
	55	55	22	78	78	78	78	79	78	79	78
Summenscore PSB MC	-,117	,055	-,429(*)	,519(**)	,362(**)	,409(**)	,562(**)	,360(**)	,587(**)	,629(**)	1,000
	,187	,337	,026	,000	,000	,000	,000	,001	,000	,000	.
	60	60	21	82	82	82	82	78	83	78	83

* Die Korrelation ist auf dem 0,05 Niveau signifikant (einseitig).
 ** Die Korrelation ist auf dem 0,01 Niveau signifikant (einseitig).

Reliabilität von Formale Folgerichtigkeit

	Cronbachs Alpha
Formale Folgerichtigkeit freie Antwort	.666
Formale Folgerichtigkeit Multiple-Choice	.704

Reliabilität von Untertest drei des PSB

	Cronbachs Alpha
PSB Untertest 3 freie Antwort	.802
PSB Untertest 3 Multiple-Choice	.811

LEBENS LAUF

Carina Heuschober, geboren am 13.05.1980

Ausbildung

2007/2008	Erasmusaufenthalt an der Universidad Autonoma in Madrid, Spanien
seit 10/2003	Diplomstudium Psychologie an der Universität Wien
2000 - 2002	Sozialpädagogisches Kolleg der Diözese Linz
1995 - 2000	Bundesgymnasium Steyr, mit Matura abgeschlossen
1990 - 1995	Stiftsgymnasium Kremsmünster
1986 - 1990	Volksschule Bad Hall

Praktika

08-09/2008	6-Wochen Praktikum im psychosozialen Rehasentrum Sonnenpark, Bad Hall
03-04/2002	Praktikum bei der Mobilien Familienbegleitung Mobilis, Wels
11/2001	Praktikum bei der Sozialpädagogischen Familienbegleitung des Jugendamtes Steyr
08/2001	Praktikum in der sozialpädagogischen Wohngruppe Heidlmair, Kremsmünster
03/2001	Praktikum im Schülerwohnheim Steyr
10-11/2000	Praktikum im Kinderhort Allendeplatz, Linz

Berufserfahrung

Seit 2/2012	stellvertretende Tagesheimleiterin und Gruppenerzieherin der Schule St. Ursula, Wien
12/2008 – 02/2011	Gruppenerzieherin in der KMS und AHS St. Ursula, Wien
01/2004 – 06/2007	Telefoninterviewerin bei Statistik Austria, Wien
11/2002 – 07/2003	Erzieherin im Schülerwohnheim Steyr
05/2002 – 02/2003	Familienbetreuerin bei Mobilis in Wels
01-07/2002	Kinderbetreuerin bei der Familienbegleitung, Steyr