



universität
wien

Diplomarbeit

Titel der Arbeit

Einzel-vs. Gruppenvorgabe der AID 3 Untertests
*Alltagswissen, Angewandtes Rechnen, Synonyme
Finden, Funktionen Abstrahieren und Soziales
Erfassen und Sachliches Reflektieren*

Verfasserin

Martina Scheffl

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im August 2011

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Univ.- Ass. Dr. Stefana Holocher-Ertl

Danksagung

An erster Stelle möchte ich meiner Familie danken, die mich während der gesamten Studienzeit emotional als auch finanziell unterstützt haben.

Ich danke Univ.- Ass. Dr. Stefana Holoher-Ertl für die Zusammenarbeit. Sie waren immer erreichbar und für alle anstehenden Fragen sehr bemüht. Ich schätze Ihr Engagement und ihre Zuverlässigkeit.

Desweiteren bedanke ich mich bei den DirektorInnen, den LehrerInnen und den SchülerInnen der Volksschule Sierndorf und Stockerau, sowie des Gymnasiums in Stockerau und Hollabrunn, die mir die Durchführung der Testung ermöglichten.

Großer Dank gebührt meinem Lebenspartner Christian Zellner und meiner guten Freundin Mag. Judith Forstner für die große Unterstützung.

Weiters danke ich allen meinen Versuchsleitern, die mich tatkräftig bei meiner Untersuchung unterstützten.

Ausdrücklich bedanken möchte ich mich bei meinen Korrekturleserinnen Mag. Sabine Poisinger und Melanie Neunteufel.

Abstract (Deutsch)

In der vorliegenden Arbeit wurde untersucht, ob die Bearbeitung von Aufgaben aus einem Intelligenztest für Kinder und Jugendliche unter zwei verschiedenen Bedingungen gleich schwierig ist. Dabei wurden die Untertests „Alltagswissen“, „Angewandtes Rechnen“, „Synonyme Finden“, „Funktionen Abstrahieren“ und „Soziales Erfassen und Sachliches Reflektieren“, aus dem Adaptiven Intelligenz Diagnostikum 2 (Version 2.2., Kubinger, 2009a), im Einzelsetting und Gruppensetting vorgegeben. Untersucht wurden die Leistungsunterschiede von 65 Volksschülern und 89 Gymnasiasten je nach Testvorgabe.

Die Testpersonen mussten die 5 Untertests einmal in der Gruppe bearbeiten und einmal im Einzelsetting, wozu die Hälfte der Stichprobe zuerst im Einzelsetting und dann im Gruppensetting und die andere Hälfte umgekehrt, dieselben Aufgaben bearbeiteten. Gemessen wurde, ob es Unterschiede zwischen diesen beiden Testarten hinsichtlich der Testleistungen gibt. Im Mittelpunkt des Interesses stand die Schwierigkeit der Items für die Testpersonen unter diesen beiden Bedingungen.

In allen Untertests gab es einen Effekt beim Hauptfaktor „Testart“. Die Testpersonen schnitten in den Einzeltestungen immer besser ab als in den Gruppentestungen. Weiters konnte ein Effekt im Untertest „Angewandtes Rechnen“ hinsichtlich des Faktors „Gruppe“, unabhängig von der Testart gefunden werden. Hier waren die Versuchspersonen unter der Bedingung „Einzeltestung zuerst“ besser, als die Versuchspersonen in der Bedingung „Gruppentestung zuerst“. Ein zusätzlicher Effekt zeigte sich im Untertest „Funktionen Abstrahieren“, betreffend der Interaktion „Testart x Gruppe“.

Abstract (English)

The paper submitted examines the question if the doing of tasks from an intelligence test for children and adolescents is equally difficult under two different conditions. In doing so the fields of “general knowledge”, “practical calculating”, “finding synonyms”, “abstracting functions”, “social understanding and factual reflecting” from Adaptive Intelligence Diagnosticum Version 2.2 (Kubinger, 2009a) were specified in individual and group setting. The different performances of 65 elementary school children and of 89 grammar school students in corresponding tests were evaluated.

All the probands had to do the same tasks of the 5 fields, but they were split up in two halves, the one half of the sample working in individual setting first and then the group setting, the other half of the sample vice versa. I analyzed if the results differ in reference to the two testing methods. The focus of interest was on the difficulty of the items for the test persons under both conditions.

In all the fields significant differences concerning the main factor test setting could be found. Throughout the test the achievements of the individual tests were better than those of the tests done in a group. In addition to that, some effect could be seen in the field practical calculating as to the factor “group”, without regard to the factor “setting”. Hereby were the test persons under the condition “individual setting first” better than the test persons under the condition “group setting first”. An additional effect arose in the field abstracting functions concerning the interrelation between “test setting and group”.

Inhaltsverzeichnis

1 Einleitung	1
2 Definition des Begriffes Intelligenz	3
2.1 Die ersten Intelligenztests und die Entwicklung des Intelligenzquotienten	5
2.1.1 Die Arbeiten von Binet und Simon.....	5
2.1.2 Terman´s Gruppentest	7
2.1.3 Wechslers Intelligenzkonzept.....	7
2.2 Intelligenzmodelle	9
2.2.1 Zwei-Faktoren-Theorie nach Spearman.....	9
2.2.2 Primary Mental Abilities nach Thurstone.....	11
2.2.3 Fluide und Kristalline Intelligenz nach Cattell	13
2.2.4 Drei-Schichten-Modell nach Carroll.....	14
2.2.5 Cattell-Horn-Carroll-Theorie (CHC-Modell)	16
3 Darbietungsmöglichkeiten von diagnostischen Verfahren	18
3.1 Vergleiche von Gruppentestung und Einzeltestung	23
3.1.1 Vorteile der Gruppentestung	23
3.1.2.Nachteile der Gruppentestung.....	25
3.1.3 Vorteile der Individualtestung	26
3.1.4 Nachteile der Individualtestung	27
4 Empirischer Teil	28
4.1 Methode	28
4.1.1 Fragestellung und Hypothesen.....	28
4.1.2 Versuchsplan	29
4.1.3 Testdurchführung	31
4.2 Verhaltensbeobachtung.....	32
4.2.1 Beobachtungen während den Gruppentestungen	32
4.2.2 Beobachtungen während den Einzeltestungen.....	33
4.3 Probleme bei der Datenerhebung.....	34
5 Untersuchungsmaterial	35
5.1 Das Adaptive Intelligenz Diagnostikum 2 (Kubinger, 2009a).....	35
5.2 Entwicklung der Gruppenvorgabe.....	38
5.3 Einzelvorgabe des AID 3	43
5.4 Testinstruktion	44

6 Darstellung der Ergebnisse	45
6.1 Stichprobenbeschreibung	45
6.2 Statistische Auswertung	48
6.3 Ergebnisse	49
6.3.1 Ergebnisse der Multivariaten Varianzanalyse	49
6.3.2 Ergebnisse der Zweifachen Varianzanalyse mit Messwiederholung.....	52
7 Interpretation der Ergebnisse, Diskussion und Ausblick	72
7.1 Interpretation der Ergebnisse.....	72
7.2 Diskussion	74
7.3 Ausblick	79
8 Zusammenfassung.....	80
9 Literaturverzeichnis	82
10 Abbildungs- und Tabellenverzeichnis	87
11 Anhang	89
11.1 Elterninformationsbrief und Einverständniserklärung	89
11. 2 Gesprächsleitfaden.....	91
12 Lebenslauf	93

1 Einleitung

Intelligenztest ist die „Bezeichnung für eine Gruppe von Tests, welche auf der Grundlage unterschiedlicher Intelligenzdefinitionen und Intelligenztheorien die intellektuelle Leistungsfähigkeit quantitativ und qualitativ bestimmen“ (Häcker & Stapf, 2004, S. 406)

Krohne und Hock (2007) sehen den Kernbereich der Intelligenz in der Fähigkeit, Probleme aufzudecken und diese zu lösen. Als eine der wichtigsten Funktionen nennen die Autoren die Anpassung und die aktive Gestaltung der Umwelt in Hinblick auf persönlich hoch bewertete Ziele. Zusätzlich können noch folgende Bereiche hinzugefügt werden: „schnelles und korrektes Erkennen von Zusammenhängen, schlussfolgerndes Denken sowie den Erwerb und die Nutzung von Wissen“ (S. 370).

Die Erfassung der intellektuellen Fähigkeiten bei Kindern und Jugendlichen ist bei schulpsychologischen Fragestellungen (z.B.: Unter- bzw. Überforderung, Abklärung einer möglichen Legasthenie/Dyskalkulie, Aufmerksamkeitsprobleme etc.) unumgänglich. Mit einem Intelligenztest werden die Schwächen und Stärken eines Kindes erfasst. Wichtig bei solchen Testungen ist eine für das Kind individuell angepasste Testsituation, damit es ihr Leistungsoptimum zeigen kann. Forschungen im Bereich der Intelligenzdiagnostik sind relevant um die Beratung im schulpsychologischen Kontext zu verbessern und dem aktuellsten wissenschaftlichen Stand anzupassen. Die Intelligenztestbatterie AID 2 (Adaptives Intelligenzdiagnostikum 2.2, Kubinger, 2009a) ist ein Einzelverfahren, welches mehrere Dimensionen an kognitiven Fähigkeiten erfasst. Sie ist ein viel eingesetztes Verfahren in schulpsychologischen Fragestellungen. Dieser Intelligenztest sieht sich als Individualverfahren, welches nach Herle (2003) nur einer Testperson vorgegeben werden kann. Könnten Psychologen diesen Test einer ganzen Klasse in einer Stunde vorgeben, würde diese Gruppenvorgabe enorme Zeit einsparen. Es würde den Unterricht weniger stören und viele Kinder könnten gleichzeitig in einer Unterrichtsstunde getestet werden.

Die Vorgabe von psychologischen Verfahren weisen im Einzelsetting, als auch im Gruppensetting, verschiedene Möglichkeiten und Probleme auf. Wie schon erwähnt, kann in Gruppentestungen Zeit eingespart werden, da mehrere Personen gleichzeitig

getestet werden können. Zudem werden Gruppentestungen nachgesagt, dass sie ein ökonomisches Verfahren sind. Bühner (2011) spricht von einem ökonomischen Test, wenn er einfach zu handhaben ist, die Testzeit angemessen ausfällt und wenig Materialien verbraucht.

Auf der anderen Seite bieten Einzelsettings eine individuell gestaltete Testsituation, die auf die Bedürfnisse der einzelnen Personen abgestimmt werden kann.

Besonders der AID 2 verwendet in einzelnen Untertests bestimmte Materialvorgaben, um manuell-visuelle Fähigkeiten zu überprüfen. Diese Materialbeschaffenheit müsste in Gruppentestungen außer Acht gelassen werden und ist nur in Einzelsettings umzusetzen.

Vor allem aber bei der Testung von Kindern, kann das Problem auftauchen, dass diese in Gruppentestungen die Instruktion nicht verstehen und ihre Verständnislosigkeit nicht vor der Gruppe bekannt geben wollen (Kubinger, 2009b). Bei Einzeltestungen ist der Testleiter anwesend und kann bei den Übungsbeispielen das Verständnis abprüfen.

In der vorliegenden Arbeit wird nun untersucht, ob die Bearbeitung von Aufgaben aus einem Intelligenztest für Kinder und Jugendliche unter zwei verschiedenen Bedingungen gleich schwierig ist. Dabei werden die Untertests „Alltagswissen“, „Angewandtes Rechnen“, „Synonyme Finden“, „Funktionen Abstrahieren“ und „Soziales Erfassen und Sachliches Reflektieren“, aus dem Adaptiven Intelligenz Diagnostikum 2 (Version 2.2., Kubinger, 2009a), im Einzelsetting und Gruppensetting vorgegeben. Untersucht werden die Leistungsunterschiede von 65 Volksschülern und 89 Gymnasiasten je nach Testvorgabe.

Eingangs möchte ich auf historische Aspekte in der Intelligenzdiagnostik eingehen. Danach wird ein Überblick über einige Erhebungstechniken gegeben wobei insbesondere ein Einblick in die Vor- und Nachteile von Gruppenverfahren und Einzelverfahren gegeben wird, um einen Grundstein für die empirische Untersuchung zu schaffen. Im empirischen Teil wird dargestellt, ob Leistungsunterschiede zwischen den Bedingungen Einzel- vs. Gruppentestung, mit ausgewählten Untertests des AID 2, bestehen.

2 Definition des Begriffes Intelligenz

Seit über 100 Jahren wird Intelligenz intensiv erforscht und es gibt noch keine eindeutige Auffassung darüber, was Intelligenz eigentlich ist. Diese Kontroversen gründen darin, dass Intelligenz „keinen allgemein anerkannten, objektiven Inhalt besitzt“ (Funke & Vaterrodt-Plünnecke, 1998, S.9).

Die folgenden Definitionen von Intelligenz, sollen dies veranschaulichen:

Binet (1905) definiert Intelligenz, als die Fähigkeit „gut urteilen, gut verstehen, gut denken“ zu können (Guthke, 1996, S.31).

1956 erklärt Wechsler Intelligenz folgendermaßen: „Intelligenz ist die zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen“ (S.13).

Nach Gardner (1991) „muss eine intellektuelle Kompetenz ein Sortiment von Fähigkeiten beinhalten, die ihrem Inhaber ermöglichen, echte Probleme oder Schwierigkeiten zu lösen und die Fähigkeit, Probleme zu entdecken oder zu schaffen, um die Basis für neues Wissen zu legen“ (S. 65).

In diesen vielfältigen Versuchen Intelligenz zu definieren, finden sich nach Neubauer (2005) zwei Aspekte die alle Definitionen gemeinsam haben:

1. Sich in neuen Situationen, auf Grund von Einsichten, zurechtzufinden,
2. Aufgaben mit Hilfe des Denkens zu lösen, ohne dass hierfür die Erfahrung, sondern vielmehr die Erfassung von Beziehungen, das Wesentliche ist.

In den neuesten Forschungsarbeiten wird Intelligenz als ein komplexes Konstrukt bezeichnet, welches sich aus mehreren kognitiven Fähigkeiten zusammensetzt (Brocke & Beauducel, 2001).

Durch die intensive Erforschung der Intelligenz ergaben sich zum Beispiel folgende Teilfähigkeiten: numerische Intelligenz, verbale Intelligenz, soziale Intelligenz (Brocke & Beauducel, 2001). Aufgrund der Erweiterungsfähigkeit und der theoretischen Entwicklung werden solche Konstrukte auch offene Konstrukte genannt (Vock & Holling, 2006).

Da das Konstrukt Intelligenz nicht direkt beobachtbar ist, so dass dann etwas über dieses Merkmal gesagt werden kann, werden nicht beobachtbare Sachverhalte (Konstrukte) mit beobachtbaren Verhaltensvariablen (zum Beispiel Anzahl der richtigen Lösungen) verknüpft (Brocke & Beauducel, 2001; Holling, Preckel & Vock, 2004).

Es ist noch nicht möglich Konstrukte durch einen einzigen Test zu erfassen. Vielmehr werden durch eine Vielzahl von Verfahren die Teilfähigkeiten erschlossen (Brocke & Beauducel, 2001).

In Abhängigkeit von der diagnostischen Fragestellung und der Interpretation der Ergebnisse ist es wichtig zu wissen, auf welchem Intelligenzmodell ein Test aufgebaut ist. Die entwickelten Intelligenztests verwenden unterschiedliches Aufgabenmaterial und beruhen auf unterschiedlichen theoretischen Konzepten (Holling et al., 2004). Um die erhaltenen Werte in einem Intelligenztest interpretieren zu können, ist es unerlässlich die Methode des Verfahrens zu kennen (Stemmler et al., 2011).

2.1 Die ersten Intelligenztests und die Entwicklung des Intelligenzquotienten

2.1.1 Die Arbeiten von Binet und Simon

James McKeen Cattell (1860-1944) prägte 1890 den Begriff „mental tests“ (Tests für geistige Prozesse). Als Begründer der ersten weltweiten Testbatterie versuchte er mittels Reaktionszeitmessungen Leistungsfähigkeit zu messen (Stemmler, Hagemann, Amelang & Bartussek, 2011). Mit seinen mental tests wollte Cattell Prognosen für einen späteren Erfolg auf dem College aufstellen. Jedoch führten diese Tests zu negativen Ergebnissen und seine Forschungen konnten nicht weitergeführt werden. (Lamberti, 2006, Guthke, 1996).

Ein Neubeginn erfolgte 1904 durch Alfred Binet (1857-1911). Gemeinsam mit seinem Kollegen Theophile Simon (1873-1961) entwickelte er den ersten Intelligenztest (Binet-Simon-Test) für Kinder im Alter von 3 bis 15 Jahren. Mit diesem Verfahren sollten all jene Kinder ermittelt werden, die Entwicklungsrückstände aufzeigten (Guthke, 1996).

Die Aufgaben wurden so gewählt, dass sie pro Altersstufe von 50 bis 75% der Kinder gelöst werden konnten (Stemmler et al., 2011). Abbildung 1 veranschaulicht dies mittels einer Grafik.

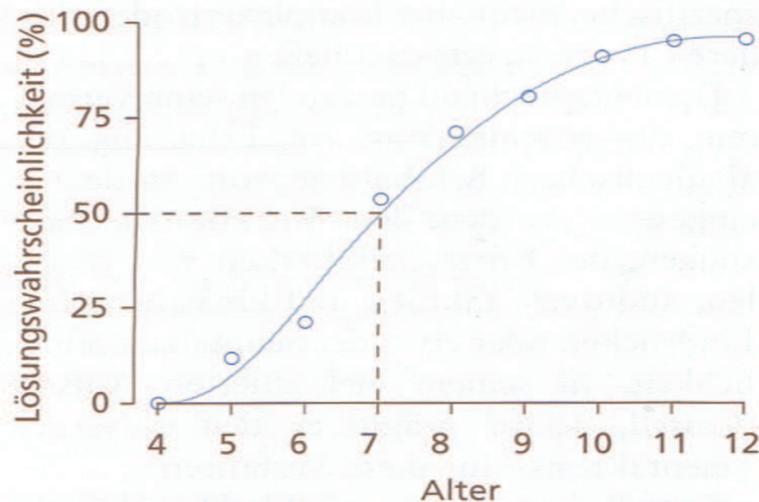


Abb. 1 Lösungswahrscheinlichkeit einer Binet-Aufgabe. 50% der Kinder im Alter von 7 Jahren lösen die Aufgabe richtig (Stemmler, Hagemann, Amelang und Bartussek, 2011, S.30)

Die entwickelten Aufgaben erforderten von den Kindern logisches Denken und nicht das Auswendiglernen eines Stoffes. Der Test diente dazu, komplexere Leistungen wie Urteilsvermögen, Verständnisfähigkeit, Aufmerksamkeit, motorische Fertigkeiten und Vorstellungskraft zu messen (Funke & Vaterrodt-Plünnecke, 1998). Abbildung 2 listet einige Aufgaben aus dem Binet Test auf.

- Altersgruppe 6:**
- (1) Kennt rechts und links, was durch Anfassen der Ohren erkennbar ist.
 - (2) Wiederholt einen Satz von 16 Silben.
 - (3) Wählt das hübschere Gesicht aus jedem von drei Paaren.
 - (4) Kennt Morgen und Nachmittag.
- Altersgruppe 8:**
- (1) Liest eine Textpassage und erinnert sich an zwei Details.
 - (2) Benennt vier Farben – rot, gelb, blau, grün.
 - (3) Zählt rückwärts von 20 auf null.
 - (4) Schreibt einen kurzen Satz nach dem Diktat unter Verwendung von Federhalter und Tinte.
 - (5) Kennt die Unterschiede zwischen zwei Gegenständen aus dem Gedächtnis.
- Altersgruppe 11:**
- (1) Findet Absurditäten in kontradiktorischen Feststellungen.
 - (2) Nennt 60 Wörter in drei Minuten.
 - (3) Definiert abstrakte Begriffe (Nächstenliebe, Gerechtigkeit, Freundlichkeit).
 - (4) Bringt zufällig angeordnete Wörter in einen sinnvollen Satz.

Abb. 2 Einige Aufgaben aus dem Binet Test (Stemmler, Hagemann, Amelang und Bartussek, 2011, S.31)

Das sogenannte Intelligenzalter ermöglichte die Bestimmung der Leistungsfähigkeit, welches wie folgt bemessen wurde: Die Testaufgaben wurden Kindern unterschiedlichen Alters vorgegeben und somit konnte ein Durchschnittsalter berechnet werden, welches normale Kinder in einem bestimmten Lebensalter erreichten (Asendorpf, 2007). Die Testleistung jedes einzelnen Kindes wurde mit diesem Durchschnittsalter verglichen. Löste nun beispielsweise ein achtjähriges Kind die Aufgaben bis zur sechsten Testaltersstufe, erhielt es ein Intelligenzalter von sechs Jahren (Zimbardo & Gerrig, 2004).

2.1.2 Terman's Gruppentest

Die ersten Intelligenztests, die entwickelt wurden, waren Individualverfahren, mit denen jeweils nur eine Testperson untersucht werden konnte. 1917 benötigte die US Army die Unterstützung von Psychologen zur Auswahl neuer Rekruten (Stemmler et al., 2011). Da die bestehenden Intelligenztests nicht ausreichten um eine große Anzahl an Erwachsenen gleichzeitig zu testen, entwickelten die beauftragten Psychologen den Alpha-und-Beta Gruppentest der US Army (Funke & Vaterrodt-Plünnecke, 1998). Der Stanford-Professor Lewis Terman (1877-1956) erweiterte den Intelligenztest von Simon und Binet und gab ihm den Namen Stanford-Binet-Intelligenztest. Unter der Hilfe von Terman und seinem abgeänderten Intelligenztest konnte ein Gruppenverfahren entwickelt werden. Zu dieser Zeit fand der Alpha-Test hohes Ansehen, da mit ihm mehrere Personengruppen auf einmal getestet werden konnten (Myers, 2004, zitiert nach Terman, 1912).

2.1.3 Wechslers Intelligenzkonzept

Jedoch ergab sich ein großer Nachteil des Intelligenzalters. Ein Entwicklungsrückstand von zwei Jahren liegt bei einem Zehnjährigen noch im Normalbereich, bei einem Vierjährigen bedeutet es eine enorm niedrige Intelligenz (Stemmler et al., 2011). Aufgrund dieser Überlegungen entwickelte 1912 der Psychologe Wilhelm Stern (1871-1938) den Intelligenzquotienten. Dabei wird das Intelligenzalter durch das Lebensalter dividiert und mit 100 multipliziert (Holling, et al., 2004).

Die Berechnung des Intelligenzquotienten nach Stern (1912) war jedoch nur bei Kindern anwendbar, und führte bei der Berechnung von Erwachsenen zu unsinnigen Ergebnissen (Guthke, 1996). Man ging davon aus, dass das Intelligenzalter über das Erwachsenenleben hinweg unverändert bleibt und berücksichtigte nicht, dass die intellektuellen Fähigkeiten ab einem gewissen Alter absinken (Gardner, 1991).

Auf Grund dieser Problematik entwickelte David Wechsler (1956) eine neue Methode zur Berechnung des Intelligenzquotienten. Bei seiner Intelligenzmessung wird die kognitive Leistung einer Person mit der Durchschnittsleistung ihrer Altersgruppe verglichen (Myers, 2004).

Wechsler (1956) nimmt an, dass sich die Intelligenz aus verschiedenen Fähigkeiten auseinandersetzt. In seinem ersten Intelligenztest für Erwachsene (Hamburger-Wechsler-Intelligenztest; 1956) unterteilt er deshalb seine Untertests in einen verbalen Teil und einen Handlungsteil. Beispiele für den Handlungsteil sind: eine Bildergeschichte richtig ordnen oder das Zusammensetzen von bekannten Figuren. Somit erlaubte dieser Test neben einer Berechnung der gesamten Intelligenz auch die eines Verbal-IQ und eines Handlungs-IQ (Holling, 2004).

Seit Wechsler spricht man daher nicht mehr von einem Intelligenzquotienten, sondern von einem Abweichungsmaß. Der Testwert einer Person wird mittels einer Tabelle mit dem Normwert der repräsentativen Vergleichsstichprobe verglichen (Kubinger, 2009b).

2.2 Intelligenzmodelle

Seit mehr als 100 Jahren werden intellektuelle Fähigkeiten analysiert, statistische Methoden angefertigt um die Leistungen zu erheben und Theorien über das Konstrukt Intelligenz entwickelt. Rost (2009, S.22ff) teilt die zahlreichen Intelligenztheorien in drei Gruppen ein:

- A priori Modelle: Ihre Grundlagen sind spekulative Überlegungen, die in ihren Erklärungen Alltagserfahrungen und Einzelfälle berücksichtigen. Die Vertreter dieser Intelligenzmodelle versuchen nicht die Intelligenzkomponenten durch empirische Befunde, wie zum Beispiel Testverfahren zu widerlegen.
- A posteriori Modelle: Theorien werden durch empirische Erkenntnisse verbessert und umgestaltet bzw. auch verworfen. Vertreter dieser Intelligenzmodelle sind Spearman, Thurstone, R. B. Cattell und Carroll.
- Die letzte Gruppe bilden Ansätze, welche zwischen den beiden oben genannten Modellen liegen. Diese Theorien gehen von einer Klassifikation der intellektuellen Fähigkeiten aus und entwickeln Testverfahren, welche die einzelnen Faktoren ihres Modells erfassen.

2.2.1 Zwei-Faktoren-Theorie nach Spearman

Zu Beginn seiner Forschungen, bemerkte Spearman (1863-1945), dass Schulleistungen, die fast nichts miteinander gemeinsam hatten, hoch miteinander korrelierten. Wer gute Noten in Deutsch hatte, zeigte auch gute Leistungen in Mathematik (Rost, 2009).

Spearman (1927) nahm an, dass alle verschiedenen kognitiven Leistungen auf eine Dimension zurückzuführen sind. Diese bezeichnete Spearman (1927) als die allgemeine Intelligenz und gab ihr den Namen „general factor“ (abgekürzt: g).

Im Laufe seiner Forschungen, machte Spearman darauf aufmerksam, dass es noch eine zweite Komponente neben dem g-Faktor gibt, nämlich den spezifischen Anteil s (Horn, 1993). Wendet man diese Theorie an Aufgaben eines Tests an, werden alle intellektuellen Leistungen durch den Faktor g bestimmt (Abb. 3). Zusätzlich werden die einzelnen Aufgaben durch die spezifische Komponente bestimmt (Hofstätter, 1957).

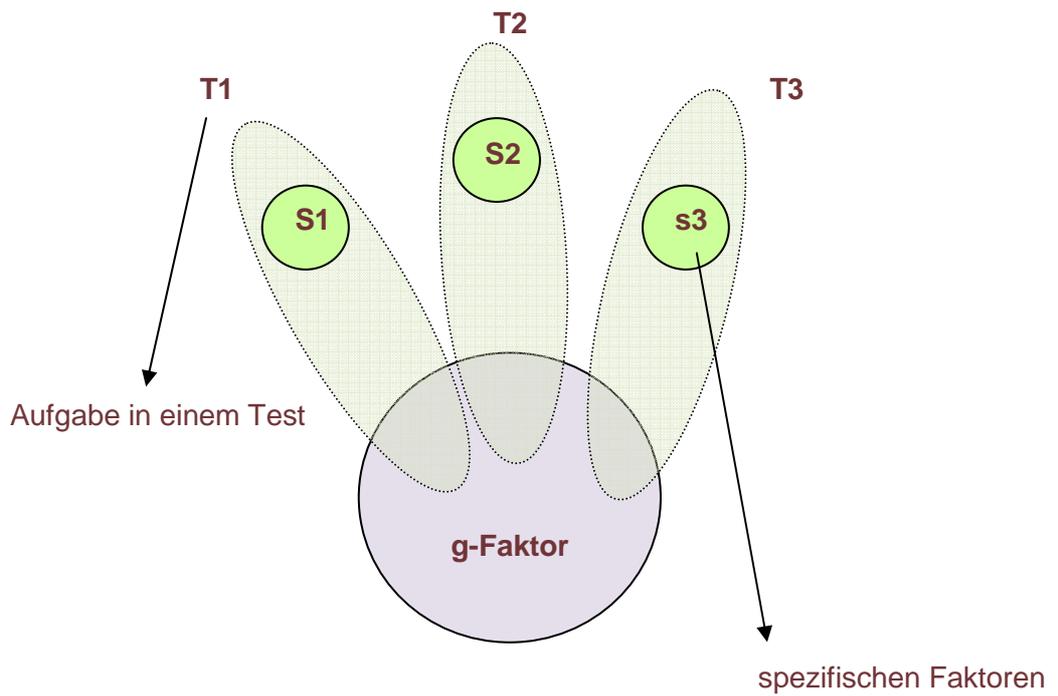


Abb. 3 Zwei-Faktoren-Theorie nach Spearman (nach Hofstätter, 1957, modifiziert)

Tests, wie die Progressiven Matrizen nach Raven (1941) und der Wiener Matrizen-Test (WMT; Forman, 1979), enthalten Aufgaben, die zur Messung von g geeignet sind (Rost, 2009).

2.2.2 Primary Mental Abilities nach Thurstone

Thurstone (1887-1955) legte Studenten 56 verschiedene Leistungstests vor, aus denen er sieben Faktoren (Primary Mental Abilities) ableitete (Myers, 2004). Diese sieben Primärfaktoren sind voneinander unabhängig (Horn, 1993), bauen nicht aufeinander auf und stehen somit auf einer Ebene (Stemmler et al., 2011).

Die Bezeichnungen für die Primary Mental Abilities und ihre Messung werden nachfolgend kurz vorgestellt (Rost, 2009, S.31ff). Im Anschluss stellt Abbildung 4 die Theorie nach Thurstone grafisch dar.

- **Memory** (Merkfähigkeit): bezeichnet ein kurzzeitiges Beibehalten von Zahlen (beim Telefonieren) oder Buchstaben-Zahlen-Kombinationen (Merken von Nummernschildern), wie es im Alltag notwendig ist. Erfasst wird der Faktor „Memory“ durch das Einprägen verschiedener Assoziationen wie zum Bsp. „Stuhl-24“ oder „Kasten-44“.
- **Number** (Rechengewandtheit): Diese Fähigkeit tritt bei allen einfachen Rechenoperationen wie Addition, Subtraktion, Multiplikation und Division auf und ist an numerisches Material gebunden. Gemessen wird diese Eigenschaft zum Beispiel anhand folgender Anweisung: „Markiere jede Zahl in der nachfolgenden Reihe, die exakt um drei größer ist als die vorausgehende Zahl“. Dieser Faktor hat nichts mit der Fähigkeit gemeinsam, schwierige mathematische Probleme zu lösen.
- **Perception** (Wahrnehmungsgeschwindigkeit): ist durch die Fähigkeit charakterisiert rasch zu erkennen und visuell zu diskriminieren. Die Geschwindigkeitskomponente spielt hier eine wichtige Rolle. Eine mögliche Aufgabe zur Messung dieser Eigenschaft könnte sein: Aus verschiedensten Bildern, die einem Bezugsbild ähnlich sind, muss jenes herausgesucht werden, dass dem Bezugsbild identisch ist.
- **Reasoning** (logisches Denken): umfasst komplexe kognitive Probleme, die materialunabhängig sind. Aufgaben zum logischen Denken erfordern Gesetzmäßigkeiten herauszufiltern, mit denen Buchstaben- oder Zahlenreihen fortgesetzt werden sollen.
- **Space** (Raumvorstellung): ist die Fähigkeit, sich Objekte im zwei- und dreidimensionalen Raum vorzustellen. Der Testperson wird eine Bezugsfigur

vorgelegt, daneben liegen identische und ähnliche Figuren, welche aber in unterschiedlichen Dimensionen präsentiert werden. Es müssen jene Figuren herausgesucht werden, welche durch Rotationen an die Bezugsfigur angepasst werden können.

- **Verbal** (Wortverständnis): bezeichnet die Fähigkeit zu verbalen Beziehungen und Bedeutungen. Die Aufgaben sind durch Analogiebildungen gekennzeichnet. Zu einem Synonym muss ein anderes Wort gefunden werden, welches die gleiche Bedeutung hat.
- **Word** (Wortflüssigkeit): im Gegensatz zu „Verbal“ steht hier nicht die Bedeutung der Wörter im Vordergrund, sondern die Mengenleistung. Die Aufgabenstellungen bei dieser Eigenschaft, verlangen von der Testperson zum Beispiel, möglichst viele Wörter zu nennen die mit „F“ beginnen.

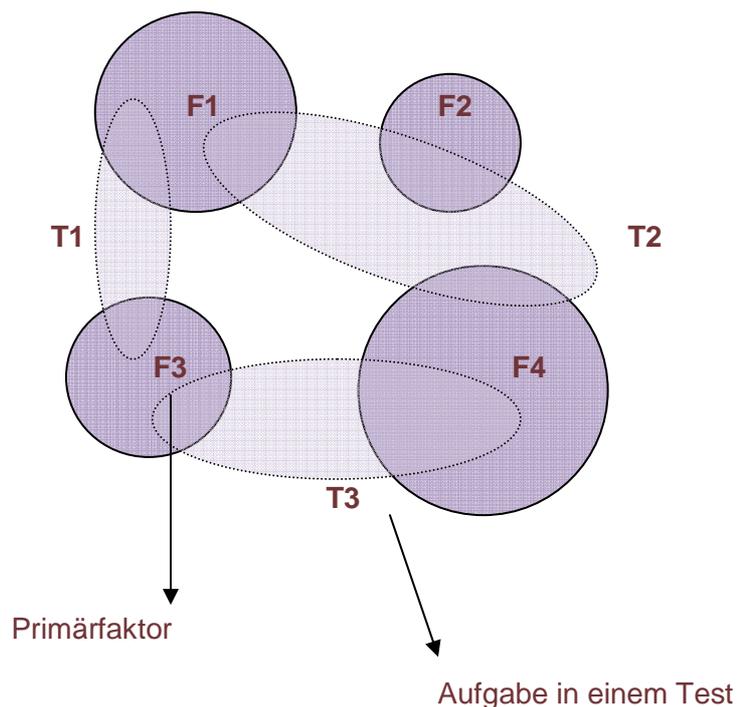


Abb. 4 Multiple-Faktoren-Theorie nach Thurstone (Hofstätter, 1957, modifiziert)

Der Intelligenz-Struktur-Test 2000 R (IST 2000 R; Amelang, Brocke, Liepmann & Beauducel, 2001), das Prüfsystem für Schul- und Bildungsberatung (PSB-R 4-6; Horn, Lukesch, Kormann & Mayrhofer, 2002), der Wilde-Intelligenz-Test (WIT-2; Kersting, Althoff & Jäger, 2008) und der Kognitive Fähigkeitstest (KFT 4-12+R; Heller & Perleth, 2000) orientieren sich an den Primärfaktoren nach Thurstone (Rost, 2009; Kubinger 2009b).

2.2.3 Fluide und Kristalline Intelligenz nach Cattell

Das Adaptive Intelligenzdiagnostikum 2 (Kubinger, 2009b) orientiert sich nicht an einem bestimmten Intelligenzmodell. Jedoch gab es Verknüpfungen zu verschiedenen Intelligenztheorien. Deshalb soll in diesem Kapitel ein kurzer Abriss von Cattells Intelligenztheorie (1963), unter der Mitarbeit von J.L. Horn (Cattell-Horn-Theorie), gegeben werden.

Die Unterteilung der Intelligenz in erworbenes Wissen und der eigentlichen Denkfähigkeit geht auf R. B. Cattell (1905-1998) zurück. Seine Arbeiten bauen auf den von Spearman (1927) definierten g-Faktor auf.

R.B. Cattell (1963) stützte sich auf die Annahme, dass sich der g-Faktor aus zwei Facetten zusammensetzte: aus einer „Fluiden“ („Flüssigen“, fluid, *gf*) Fähigkeit und aus einer „Kristallinen“ („Kristallisierten“, crystallized, *gc*) Fähigkeit. Seiner Meinung nach bestand die Intelligenz von Erwachsenen aus diesen beiden Teilen.

Die fluide Intelligenz umfasst Leistungen des Denkens, der Wahrnehmung, des Gedächtnisses und der Motorik (Oerter & Montada, 2008). Mit der fluiden Fähigkeit kann man sich an neue Situationen anpassen und neuartige Probleme lösen ohne, dass dazu Lernerfahrungen notwendig sind (Stemmler et al., 2011). Man geht davon aus, dass sie von kulturellen und gesellschaftlichen Einflüssen unabhängig ist (Holling et al., 2004). Die fluide Intelligenz wird mittels figuralen Aufgaben und räumlichen Anforderungen gemessen (Zimbardo & Gerrig, 2004). Der fluide Faktor wächst ab der Geburt, bis zu einem Alter von ca. 15-20 Jahren, an und nimmt dann, individuell stark variierend, langsam wieder ab (Rost, 2009).

Die kristalline Intelligenz wird durch Wissen und Sprache repräsentiert und bleibt über das Alter hinweg bestehen (Oerter & Montada, 2008). Sie wird durch Wissen erworben

und umfasst die Fähigkeit, dieses Wissen abzurufen (Zimbardo & Gerrig, 2004). Die kristalline Fähigkeit wächst bis zum 28. Lebensjahr und nimmt erst viel später ab. Der Abbau vollzieht sich langsamer als bei der fluiden Intelligenz. Durch das ständige Aneignen von Wissen ist sie stabiler als die fluide Intelligenz und kann sogar noch anwachsen (Rost, 2009).

Folgende Tests basieren auf der Theorie von Cattell: der Grundintelligenztest Skala 1 (CFT 1; Weiß & Osterland, 1997), der Intelligenz-Struktur-Test 2000 R (IST 2000 R; Amelang, Brocke, Liepmann & Beauducel, 2001) (Kubinger, 2009b).

2.2.4 Drei-Schichten-Modell nach Carroll

Carroll entwickelte ein Modell, das derzeit große Akzeptanz erreicht hat. Es besteht aus 3 Schichten, welche hierarchisch angeordnet sind (Schweizer, 2005).

Auf der obersten Schicht (Stratum III) befindet sich die allgemeine Intelligenz. Sie ist stabil, charakteristisch für komplexere kognitive Prozesse und umfasst Aufgaben zum schlussfolgernden Denken (Rost, 2009).

In der mittleren Ebene (Stratum II) sind acht Faktoren angeordnet, die Komponenten der allgemeinen Intelligenz sind (Abb. 5): Fluide Intelligenz und Kristalline Intelligenz aus dem Modell von Cattell (1963), Allgemeines Gedächtnis und Lernen, Allgemeine visuelle Wahrnehmung, Allgemeine auditive Wahrnehmung, Allgemeine Zugriffsfähigkeit, Allgemeine kognitive Geschwindigkeit und Verarbeitungsgeschwindigkeit (Schweizer, 2005, S.6).

Auf der untersten Ebene (Stratum I) sind 68 Teilkomponenten angesiedelt (Süß, 2003).

Die allgemeine Intelligenz steht zu den acht Faktoren auf unterschiedliche Weise in Beziehung. Zu einigen Faktoren, wie zum Beispiel der fluiden und kristallinen Intelligenz, besteht eine enge Beziehung, zu anderen Faktoren eine weite Beziehung (Schweizer, 2005).

Stemmler et al. (2011, S.161ff) geben eine grobe Übersicht über die acht Faktoren:

- **Fluide Intelligenz:** entspricht dem Faktor *gf* nach der Theorie von Cattell (1963).
- **Kristallisierte Intelligenz:** entspricht dem Faktor *gc* nach Cattell (1963).
- **Gedächtnis und Lernen:** umfasst das Nachsprechen einer Zahlenreihe, nichtzusammenhängende Wortpaare müssen gelernt werden, aus einer gelernten Wortliste so viele wie möglich reproduzieren, Lernen zusammenhängender Wortpaare, Details eines Filmes müssen erinnert werden usw.
- **Visuelle Wahrnehmung:** bezeichnet die Fähigkeit Längen einzuschätzen, bildliches Vorstellungsvermögen und das Erkennen von räumlichen Beziehungen.
- **Auditive Wahrnehmung:** Unterschiede in der Lautstärke, Höhe und Dauer von Tönen müssen identifiziert werden
- **Allgemeine Zugriffsfähigkeit:** Längst gelerntes Material muss aus dem Gedächtnis verfügbar sein, wie zum Beispiel: Wörter nennen, die mit „H“ beginnen; Farbnamen produzieren oder runde Objekte aufzählen.
- **Kognitive Schnelligkeit:** beinhaltet die Bearbeitungszeit von Leistungstests.
- **Verarbeitungsgeschwindigkeit:** charakterisiert die Geschwindigkeit einer Reaktion nach Reizdarbietung.

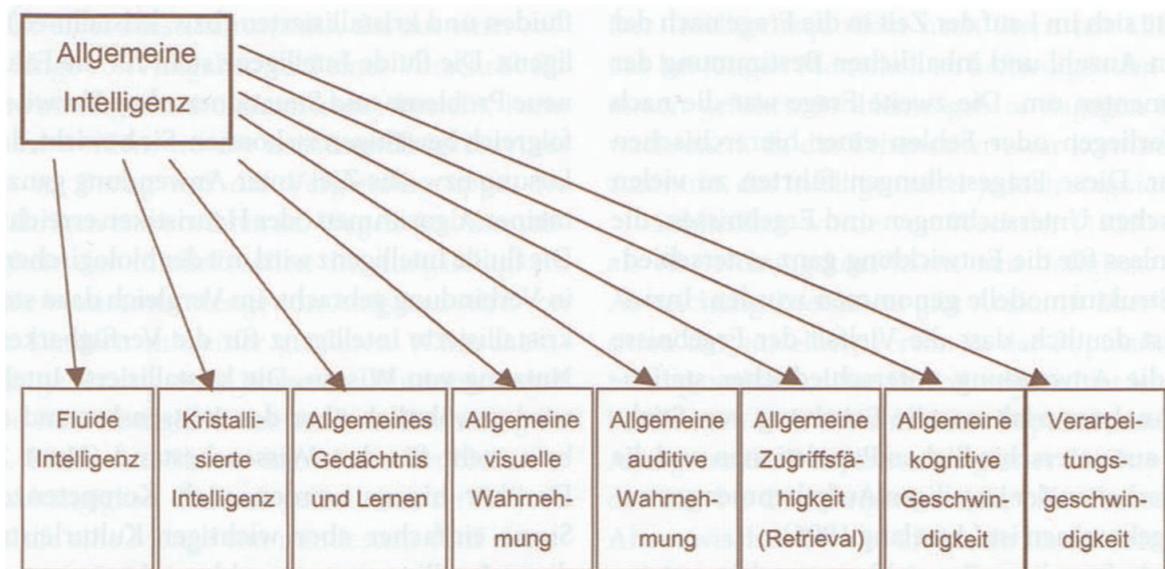


Abb. 5 Schematische Darstellung der oberen und mittleren Ebenen von Carrolls Drei-Schichten-Modell (Schweizer, 2006, zitiert nach Carroll, 1993)

2.2.5 Cattell-Horn-Carroll-Theorie (CHC-Modell)

Neueste Ansätze versuchen die Cattell-Horn Theorie, der fluiden und kristallinen Intelligenz, mit der Drei-Schichten Theorie von Carroll zur Cattell-Horn-Carroll-Theorie zu verbinden (Rost, 2009).

Das CHC Modell hat eine hierarchische Anordnung der „cognitive abilities“ die einem Generalfaktor untergeordnet sind: auf der untersten Ebene (Stratum I) befinden sich die „narrow abilities“, auf der mittleren Ebene (Stratum II) sind die „broad abilities“ angeordnet und auf der höchsten Ebene (Stratum III) befindet sich der Generalfaktor *g* (Floyd, Keith, Taub & McGrew, 2007).

Die zehn „broad abilities“ umfassen: Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-term Memory (*Gsm*), Visual Processing (*Gv*), Auditory Processing (*Ga*), Long-term Retrieval (*Gl*), Processing Speed (*Gs*), Reading and Writing (*Grw*), Quantitative Knowledge (*Gq*) and Decision/Reaction Time or Speed (*Gt*) (Taub, Floyd, Keith & McGrew, 2008).

Die „broad abilities“ wurden im CHC-Modell auf zehn Faktoren erweitert. Die „broad abilities“ umfassen 73 „narrow abilities“ (Phelps, McGrew, Knopik and Ford, 2005).

Aktuell gibt es zwei Methoden, die die Cattell-Horn-Theorie ausführen: der Woodcock-Johnson III-Test (Woodcock, McGrew & Mather, 2001) und der WISC III (Wechsler Intelligence Scale (Wechsler, 1991) (Phelps et al., 2005).

3 Darbietungsmöglichkeiten von diagnostischen Verfahren

Ein diagnostischer Test ist „ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (Lienert & Raatz, 1994, S.1).

Lienert und Raatz (1994, S.13ff) machen auf die bestimmten Gütekriterien aufmerksam, welche ein guter Test erfüllen soll. Sie unterscheiden zwischen den Hauptgütekriterien

- Objektivität meint den „Grad in dem die Ergebnisse eines Testes unabhängig vom Untersucher sind“ (S.13).
- Reliabilität ist der Grad der Genauigkeit, mit dem ein Test ein bestimmtes Merkmal misst.
- Validität meint den Grad der Genauigkeit, mit dem ein Test das Merkmal, welches er messen soll oder vorgibt zu messen, auch tatsächlich misst.

und den vier Nebengütekriterien

- Normierung: Dieses Nebengütekriterium verlangt ein Bezugssystem in welchem, individuelle Ergebnisse des Testes eingeordnet werden können.
- Vergleichbarkeit liegt dann vor, wenn eine Parallelförmigkeit eines Testes mit dem Test selbst möglich ist.
- Ökonomie: Unter Ökonomie versteht man eine kurze Durchführungszeit, wenig Materialverbrauch, eine gute Handhabung, eine schnelle und bequeme Auswertung und wenn er als Gruppentest durchführbar ist.
- Nützlichkeit: Ein Test ist dann nützlich, wenn für das zu messende Merkmal eine praktische Bedeutung besteht.

In der psychologischen Diagnostik gibt es verschiedene Methoden, wie diagnostische Verfahren gestaltet und vorgegeben werden können. In Abhängigkeit von dem zu messenden Konstrukt kommen die unterschiedlichen Darbietungsformen zum Einsatz (Kubinger, 2009b).

Bei Fähigkeits- und Leistungstests, zu denen die Intelligenztests gehören, wird das Verhalten der Testpersonen an den Gütemaßstäben der Richtigkeit, Qualität und Schnelligkeit gemessen. Somit sind in diesem Bereich der diagnostischen Verfahren Aufgaben notwendig die einen korrekten Lösungsvorschlag besitzen mit denen die Leistungen als richtig oder falsch gewertet werden können (Krohne & Hock, 2007).

Kubinger (2009b) unterscheidet vier verschiedene Darbietungsformen psychologisch-diagnostischer Verfahren:

- Freies Antwortformat vs. Multiple-Choice-Format,
- Power- vs. Speed-(and-Power)test,
- Gruppen- vs. Individualverfahren und
- Papier-Bleistift-Verfahren vs. Computerverfahren

Freies Antwortformat vs. Multiple-Choice-Format

Beim Multiple-Choice-Format werden der Testperson mehrere Antwortmöglichkeiten dargeboten, aus welchen sie dann die richtige Lösung herausfinden muss (Bühner, 2011). Im Gegensatz zum freien Antwortformat, bei welchem die Testpersonen die Antwort selbst formulieren müssen (Jonkisz und Moosbrugger, 2007). Multiple-Choice-Formate haben eine hohe Verrechnungssicherheit. Darunter versteht man, dass jeder Testleiter beim Auswerten des Tests zu den gleichen Ergebnissen gelangt. Zusätzlich bietet dieses Format eine kurze Bearbeitung der Aufgaben und eignet sich besonders für Gruppentestungen (Kubinger, 2009b).

Multiple-Choice-Formate versprechen eine gute Verrechnungssicherheit, da die Auswertung leicht zu handhaben ist und auch mittels Computer möglich ist. Sie eignen sich besonders als Gruppenvorgabe, wodurch ihnen hohe Testökonomie nachgesagt wird (Seiwald, 2003).

Bei Tests mit freiem Antwortformat besteht die Gefahr, einer eingeschränkten Verrechnungssicherheit (Bühner, 2011). Unterschiedliche Testleiter könnten, aufgrund der unterschiedlich ausformulierten Antworten, zu denen es aufgrund des großen Freiraumes an Antwortmöglichkeiten kommt, zu verschiedenen Ergebnissen gelangen (Kubinger, 2009b, Seiwald 2003).

Jonkisz und Moosbrugger (2007) merken bei diesem Aufgabentyp den großen Zeitaufwand, für den Auswerter aber auch für den Probanden, als sehr kritisch an. Andererseits erhöht sich durch das freie Antwortformat der Informationsgehalt einzelner Items (Seiwald, 2003).

Problematisch beim Multiple-Choice-Format ist der Rateeffekt. Testpersonen wird hier die Möglichkeit gegeben, die richtige Antwort zu erraten. Bei fünf Antwortmöglichkeiten besteht eine Wahrscheinlichkeit von 20%, dass die Aufgabe richtig gelöst wird, wodurch die Reliabilität eines Testes verringert wird (Kubinger, 2009b). Jonkisz und Moosbrugger (2007) führen hier die Auswahl mehrerer Richtigantworten als Verringerung der Ratewahrscheinlichkeit an.

Bühner (2011) betont, dass diesem Aufgabentyp eher ein Wiedererkennen des Materials im Vordergrund steht, als eine Reproduktion.

Power- vs. Speed-(and-Power)test

Power-Tests sind durch die aufsteigenden Schwierigkeiten der Aufgaben und durch die unbegrenzte Zeitvorgabe gekennzeichnet (Amelang & Schmidt-Atzert, 2006). Bei dieser Art von Test kann die Person, trotz unbegrenzter Zeitvorgabe, nicht alle Aufgaben lösen (Krohne & Hock, 2007). Bei Speed-Tests steht die Aufgabenmenge im Vordergrund, die die Testperson in einer vorgegebenen Zeit löst. Die Aufgaben sind hier so gestaltet, dass die Probanden meistens ausreichend Zeit haben die Aufgaben zu lösen (Pospeschill & Spinath, 2009).

Eine Kombination aus beiden Vorgaben sind die Speed-and-Power-Tests, welche durch den aufsteigenden Schwierigkeitsgrad und die Zeitbegrenzungen charakterisiert sind (Kelava & Moosbrugger, 2007). Hier spielen die Richtigkeit und die Geschwindigkeit eine bedeutende Rolle (Krohne & Hock, 2007). Gruppentests sind meistens als Speed-and-Power-Tests konzipiert, während bei Individualverfahren, die Möglichkeit besteht, zusätzlich jede Aufgabe mit einer zeitlichen Begrenzung zu versehen. (Kubinger 2009b).

Gerade bei Gruppentestungen kann es vorkommen, dass einige Personen die Fragen schneller beantworten, wodurch es bei diesen Personen zu Wartezeiten kommen kann. Ursachen für schlechte Leistungen in einem Speed-and-Powertest können dem

Mangel an Schnelligkeit oder dem Mangel an der Fähigkeit zugeschrieben werden. (Kubinger, 2009b).

Das Erbringen von schlechten Leistungen in Power-Tests kann nicht auf den Zeitdruck zurückgeführt werden, sondern liegt hier eindeutig an der mangelnden Fähigkeit (Wagener, 2003). Power-Tests orientieren sich an der Fähigkeit und nicht dem Schnelligkeitsfaktor (Lienert & Raatz, 1994).

Gruppen- vs. Individualverfahren

Eine Gruppentestung erlaubt dem Testleiter, mehreren Testpersonen gleichzeitig ein Verfahren vorzugeben. Eine Individualtestung lässt die Untersuchung einer einzelnen Person zu. Herle (2003) weist explizit darauf hin, dass Einzelverfahren „nur im Rahmen einer Individualtestung zum Einsatz kommen“ (S.205) und Gruppenverfahren auch als Individualtestung vorgeben werden können.

Desweiteren sind Einzelverfahren eher aufwendig, da pro Testung ein Testleiter zur Stelle sein muss. Auf der anderen Seite bieten sie eine ausführliche Verhaltensbeobachtung, was den Gruppentestungen nicht nachgesagt werden kann (Amelang & Schmidt-Atzert, 2006).

Kubinger (2009b) führt weiter an, dass Gruppentests oft mit einer Speedkomponente versehen sind und somit keine ungestörte Testsituation gewährleistet werden kann. Zudem sind Individualverfahren fehleranfälliger, wenn es um die Auswertung geht.

Die Vorteile und Nachteile von Gruppen- und Individualverfahren werden im Kapitel 3.1. noch ausführlicher beschrieben.

Papier-Bleistift-Verfahren vs. Computerverfahren

Papier-Bleistift-Verfahren sind durch die schriftliche Bearbeitung von Aufgaben mittels Bleistift auf Papier charakterisiert (Kubinger, 2009). Sie sind die am weit verbreitetsten Verfahren (Jonkisz & Moosbrugger, 2007). Seit den 80er Jahren werden fast alle Papier-Bleistift-Verfahren mittels einer computergestützten Version angeboten (Kubinger, 2009b).

Jonkisz und Moosbrugger (2007) erläutern, dass Computertests ökonomischer sind, da die Auswertung einfach zu handhaben ist. Amelang und Schmidt-Atzerz (2006) fügen dem noch hinzu, dass dem Testleiter eine korrekte Auswertung, ohne zusätzliche Arbeitszeit, sofort nach der Testung bereitgestellt wird.

Ferner verfügen Computerverfahren über einen höheren Grad an Objektivität als Papier-Bleistift-Verfahren. Die Testpersonen müssen die Instruktionen lesen und anhand von Übungsbeispielen wird überprüft, ob die Instruktion korrekt verstanden wurde (Wagner-Menghin, 2003).

Personen mit mangelndem Leseverständnis könnten hier benachteiligt sein. Nehmen Testpersonen an Computertests teil bei denen es um Schnelligkeit geht, könnten jene, die im taktil-kinesthischen¹ Bereich Probleme haben, gehandicapt sein (Wagner-Menghin, 2003; Kubinger, 2009b).

Die Nutzung des Computers ebnet den Weg zur Erfassung neuer Eigenschaften, wie den organisatorischen Fähigkeiten, mithilfe von Computer-Simulationen. Mittels realitätsnahen komplexen Aufgabenstellungen soll das Verhalten von Personen erfasst werden (Kubinger, 2009a). Solche Verfahren ermöglichen, unter Darbietung bewegender Reize, die Gelegenheit Reaktionsschnelligkeit unter komplexeren Bedingungen zu messen (Amelang & Schmidt-Atzert, 2006).

Trotz bestimmter Kritikpunkte der einzelnen Vorgabemöglichkeiten, sind oft die Fragestellungen bzw. Ziele und die vorhandenen Ressourcen entscheidend welche Methode einzusetzen ist.

¹ kinesthatisch: Bewegungswahrnehmung des eigenen Körpers (Häcker & Stapf, 2004)

3.1 Vergleiche von Gruppentestung und Einzeltestung

3.1.1 Vorteile der Gruppentestung

Alle einsetzbaren Verfahren in der Diagnostik bieten Vorteile und Nachteile. Die Daten in dieser Arbeit wurden mit den beiden Vorgabemodi Gruppen- und Individualtestung erhoben. Deshalb wird in diesem Kapitel genauer auf die Vorteile und Nachteile dieser Verfahren eingegangen.

Ein bedeutender Unterschied zwischen den beiden Vorgabemodi liegt in der Ökonomie.

Ein Test erfüllt das Gütekriterium der Ökonomie, wenn die finanziellen und zeitlichen Ressourcen minimal gehalten werden (Kubinger & Jäger, 2003).

Verschiedentlich wurde betont (Lienert und Raatz, 1994; Woike 2003; Bühner, 2011), dass ein Verfahren ökonomisch ist bzw. es zu einer Verbesserung der Testökonomie kommt, wenn es als Gruppentest durchführbar ist.

Bühner (2011) spricht von einem ökonomischen Test, wenn er einfach zu handhaben ist, die Testzeit angemessen ausfällt und wenig Materialien verbraucht. Woike (2003) betont neben finanziellen Kosten noch die zeitlichen Ressourcen, die bei der Ökonomie zu berücksichtigen sind. Auf der einen Seite muss der Testleiter Zeit für die Testung selbst, für die Auswertung und Rückmeldung aufbringen. Auf der anderen Seite ist auch die Testzeit für die Testperson entscheidend.

Hornke (2006) weist explizit darauf hin, dass die Ökonomie den Hauptgütekriterien Validität, Reliabilität und Objektivität unterzuordnen ist. Deshalb darf die Testlänge nicht zu kurz sein, da das Verfahren sonst an Reliabilität und Validität verliert (Woike, 2003). Jedoch sollte eine Untersuchung die Testperson betreffend zeitlicher, psychischer und physischer Betrachtungsweise nicht belasten (vgl. Kubinger, 2009b, S.116). Wie Bühner (2011) richtig bemerkt, muss die Testzeit „im Verhältnis zur Wichtigkeit der diagnostischen Fragestellung stehen“ (S.72).

Amelang und Schmidt-Atzert (2006) heben einen weiteren Vorteil der Gruppenverfahren hervor. Durch die Testung in Gruppen können mehrere Personen gleichzeitig getestet werden, was zu einem geringeren Zeitaufwand führt und zu einer besseren Aktualisierung von Testnormen.

Wie soeben beschrieben wurde, kann durch das gleichzeitige Testen mehrerer Personen Zeit eingespart werden. In einer Gruppentestung ist es unmöglich jeder Person die Zeit zur Verfügung zu stellen, die es zum Beantworten der Aufgaben benötigt. Somit ist es günstig ein Zeitlimit, für das Lösen der Aufgaben, aufzustellen (Kubinger, 2009b), was zu einer Vermengung der bereits erwähnten Power-/und Speedkomponenten führt.

Gruppenverfahren sind nicht nur ökonomisch, sie sind auch objektiv.

„Unter Objektivität eines Tests ist zu verstehen, dass die mit ihm gewonnenen Ergebnisse unabhängig vom Untersucher sind“ (Kubinger, 2009b, S.38).

Innerhalb der Objektivität unterscheidet man zwischen der Testleiterunabhängigkeit, der Verrechnungssicherheit und der Interpretationseindeutigkeit. Testleiterunabhängigkeit ist in einem Verfahren gegeben, wenn das Verhalten der Testperson während der Testung und ihr Testwert unabhängig sind von zufälligen Verhaltensaspekten des Testleiters (Kubinger, 2009b).

Desweiteren muss der Testleiter die Regeln einhalten, die für die Durchführung eines Verfahrens festgelegt sind (Holling et al., 2004). Gruppentestungen ziehen hier ihren Vorteil, da es seitens des Testleiters unmöglich ist, sich einigen Personen gegenüber anders zu verhalten und die Kommunikation gering ausfällt, da die Instruktionen zumeist schriftlich verlaufen (Herle, 2003; Kubinger, 2009b).

Da Gruppentestungen nach dem Multiple-Choice-Format gestaltet werden, kann diesem Verfahren eine Verrechnungssicherheit nachgesagt werden (Kubinger, 2009b). Man spricht von Verrechnungssicherheit, wenn jeder Testleiter die gleichen Testwerte eines Probanden ermittelt (Bühner, 2011). Im Adaptiven Intelligenz Diagnostikum wird der Testperson ein freies Antwortformat geboten. Dadurch könnte es seitens des Testleiters zu Unsicherheiten bei der Vergabe der Testwerte kommen und somit zu einer geringeren Verrechnungssicherheit (Kubinger, 2009b).

Einige Testpersonen begünstigen Testungen in der Gruppe, da diesbezüglich ihre Leistungen anonym bleiben weil sie nur „einer von vielen“ sind (Herle, 2003). Jonkisz & Moosbrugger (2007) beschreiben einen schwächeren Effekt der sozialen Erwünschtheit im Zusammenhang mit der Anonymität. Soziale Erwünschtheit stellt ein Antwortverhalten dar, in welchem Personen absichtlich falsche oder richtige Antworten geben, um „sich in einem möglichst günstigem Licht darzustellen“ (Jonkisz & Moosbrugger, 2007, S.59).

Laien, die sich zum ersten Mal dem Prozess des psychologischen Diagnostizierens unterziehen, stehen dem eher kritisch gegenüber. Sie fühlen sich im Gespräch mit dem Psychologen wie ein offenes Buch aus dem dieser die intimsten Geheimnisse herauslesen könne (Kubinger, 1997).

3.1.2.Nachteile der Gruppentestung

Nachteile in Gruppenverfahren finden sich, in der Gruppensituation selbst, in der eventuell Testpersonen dazu neigen, bei ihrem Nachbar abzuschreiben oder sich durch die anderen anwesenden Personen gestört fühlen (Amelang & Schmidt-Atzert, 2006).

Gruppentestungen erzwingen eine Vermischung der beiden Verfahrensmöglichkeiten Powertests und Speedtests zu der Mischform Speed-and-Powertests. Diese besondere Gestaltungsmöglichkeit weist gewisse Mängel auf. Schlechte Leistungen einer Testperson könnten entweder durch den Mangel an Schnelligkeit oder aber durch den Mangel an Fähigkeit erklärt werden (Kubinger, 2009b).

Zusätzlich wird in Gruppensituationen eine genaue Verhaltensbeobachtung erschwert (Jonkisz & Moosbrugger, 2007).

Amelang und Schmidt-Atzert (2006) führen hier noch die gute Lesefertigkeit an, die in Gruppentestungen wichtig ist, um die schriftlichen Instruktionen zu verstehen. Personen mit geringer Lesefertigkeit werden in diesem Verfahren benachteiligt.

Darüber hinaus ist es gerade bei der Testung von Kindern bei Gruppentestungen oftmals nicht möglich, die – wie im AID 2 Manual geforderten – optimalen Rahmenbedingungen herzustellen, unter denen Kinder ihr Leistungsoptimum zeigen können.

Unter optimalen Rahmenbedingungen versteht es sich, laut Manual (Kubinger, 2009a), eine angenehme Atmosphäre zu schaffen, die es dem Kind erlaubt, sich bestmöglich zu entfalten. Das Kind soll zu angemessenen Leistungsstreben motiviert werden. Der Testleiter sollte eine vertrauensvolle Beziehung herstellen.

Gleichermaßen ist die Erfassung manuell-visueller Fähigkeiten in Gruppentests gehandicapt. Solche Fertigkeiten, die sich im Kinder- und Jugendbereich als wichtig herausstellen, können aufgrund der Materialbeschaffenheit nicht untersucht werden (Kubinger, 2009a).

3.1.3 Vorteile der Individualtestung

Die Individualtestung erlaubt dem Testleiter einen besseren Überblick über die Testsituation. Während der Testung können Beobachtungen über das Verhalten, der zu testenden Person, getätigt werden (Jonkisz & Moosbrugger, 2007). In den Beiträgen von Amelang & Schmidt-Atzer (2006), Herle (2003) und Eichhorn (2011) wird deutlich, dass eine genauere Verhaltensbeobachtung möglich ist und bestimmte Aspekte, wie Motivation oder Frustration, beobachtet werden können.

Zudem ist es interessant einen Einblick in die Strategien zu bekommen, welche die Person anwendet um die Aufgaben zu lösen. Fühlt sich die Person, zum Beispiel bei schwierigeren Aufgaben frustriert oder unmotiviert, ist ein Agieren des Testleiters möglich (Herle, 2003).

Da es in der Individualtestung zu einer psychologischen Untersuchung einer Einzelperson kommt, wird die Testperson nicht dazu verleitet, bei anderen Personen abzuschreiben. Ebenso kann es hier nicht zu Störungsfaktoren durch andere Testpersonen kommen. (Amelang & Schmidt-Atzert, 2006).

Durch die Konzentration des Testleiters auf eine Person, fällt diesem eher auf, ob die Testperson die Instruktion verstanden hat oder nicht. Desweiteren wird auch das individuelle Eingehen auf die Testperson bei der Einzeltestung gefördert (Herle, 2003).

Einen weiteren Vorteil der Einzeltestung bringt der Einsatz von Testmaterialien. Untertests des AID 2 sind zum Beispiel mit Würfeln, Bildkarten oder Puzzles

ausgestattet. Neben dem Einsatz des vielfältigen Testmaterials, bietet die adaptive Testvorgabe² einen ökonomischen Vorteil (Kubinger, 2009a).

3.1.4 Nachteile der Individualtestung

Betreffend das Gütekriterium Objektivität sind Individualverfahren bei der Testleiterunabhängigkeit und Verrechnungssicherheit im Nachteil (Herle, 2003). Kubinger (2009b) weist darauf hin, dass ohne eine Interaktion mit der Testperson eine Einzeltestung nicht möglich wäre. Im Gegensatz zum Gruppenverfahren kann der Testleiter die Testperson absichtlich oder unabsichtlich beeinflussen.

Kelava und Moosbrugger (2007) betonen, dass bei Individualverfahren der Effekt der sozialen Erwünschtheit höher ist als bei Gruppentestungen. Kubinger (2009b) erklärt diesen Effekt folgendermaßen: anhand des Eingehens auf die Testperson kann es seitens des Testleiters zu bestimmten Erwartungen an die Testperson kommen.

Die meisten Einzelverfahren sind nach dem freien Antwortformat gestaltet. Bei manchen Aufgaben werden dem Testleiter mehrere Antwortmöglichkeiten als richtig dargeboten, wodurch ihm gewisse Ermessensfreiheit ermöglicht wird (Kubinger, 2009b).

In der vorliegenden Arbeit geht es nun darum, zu untersuchen, ob die Vorgabe ausgewählter Untertests, des eigentlich als Individualtest angelegten Intelligenztests AID 2, im Gruppensetting zu signifikant anderen Leistungen führt.

² Den Testpersonen werden Fragen gestellt, welche ihren Leistungen entsprechen (Kubinger, 2009b). Das Prinzip des adaptiven Testens wird im Kapitel 5.1. näher beschrieben.

4 Empirischer Teil

4.1 Methode

4.1.1 Fragestellung und Hypothesen

Wie dem theoretischen Teil zu entnehmen ist, bieten Gruppentestungen eine ökonomische und objektive Vorgabe von Verfahren (Kubinger, 2009b; Amelang & Schmidt, 2006; Woike, 2003; Kubinger & Jäger, 2003; Lienert und Raatz, 1994). Das Adaptive Intelligenz Diagnostikum ist als Individualtestung konzipiert (Kubinger, 2009a). Die vorliegende Arbeit soll untersuchen, ob diese Intelligenztestbatterie auch als Gruppentestung einsetzbar ist. Für die Untersuchung wurden die Items der Untertests „Alltagswissen“, „Angewandtes Rechnen“, „Synonyme Finden“, „Funktionen Abstrahieren“ und „Soziales Erfassen und Sachliches Reflektieren“ als Gruppenversion zusammengestellt.

Deshalb steht folgende Fragestellung im Mittelpunkt meiner Untersuchung:

Gibt es Unterschiede zwischen den beiden Verfahren Gruppentestung und Einzeltestung in Bezug auf die Leistungen der Testpersonen?

Folgende Hypothesen können aus der Fragestellung heraus formuliert werden:

Hypothese 1:

HO₁: Die Testart, Gruppentestung oder Einzeltestung, hat keinen Einfluss auf die Testleistung, unabhängig von der Gruppe (Gruppe A: Einzeltestung zuerst; Gruppe B Gruppentestung zuerst).

H1₁: Die Testart, Gruppentestung oder Einzeltestung, hat einen Einfluss auf die Testleistung, unabhängig von der Gruppe.

Hypothese 2:

HO₂: Die Gruppe hat keinen Einfluss auf die Testleistung, unabhängig von der Testart.

H1₂: Die Gruppe hat einen Einfluss auf die Testleistung, unabhängig von der Testart.

Hypothese 3:

HO₃: Es gibt keine Interaktionseffekte zwischen Testart und Gruppe.

H1₃: Es gibt Interaktionseffekte zwischen Testart und Gruppe.

4.1.2 Versuchsplan

Nach Festlegung der Fragestellung und des Testmaterials wurde das Konzept den Direktoren aus niederösterreichischen Volksschulen und Gymnasien vorgestellt und um deren Erlaubnis zu der Testung gebeten. Anschließend wurde der Landesschulrat NÖ kontaktiert.

Die Testungen fanden in folgenden Schulen statt:

- Josef Wondrak Volksschule, Schulweg 4, 2000 Stockerau
- Volksschule Sierndorf, Schulstraße 5, 2011 Sierndorf
- Bundesgymnasium und Bundesrealgymnasium,
Unter den Linden 16, 2000 Stockerau
- Erzbischöfliches Gymnasium und Realgymnasium,
Kirchenplatz 2, 2020 Hollabrunn

Aufgrund von Kontakten mit den Volksschuldirektoren und als ehemalige Absolventin der Gymnasien Stockerau und Hollabrunn wurden diese Schulen für die Datenerhebung ausgewählt.

Jede Testperson wurde entweder der Gruppe A oder der Gruppe B zugeteilt und musste sich einer Gruppentestung und einer Einzeltestung unterziehen (Abb. 6). Als Versuchsbedingung galt, dass ein Abstand von zwei bis drei Tagen zwischen den

beiden Testarten eingehalten werden muss, um Übungseffekte³ auszuschließen. Zusätzlich war es notwendig einen balancierten Vorgabemodus zu berücksichtigen, um Reihenfolgeeffekte auszuschließen. Das „balancierte Testen“ soll anhand eines Beispiels erklärt werden:

Testperson A erhält zuerst die Testart „Einzel“ und nach Einhaltung des Abstandes die Testart „Gruppe“.

Testperson B erhält zuerst die Testart „Gruppe“ und nach Einhaltung des Abstandes die Testart „Einzel“.

Die Reihenfolge der Vorgabe von Gruppentestung und Einzeltestung wurde vollständig ausbalanciert. Die Untersuchungsbedingungen setzten voraus, dass eine adaptive Vorgabe der Items nicht möglich war, somit wurden die mittelschwierigen Items vorgegeben. Diese Vorgehensweise war notwendig, weil Items in der Gruppe nicht adaptiv vorgegeben werden können und beide Vorgaben die gleichen Items beinhalten sollten.

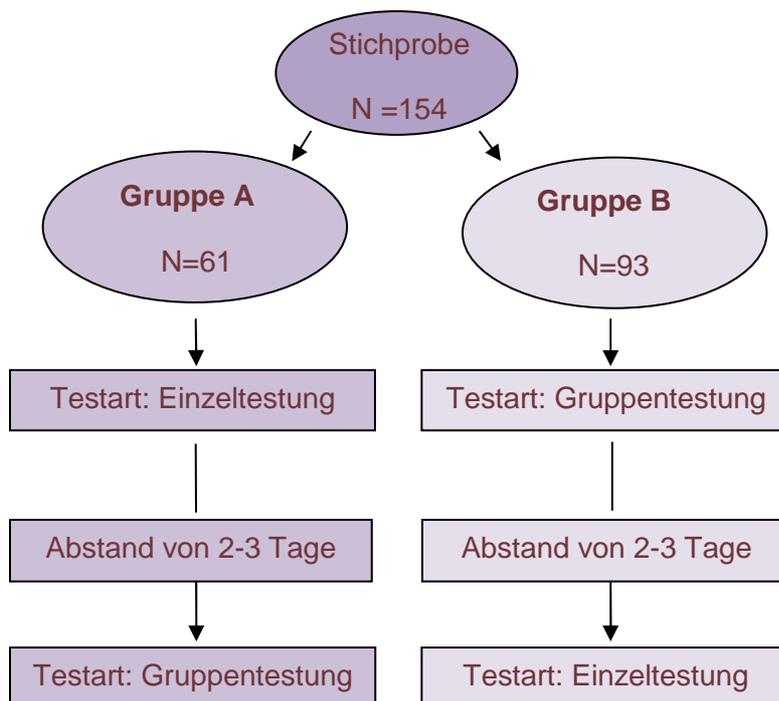


Abb. 6 Grafische Darstellung der Versuchsbedingungen

³ Durch wiederholte Reizdarbietung können Inhalte gelernt werden (Haseloff, 1970).

4.1.3 Testdurchführung

Nach der Zustimmung des Landesschulrates sowie der Direktoren und Lehrer erhielten die Eltern einen Informationsbrief (siehe Anhang). Zusätzlich wurde ausdrücklich auf die Anonymität der Testpersonen hingewiesen. Für etwaige Rückfragen wurde Telefonnummer und Mailadresse der Diplomandin angegeben. Jeder Testperson wurde ein Code zugeteilt um die Anonymität zu gewährleisten.

Die Untersuchung fand im Zeitraum von Mai 2010 bis Juni 2010 statt.

Für die schnellere Durchführung der Untersuchung wurden Testleiterinnen herangezogen, die den AID 2 Zertifizierungskurs abgelegt haben. Der Zertifizierungskurs fand an zwei Tagen statt und umfasste eine theoretische sowie eine praktische Darstellung des AID 2.

Die Gruppentestungen und Einzeltestungen fanden in den Schulen statt. Für die Gruppentestungen stellten die Direktoren leerstehende Klassen, Bibliotheken und Konferenzräume zur Verfügung. In diesen Räumlichkeiten hatte der Testleiter genügend Platz um mehrere Schüler gleichzeitig zu testen. Es bot sich die Möglichkeit, die Kinder auseinander zu setzen, um ein Abschreiben zu verhindern. An der Tür wurde ein Schild montiert, welches Schüler und Lehrer darauf hinwies nicht einzutreten, da eine Testung stattfand. Die Einzeltestungen wurden ebenfalls in diesen Räumlichkeiten durchgeführt. Die Kinder wurden immer einzeln bzw. in Gruppen von den Testleitern aus ihren Klassen geholt. Wie im Manual (Kubinger, 2009a) beschrieben, wurden die Testungen ausschließlich am Vormittag durchgeführt. Der Testleiter stellte sich zu Beginn jeder Testung mit Namen vor und erklärte den Schülern, in ein paar Sätzen, wie die nächsten 20 Minuten ablaufen werden. Als Dankeschön für die Teilnahme an der Untersuchung durften sich die Kinder am Ende jeder Testung eine Süßigkeit aussuchen.

4.2 Verhaltensbeobachtung

4.2.1 Beobachtungen während den Gruppentestungen

Während des gesamten Testzeitraumes kam es bei den Gruppentestungen zu keinen gravierenden Störungen durch andere, so dass Testpersonen das Bearbeiten der Aufgaben abbrechen und den Raum verlassen mussten. Die Probanden wirkten konzentriert und bemüht. Vereinzelt gab es Schüler, die die Instruktion nicht verstanden haben und unsicher wirkten. Sie mussten vom Testleiter nochmal aufgefordert werden, die Instruktion zu lesen. Einige Probanden wirkten abwesend und wurden daher vom Testleiter gebeten, sich auf die Aufgaben zu konzentrieren.

Insbesondere Schüler mit einer anderen Muttersprache waren mit der Testsituation überfordert und haben die Instruktion durch den Testleiter als auch die schriftliche Instruktion fast gar nicht verstanden. Dadurch waren sie in ihren Leistungen eingeschränkt. Im Vergleich zu den älteren Schülern stellten die Volksschüler und auch die Gymnasiasten der 1. Klasse keine Fragen zur Anonymität. Einige Schüler schrieben, weil sie es so vom Unterricht gewohnt sind, ihren Namen auf das Testheft, welcher dann vom Testleiter wieder entfernt wurde. Für die jüngeren Schüler stellte die Testsituation keine Wettbewerbssituation dar. Sie versuchten nicht, in das Testheft von anderen Schülern zu schauen. Es konnte beobachtet werden, dass Schüler in die Luft starrten oder auf ihrem Sessel herumrutschten, wenn eine Aufgabe eine besonders schwierige Herausforderung darstellte oder sie die Testung langweilte. In solchen Situationen befassten sich die jüngeren Kinder eher mit sich selbst, anstatt auf die Tätigkeiten der anderen zu schauen. Nach Ende der Testung war für diese Schüler die Süßigkeit wichtiger als die soeben bearbeiteten Aufgaben. Dies war vor allem dadurch erkennbar, indem sie einen Blick auf die Süßigkeiten erhaschten, die sich andere Mitschüler aussuchten.

Die Testpersonen aus der 3. Klasse Gymnasium und 1. Klasse Oberstufe verhielten sich ganz anders als die jüngeren Testpersonen. Sie fragten vor Beginn der Testung mehrmals nach, ob der ganze Test anonym bleibt. Sie brachten zum Ausdruck, dass es ihnen peinlich wäre, falsche Antworten aufzuschreiben. Deshalb mussten die älteren Schüler vom Testleiter beruhigt werden, dass alles anonym ist und sie alles hinschreiben durften, was für sie richtig erscheint. Es wurde während der Testung einige Male geflucht, gekichert und getratscht. Der Testleiter bat die betreffenden

Personen, nicht mit den anderen zu reden und es kam zu keinen weiteren Störungen. Im Vergleich zu den jüngeren Testpersonen fühlten sich, ganz besonders, die 15-jährigen in einer Wettbewerbssituation. Sie versuchten so schnell wie möglich, ihre Aufgaben zu lösen, warfen immer wieder einen Blick zum Nachbar und tauschten sich sofort nach Einsammeln der Hefte über die Lösungen aus.

Bei den schriftlichen Instruktionen waren einige Kinder mit nicht deutscher Muttersprache benachteiligt. Sie saßen vor allem bei den Untertests „Synonyme Finden“ und „Funktionen Abstrahieren“ vor dem Testheft, ohne eine einzige Aufgabe zu lösen. Ein weiteres Problem bei der Gruppentestung war die Zeitvorgabe. Besonders für leseschwache Schüler war sie ein Hindernis. Während die Zeit schon fast um war, hielten sich diese Kinder noch mit dem Lesen der Instruktion auf. Die beschriebenen Probleme decken sich mit den im Theorieteil aus gängiger Forschung abgeleiteten Nachteilen von Gruppentestungen.

4.2.2 Beobachtungen während den Einzeltestungen

Die Schüler der 3. Klasse Volksschule und des 1. Gymnasiums zeigten große Freude, als sie zur Einzeltestung aufgerufen wurden. Die Schüler, die noch nicht an der Reihe waren, konnten es kaum abwarten bis ihr Name aufgerufen wurde. Sie bearbeiteten die Aufgaben mit Interesse und hatten Spaß daran, der Testleiterin ihr Wissen zu präsentieren. Die Schüler der 3. Klasse Gymnasium und der 1. Klasse Oberstufe waren ebenfalls motiviert, wobei die 15-jährigen Burschen mehr Schamgefühl zeigten als alle anderen Altersgruppen. Konnten die Testpersonen die Aufgabe nicht optimal ausführen, war eine verringerte Konzentrationsfähigkeit festzustellen.

4.3 Probleme bei der Datenerhebung

Die Motivation war bei fast allen Schülern der Volksschule und des Gymnasiums sehr gut. Der Kontakt mit den Lehrern und Direktoren verlief reibungslos. Vom Lehrpersonal konnten genügend Zimmer zur Verfügung gestellt werden. An einigen Tagen wurde von den Lehrern ein wichtiger Schulstoff durchgemacht. Deshalb war es unmöglich in dieser Stunde zu testen. Dadurch ging wertvolle Zeit verloren und man musste bis zur nächsten Stunde warten.

Bis auf die 15-jährigen Schüler tauschte sich niemand über die Lösungen der Aufgaben aus.

5 Untersuchungsmaterial

5.1 Das Adaptive Intelligenz Diagnostikum 2 (Kubinger, 2009a)

Das Adaptive Intelligenz Diagnostikum 2 (AID 2) dient der Erfassung kognitiver Fähigkeiten von Kindern und Jugendlichen im Alter von 6,0 bis 15,11 Jahren. Es wurde speziell für die Schulpsychologie und zur Diagnostik von Teilleistungsstörungen und Hochbegabung entwickelt und als Einzelvorgabe konzipiert. Der AID 2 setzt sich aus Power-Tests zusammen und verwendet ein freies Antwortformat (Kubinger, 2009a).

Laut Manual (Kubinger, 2009a) besteht das Adaptive Intelligenz Diagnostikum aus insgesamt elf Untertests und drei Zusatztests, welche die zwei Bereiche „Verbalakustische Fähigkeiten“ und „Manuell-visuelle Fähigkeiten“ abdecken. Mit diesen Untertests werden viele basale und kognitive Fähigkeiten erhoben, welche für „intelligentes“ Verhalten verantwortlich sein sollen.

Mittels Verwendung von Leistungsprofilen in allen Untertests bzw. mittels Interpretationen von Leistungshöhen und Leistungstiefen ist eine förderungsorientierte Diagnostik möglich. Im Sinne von Cattells (1963) fluider und kristalliner Intelligenz unterscheidet man zwischen förderungsabhängigen und förderungsunabhängigen Untertests. Somit wird im AID 2 kein Intelligenzquotient berechnet (Kubinger, 2009a).

Als globales Maß der kognitiven Fähigkeiten werden laut Kubinger (2009a) das Leistungsminimum und das Leistungsmaximum festgelegt. Das Leistungsminimum wird als „Mindestfähigkeit“ einer Person definiert und dient zur globalen Beurteilung der „Intelligenz“.

Beim Vergleich unterschiedlicher Intelligenztests für Kinder heben Preusche und Leiss (2003) die zusätzliche Berechnung des „Ranges“ hervor, welcher als Indikator für Teilleistungsstörungen eingesetzt werden kann.

Inhaltlich orientiert sich das AID 2 an dem Testkonzept von David Wechsler (1956). Die einzelnen Aufgabentypen sowie die Einteilung in zwei Testteile sind mit dem Hamburger-Wechsler-Intelligenztest identisch. Zusätzlich wurden aber einige Untertests erweitert (Jacobs, Heubrock & Petermann, 2003).

Methodisch führt die Testbatterie das Prinzip des adaptiven Testens, durch Geltung des Rasch-Modells, aus.

Der Begriff „adaptives Testen“ zielt darauf ab, „nicht allen Testpersonen dieselben Aufgaben vorzugeben, sondern nur diejenigen, welche der individuellen Fähigkeit angemessen sind“ (Kubinger, 2009b, S.102)

Eine Variante des adaptiven Testens ist die Methode des „branched-testing“ (Abb.7). Dabei werden die einzelnen Aufgaben zu Aufgabengruppen oder -blöcken zusammengefasst (Kubinger, 2009b). Die Vorgabe der Aufgabengruppen ist davon abhängig, welche Leistungen die Testperson in den vorangegangenen Aufgaben erbringt. Die Anzahl der gelösten Aufgaben in einem Block begründet, ob zu einer leichteren oder schwereren Aufgabengruppe weiter verzweigt wird (Kubinger, 2009a).

Einen Vorteil in der Methode des adaptiven Testens sehen Preusche und Leiss (2003) darin, dass die Enttäuschung bei Kindern seitens ihrer Leistungen relativ gering gehalten werden kann. Jacobs et al. (2003) stellen eine bessere Durchführungsdauer und Messgenauigkeit gegenüber anderen vergleichbaren Verfahren fest.

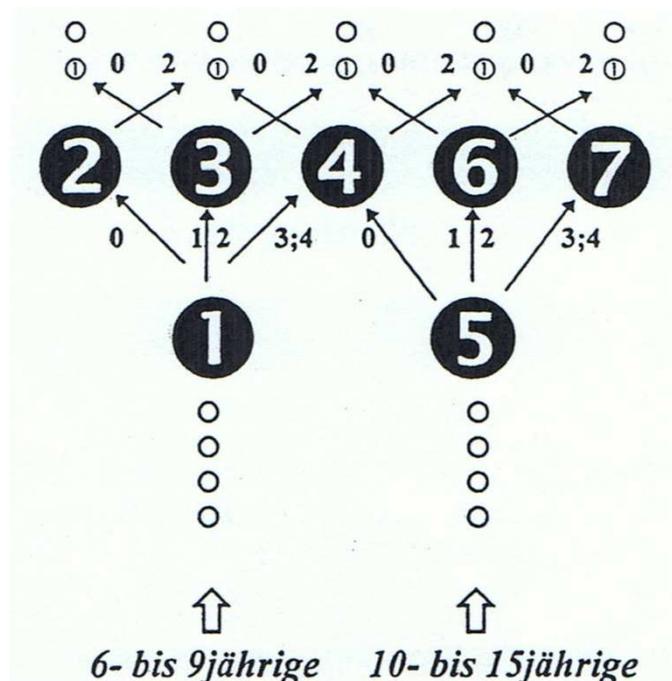


Abb. 7 Testvorgabe des Untertests Soziale und Sachliche Folgerichtigkeit und Analysieren und Synthetisieren-abstrakt (Kubinger & Wurst, 2000)

In der vorliegenden Studie wurde auf eine aktualisierte, zum Zeitpunkt der Diplomarbeitverfassung noch nicht publizierte Version des AID 2, Arbeitstitel AID 3, zurückgegriffen. Dabei wurden folgende Untertests vorgegeben:

Untertest A „Alltagswissen“: Über Fragen, soll die Fähigkeit überprüft werden, sich Sachkenntnisse und Inhalte anzueignen, die aus der heutigen Gesellschaft alltäglich sind.

Untertest E „Angewandtes Rechnen“: Mittels alltäglicher Rechenbeispiele sollen numerisch einfach gehaltene Aufgaben gelöst werden. Hier wird die Fähigkeit zur Anwendung von Rechenoperationen untersucht.

Untertest F „Synonyme finden“: Für einen vorgegeben Begriff soll ein anderes Wort mit derselben Bedeutung gefunden werden. In diesem Untertest wird der Wortschatz überprüft.

Untertest H „Funktionen Abstrahieren“: Das Kind soll die gemeinsame Funktion von jeweils zwei genannten Objekten erkennen und beschreiben. Hier handelt es sich um die Fähigkeit, durch Abstraktionen zu einer Begriffsbildung zu gelangen.

Untertest I „Soziales Erfassen und Sachliches Reflektieren“: Durch Fragen soll überprüft werden, inwieweit das Kind über gesellschaftliche Zusammenhänge und soziale Verhaltensweisen informiert ist.

Die Untertests „Alltagswissen“, „Synonyme Finden“ und „Soziales Erfassen und Sachliches Reflektieren“ sind förderungsabhängig. Die anderen Untertests sind im Sinne der fluiden Intelligenz. „Soziales Erfassen und Sachliches Reflektieren“ sind nach Cattells (1963) kristallinen Intelligenz.

5.2 Entwicklung der Gruppenvorgabe

Hier soll nun auf die Entwicklung der Gruppenvorgabe in Form eines Testhefts näher eingegangen werden sowie auf die abgeänderte Einzelvorgabe.

Da bei einer Gruppenvorgabe die Instruktion schriftlich erfolgt, sollte die Testperson in der Lage sein diese zu lesen. Deshalb wurde der ursprüngliche Altersbereich auf 8,0 bis 15,11 eingeschränkt. Die erste Überlegung bestand darin, die Kinder durch eine zulange Instruktion nicht zu überfordern. Weiters sollte eine Interaktion zwischen den Testpersonen und dem Testleiter so gering wie möglich gehalten werden. Somit musste gleichzeitig alles so verpackt sein, dass die Kinder verstehen was zu tun ist und nicht dazu verleitet werden Fragen zu stellen.

Schließlich wurden die Instruktionen zu den fünf Untertests aus dem Manual übernommen und ergänzt bzw. verkürzt. Es war vorgesehen, dass die Testpersonen die Instruktionen selbst lesen mussten und es keine Hilfestellung durch den Testleiter gibt.

Bei den beiden Untertests „Synonyme Finden“ und „Funktionen Abstrahieren“ wurde das Übungsbeispiel mit Lösung angeführt. Alle, für die Gruppentestungen abgewandelten, Instruktionen werden in Abb. 8 bis Abb. 12 dargestellt.

*Hier sind einige Aufgaben und Fragen für dich.
Manche davon sind leicht, manche sind schwieriger.
Du hast genug Zeit zum Nachdenken.
Lies die Fragen genau durch und schreibe uns deine Antwort auf.
Bitte blättere erst um, wenn du dazu aufgefordert wirst.*

Abb. 8 Testinstruktion „Alltagswissen“

*Nun kommen einige Textrechnungen.
Lies jede Aufgabe sorgfältig durch.
Du hast genug Platz für Nebenrechnungen.
Trage dein Ergebnis **IMMER** ins leere Kästchen ein.*

Abb. 9 Testinstruktion „Angewandtes Rechnen“

Hier stehen einige Wörter aufgeschrieben und wir wollen sehen, wie viele du davon kennst. Für jedes Wort sollst du ein anderes finden, das das gleiche bedeutet.

Zum Beispiel:

1. finster

dunkel

Finster und dunkel bedeutet das gleiche.

Warte ab bis du beginnen darfst.

Abb. 10 Testinstruktion des Untertests „Synonyme Finden“

Es werden immer zwei Dinge genannt und du sollst herausfinden, was sie gemeinsam haben oder was man mit beiden tun kann.

Ein Beispiel:

1.Korb – Tasche

Man kann mit beiden etwas tragen

Auf ein Zeichen dürft ihr beginnen.

Abb. 11 Testinstruktion „Funktionen Abstrahieren“

Nun kommen einige Fragen zu denen du alles aufschreiben kannst, was dir dazu einfällt.

Bitte blättere erst um, wenn du dazu aufgefordert wirst.

Abb. 12 Testinstruktion „Soziales Erfassen und Sachliches Reflektieren“

Im Untertest „Synonyme Finden“ wurden die Schüler dazu aufgefordert, ihre Lösungen in das dafür vorgesehene Kästchen zu schreiben. Befand sich die Lösung außerhalb des Kästchens, wurde die Antwort als „nicht gelöst“ gewertet, darüber wurden die Probanden vor Testbeginn informiert.

Um einen Vergleich zwischen Gruppen- und Originalvorgabe zu erhalten, musste auch das Gruppenverfahren nach dem freien Antwortformat gestaltet werden.

Um der korrekten Gestaltungsweise einer Gruppenvorgabe gerecht zu werden, wurde jeder Untertest inklusive Instruktion mit einem Zeitlimit versehen:

Untertest A „Alltagwissen“: 2 Minuten

Untertest E „Angewandtes Rechnen“: 6 Minuten

Untertest F „Synonyme Finden“: 2 Minuten

Untertest H „Funktionen Abstrahieren“: 2 Minuten

Untertest I „Soziales Erfassen und Sachliches Reflektieren“: 2 Minuten und 30 Sekunden

Vor Beginn der Untersuchung wurde die Gruppentestung mit dem festgelegten Zeitlimit an ungefähr zehn Volksschülern getestet. Dabei stellte sich heraus, dass die Zeitvorgabe nicht ausreichend war. Der Großteil der Testpersonen war nach 2 Minuten erst beim Beantworten der ersten Aufgabe.

Folgende Änderungen wurden danach vorgenommen:

Untertest A „Alltagwissen“: 3 Minuten

Untertest E „Angewandtes Rechnen“: 6 Minuten

Untertest F „Synonyme Finden“: 3 Minuten

Untertest H „Funktionen Abstrahieren“: 3 Minuten

Untertest I „Soziales Erfassen und Sachliches Reflektieren“: 5 Minuten

Jeder Untertest wurde in einer anderen Farbe gedruckt. Auf diese Weise soll der Testleiter darauf aufmerksam gemacht werden, wenn Testpersonen im Testheft (siehe Anhang) herumblättern, um Aufgaben eines anderen Untertests vorzeitig zu bearbeiten. Deshalb gab es am Ende jedes Untertests folgenden schriftlichen Hinweis: *„Warte bitte bis du umblättern darfst!“*

Die Durchführung der Gruppentestung dauerte ca. 20 Minuten und umfasste vier bis sechs Personen.

5.3 Einzelvorgabe des AID 3

Im Kapitel über die Testbeschreibung des AID 2 wurde kurz erwähnt, dass die Vorgabe im AID 2 nach der Methode des adaptiven Testens erfolgt. Die adaptive Vorgabe ist in der Gruppentestung nicht realisierbar. Aus diesem Grund wird in der Einzelvorgabe nicht adaptiv getestet. Es wird davon ausgegangen, dass die Kinder im Durchschnitt zwei bis drei Items pro Aufgabenblock lösen. Folglich sind bei beiden Vorgaben (einzeln und in der Gruppe) die Items gleich. Die Einzeltestung nahm etwa 20 Minuten in Anspruch.

Die Antworten in beiden Vorgabemodi wurden nach „gelöst“ oder „nicht gelöst“ ausgewertet.

5.4 Testinstruktion

Zu Beginn der Gruppentestung und auch der Einzeltestung wurden die Teilnehmer darüber informiert, dass niemand Einsicht in die Ergebnisse haben wird. Danach wurden die Schüler gebeten, sich während der Testung leise zu verhalten.

Würde es zu einer massiven Störung durch eine Person kommen, müsse sie den Raum verlassen und die Testung beenden. Danach instruierte der Testleiter, anhand eines festgelegten Gesprächsleitfadens (siehe Anhang), den Ablauf der Gruppentestung. Anschließend wurden die Testhefte ausgeteilt und auf ein Zeichen „*Ihr dürft jetzt anfangen*“, durfte die Instruktion für den Untertest A durchgelesen werden. Hatten alle Schüler die Instruktion gelesen, wurden sie dazu angeregt mit dem Bearbeiten der Aufgaben zu beginnen. Der Testleiter stoppte die Zeit. Nach Ablauf der Zeit mussten die Schüler das Lösen der Aufgaben beenden und zur Instruktion des nächsten Untertests blättern. Der Testleiter machte die Testpersonen auf das Ablaufen der Zeit aufmerksam und gab die Anweisung weiterzublättern.

Während der Testung sollte sich der Testleiter an den Gesprächsleitfaden (siehe Anhang) halten und so wenig wie möglich davon abweichen.

Die Instruktionen der Einzeltestung erfolgten nach dem Manual (Kubinger, 2009b).

6 Darstellung der Ergebnisse

6.1 Stichprobenbeschreibung

In diesem Abschnitt wird die Stichprobe anhand der soziodemografischen Daten beschrieben. Erhoben wurde das Geschlecht, das Alter, der Schultyp und die Muttersprache. Es wurde untersucht, ob sich die beiden Gruppen anhand der erhobenen Merkmale signifikant voneinander unterscheiden oder ob die Merkmale annähernd gleich verteilt sind.

Geschlecht

Für die vorliegende Untersuchung konnten die Daten von insgesamt 154 Schülern und Schülerinnen aus Niederösterreich herangezogen werden, die zu beiden Zeitpunkten der Testung anwesend waren. Davon waren 59 männlich und 95 weiblich.

In Gruppe A befanden sich 20 Schüler und 41 Schülerinnen, in Gruppe B 39 Schüler und 54 Schülerinnen.

Der Chi-Quadrat-Test nach Pearson ergab mit $p=.253 (>.05)$ kein signifikantes Ergebnis. Somit kann angenommen werden, dass das Geschlecht in beiden Versuchsbedingungen annähernd gleich verteilt ist.

Tabelle 1 veranschaulicht die Verteilung des Geschlechts.

Tab. 1 Verteilung des Geschlechts

	Männlich	Weiblich	Gesamt
Gruppe A	20	41	61
Gruppe B	39	54	93
Gesamt	59	95	154

Schultyp

Insgesamt nahmen 65 VolksschülerInnen und 89 GymnasiastInnen an der Untersuchung teil.

Die Verteilung hinsichtlich des Schultyps in den beiden Gruppen fällt unterschiedlich aus. Auffällig ist, dass die Volksschüler in der Versuchsbedingung 1 unterrepräsentiert waren.

In der Volksschule waren 19 SchülerInnen der Gruppe A zugeteilt und 46 SchülerInnen Gruppe B. Im Gymnasium wurden 42 SchülerInnen Gruppe A und 47 SchülerInnen Gruppe B zugeordnet (Tab. 2)

Der Chi-Quadrat-Test nach Pearson ergab mit $p=.024$ ($<.05$) einen signifikanten Unterschied bezüglich der Verteilung des Schultyps in den beiden Gruppen.

Tab. 2 Verteilung des Schultyps

	Volksschule	Gymnasium	Gesamt
Gruppe A	19	42	61
Gruppe B	46	47	93
Gesamt	65	89	154

Alter

Das Alter wurde mittels Geburtsdatum der SchülerInnen erhoben.

Der Mittelwert des Alters lag über die gesamte Stichprobe verteilt bei 11,3 Jahren.

In Gruppe A liegt der Mittelwert bei 11,73 Jahren, das Minimum bei 8 Jahren und das Maximum bei 17 Jahren. In Gruppe B liegt der Mittelwert bei 11,12 Jahren, das Minimum bei 6 Jahren und das Maximum bei 15 Jahren.

Der *t*-Test für unabhängige Stichproben ergab mit $p=.448$ ($>.05$) keinen signifikanten Unterschied zwischen den beiden Versuchsbedingungen bezogen auf das Alter.

Muttersprache

Von den insgesamt 154 SchülerInnen hatten 143 Kinder Deutsch als Muttersprache, 1 Kind Türkisch, 5 Kinder Kroatisch/Serbisch/Slowenisch und 1 Kind Polnisch als Muttersprache (Tab. 3). Der Großteil der Kinder mit nicht deutscher Muttersprache befand sich in der Volksschule.

Tab. 3 Verteilung der Muttersprache

	Gruppe A	Gruppe B	Gesamt
Deutsch	58	85	143
Türkisch	0	1	1
Kroatisch/Serbisch/ Slowenisch	1	4	5
Polnisch	0	1	1
Sonstige	2	2	4
Gesamt	61	93	154

6.2 Statistische Auswertung

Für die statistischen Auswertungen wurde das PC-Programm „SPSS Statistics (Statistical Packages for the Social Sciences)“ in der Version 18.0 verwendet.

Um das Alter in den Untersuchungsgruppen zu überprüfen, wurde ein *t*-Test für unabhängige Stichproben durchgeführt, damit mögliche Unterschiede in den untersuchten abhängigen Variablen bzw. Effekte nicht auf Unterschiede in den beiden Versuchsbedingungen zurückzuführen sind.

Für die Überprüfung der Forschungshypothesen wurde eine **Multivariate Varianzanalyse** (MANOVA) und eine **Zweifache Varianzanalyse** mit Messwiederholung errechnet. Die Multivariate Varianzanalyse überprüft, ob sich zwei oder mehrere Stichproben hinsichtlich mehrerer abhängiger Variablen unterscheiden. Darüber hinaus können eventuelle wechselseitige Beziehungen aufgezeigt werden. Im Anschluß an eine signifikante MANOVA werden univariate Varianzanalysen berechnet. Eine Zweifache Varianzanalyse überprüft, ob sich eine unabhängige Variable hinsichtlich eines Messwiederholungsfaktors (abhängige Variable) unterscheidet (Rasch & Kubinger, 2006). Als unabhängige Variable können in dieser Studie die beiden Arten der Vorgabe genannt werden. Die unabhängigen Variablen werden in die beiden Gruppen A und B unterteilt. Dabei wurden die Versuchspersonen der Gruppe A zuerst mit dem Einzelverfahren getestet und die Versuchspersonen der Gruppe B zuerst mit dem Gruppenverfahren. Als abhängige Variable werden die Testleistungen der Personen in den Subtests zu den zwei Testarten (Gruppentestung oder Einzeltestung) definiert, hierfür wurden die Fähigkeitsparameter zur Berechnung verwendet.

Als Signifikanzniveau wurde ein Alpha Niveau von 5% festgelegt, d.h. bei einem p kleiner oder gleich .05 spricht man von einem signifikanten Ergebnis. Die Effektgröße ist ein Maß für die praktische Relevanz von signifikanten Ergebnissen (Bortz, 2005). Bei Berechnung einer Varianzanalyse wird das Partielle Eta- Quadrat (η^2) als Maß für die Effektstärke herangezogen. Die Effektgröße lässt sich auch als prozentuellen Anteil der erklärten Varianz ausdrücken (Bortz, 2005).

6.3 Ergebnisse

6.3.1 Ergebnisse der Multivariaten Varianzanalyse

Der Box-M-Test zur Überprüfung der Varianz-Kovarianz-Matrix war nicht signifikant ($F=1,395$, $p=0,058$).

H_0 = Über alle Subtests hinweg gibt es keine Unterschiede zwischen den Gruppen.

H_1 = Über alle Subtests hinweg gibt es Unterschiede zwischen den Gruppen.

Die Höhe der Personenparameter, in ihren einzelnen Subtests, unterscheidet sich signifikant ($F=46,129$; $p<0,001$; $\eta^2=0,763$), in Abhängigkeit von den beiden Gruppen. Die H_1 wird angenommen und besagt, dass über alle Subtests hinweg signifikante Unterschiede zwischen den Gruppen A und B bestehen.

Tabelle 4 repräsentiert die univariaten Analysen die aufgrund des signifikanten Ergebnisses in der MANOVA errechnet werden. Aus der Tabelle ist zu entnehmen, dass in fast allen Untertests signifikante Unterschiede zwischen den Gruppen A und B vorliegen. Nur im Untertest „Alltagswissen“ zum Testzeitpunkt 2 ($F=1,201$; $p=0,275$) und im Untertest „Synonyme Finden“ zum Testzeitpunkt 1 ($F=2,005$; $p=0,159$) besteht kein signifikanter Unterschied zwischen den beiden Gruppen.

Tab. 4 Ergebnisse der Zwischensubjekteffekte der multivariaten Varianzanalyse

	df	Mittel der Quadrate	F-Wert	Signifikanz	Partielles η^2
PP „Alltagswissen“ Gruppe A Einzel Gruppe B Gruppe	1	22,973	4,377	0,038	0,028
PP „Alltagswissen“ Gruppe A Gruppe Gruppe B Einzel	1	5,931	1,201	0,275	0,008
PP „Angewandtes Rechnen“ Gruppe A Einzel Gruppe B Gruppe	1	168,955	34,727	<0,001	0,186

PP „Angewandtes Rechnen“ Gruppe A Gruppe Gruppe B Einzel	1	20,751	4,896	0,028	0,31
PP „Synonyme Finden“ Gruppe A Einzel Gruppe B Gruppe	1	8,712	2.005	0,159	0,013
PP „Synonyme Finden“ Gruppe A Gruppe Gruppe B Einzel	1	51,461	12,897	<0,001	0,078
PP „Funktionen Abstrahieren“ Gruppe A Einzel Gruppe B Gruppe	1	101,653	38,524	<0,001	0,202
PP „Funktionen Abstrahieren“ Gruppe A Gruppe Gruppe B Einzel	1	172,335	59,701	<0,001	0,282
PP „Soziales Erfassen und Sachliches Reflektieren“ Gruppe A Einzel Gruppe B Gruppe	1	54,004	19,270	<0,001	0,113
PP „Soziales Erfassen und Sachliches Reflektieren“ Gruppe A Gruppe Gruppe B Einzel	1	94,874	30,746	<0,001	0,168

*signifikant $p < 0,05$; df=Freiheitsgrade; η^2 = Effektstärke; PP=Personenparameter

Aus Tabelle 5 ist zu entnehmen, dass in allen Untertests, unabhängig davon ob zuerst in der Gruppe oder im Einzelsetting getestet wurde, die Einzeltestungen besser waren als die Gruppentestungen.

Tab. 5 Beschreibung der Mittelwerte in den Gruppen

		MW der Personenparameter	SA
Untertest „Alltagwissen“	Gruppe A Einzeltestung	1,045	2,037
	Gruppe B Gruppentestung	0,255	2,441
Untertest „Alltagwissen“	Gruppe A Gruppentestung	0,367	2,083
	Gruppe B Einzeltestung	0,768	2,307
Untertest „Angewandtes Rechnen“	Gruppe A Einzeltestung	2,644	1,946
	Gruppe B Gruppentestung	0,502	2,359
Untertest „Angewandtes Rechnen“	Gruppe A Gruppentestung	1,159	2,166
	Gruppe B Einzeltestung	1,910	1,985
Untertest „Synonyme Finden“	Gruppe A Einzeltestung	-0,478	1,834
	Gruppe B Gruppentestung	-0,964	2,232
Untertest „Synonyme Finden“	Gruppe A Gruppentestung	-1,188	1,980
	Gruppe B Einzeltestung	-0,007	2,008
Untertest „Funktionen Abstrahieren“	Gruppe A Einzeltestung	0,291	1,340
	Gruppe B Gruppentestung	-1,369	1,785
Untertest „Funktionen Abstrahieren“	Gruppe A Gruppentestung	-1,347	1,646
	Gruppe B Einzeltestung	0,815	1,732
Untertest „Soziales Erfassen und Sachliches Reflektieren“	Gruppe A Einzeltestung	0,809	1,567
	Gruppe B Gruppentestung	-0,401	1,740
Untertest „Soziales Erfassen und Sachliches Reflektieren“	Gruppe A Gruppentestung	-0,521	1,790
	Gruppe B Einzeltestung	1,083	1,733

MW=Mittelwert; SA=Standardabweichung

6.3.2 Ergebnisse der Zweifachen Varianzanalyse mit Messwiederholung

Ergebnisse - Untertest „Alltagswissen“

Der Untertest „Alltagswissen“ erfasst die Fähigkeit sich Sachkenntnisse anzueignen, die in der heutigen Gesellschaft alltäglich sind (Kubinger, 2009a).

HO₁: Die Testart hat keinen Einfluss auf die Testleistung im Untertest „Alltagwissen“, unabhängig von der Gruppe.

H1₁: Die Testart hat einen Einfluss auf die Testleistung im Untertest „Alltagwissen“, unabhängig von der Gruppe

HO₂: Die Gruppe hat keinen Einfluss auf die Testleistung im Untertest „Alltagwissen“, unabhängig von der Testart.

H1₂: Die Gruppe hat einen Einfluss auf die Testleistung im Untertest „Alltagwissen“, unabhängig von der Testart.

HO₃: Es gibt keine Interaktionseffekte zwischen Testart und Gruppe im Untertest „Alltagwissen“.

H1₃: Es gibt Interaktionseffekte zwischen Testart und Gruppe im Untertest „Alltagwissen“.

Berechnungen der Mixed ANOVA (Tab. 6) zeigen einen signifikanten Haupteffekt der Testart ($F=49,677$; $p<0,001$; $\eta^2=0,246$). Es kann davon ausgegangen werden, dass die Art der Testung Einfluss auf die Leistungen im Untertest „Alltagswissen“ hat. Unabhängig von der Gruppe sind in der Einzeltestung die Leistungen generell besser als in der Gruppentestung. Die Interaktion zwischen der Testart und der Gruppe ($F=0,943$; $p=0,333$) hat keinen signifikanten Einfluss auf die Leistung.

Tab. 6 Ergebnisse Mixed ANOVA des Untertests „Alltagswissen“

	df	Mittel der Quadrate	F-Wert	Signifikanz	Partielles Eta ²
Testart * Gruppe	1	0,496	0,943	0,333	0,006
Gruppe	1	2,779	0,288	0,592	0,002
Testart	1	26,124	49,677	<0,001	0,246

*signifikant $p < 0,05$; df=Freiheitsgrade; Eta²= Effektstärke

Aus der Grafik (Abb.13) ist zu entnehmen, dass sich in Gruppe A, vom 1. zum 2. Testzeitpunkt, also zwischen Einzel- und Gruppenvorgabe, die Leistungen verschlechtern und in Gruppe B, in der zuerst in der Gruppe und dann im Einzelsetting getestet wurde, verbessern sich die Leistungen. D.h., dass in beiden Gruppen die Versuchspersonen in der Gruppenvorgabe schlechtere Werte erzielen als in der Einzelvorgabe, unabhängig davon, ob zuerst in der im Einzelsetting oder in der Gruppe getestet wurde.

Desweiteren ist ersichtlich, dass die Leistungen in der Gruppe A generell besser sind als in der Gruppe B.

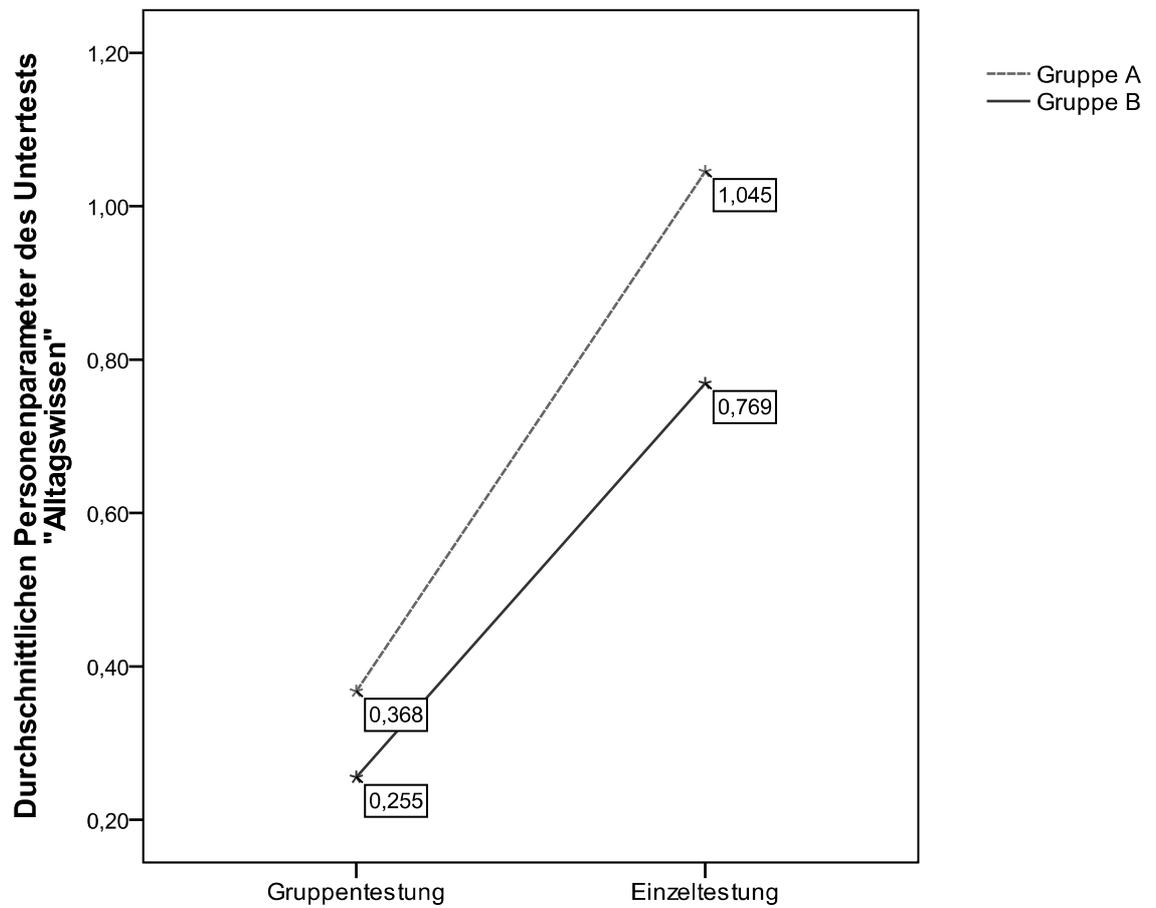


Abb. 13 Mittelwerte des Untertests „Alltagwissen“ Testart x Gruppe

Die Mittelwerte in Tabelle 7 zeigen nochmals, dass die Bearbeitung der Items in der Einzeltestung leichter war als in der Gruppentestung.

Tab. 7 Darstellung der geschätzten Randmittel des Untertests „Alltagswissen“

		MW der Personenparameter	SE
Gruppe x Testart	Gruppe A Einzeltestung	1,045	0,293
	Gruppe B Gruppentestung	0,368	0,284
	Gruppe A Gruppentestung	0,255	0,238
	Gruppe B Einzeltestung	0,769	0,230

MW=Mittelwert; SE=Standardfehler

Kurze Zusammenfassung der Ergebnisse

Im Untertest A „Alltagswissen“ werden die beiden H_{02} und H_{03} anlässlich der nicht signifikanten Ergebnisse beibehalten. Das signifikante Ergebnis beim Hauptfaktor „Testart“ führt zu einem Verwerfen der H_{01} und zu einer Annahme von H_{11} , welche besagt, dass in der Gruppentestung schlechtere Leistungen erzielt wurden als im Einzelsetting.

Ergebnisse – Untertest „Angewandtes Rechnen“

Der Untertest „Angewandtes Rechnen“ misst die Fähigkeit Rechenaufgaben durch logische Schlussfolgerungen zu lösen (Kubinger, 2009a).

HO₁: Die Testart hat keinen Einfluss auf die Testleistung im Untertest „Angewandtes Rechnen“, unabhängig von der Gruppe.

H1₁: Die Testart hat einen Einfluss auf die Testleistung im Untertest „Angewandtes Rechnen“, unabhängig von der Gruppe.

HO₂: Die Gruppe hat keinen Einfluss auf die Testleistung im Untertest „Angewandtes Rechnen“, unabhängig von der Testart.

H1₂: Die Gruppe hat einen Einfluss auf die Testleistung im Untertest „Angewandtes Rechnen“, unabhängig von der Testart.

HO₃: Es gibt keine Interaktionseffekte zwischen Testart und Gruppe im Untertest „Angewandtes Rechnen“.

H1₃: Es gibt Interaktionseffekte zwischen Testart und Gruppe im Untertest „Angewandtes Rechnen“.

Die Ergebnisse der Mixed ANOVA (Tab. 8) veranschaulichen einen Trend beim Faktor „Vorgabemodus“ ($F=4,205$; $p=0,042$; $\text{Eta}^2= 0,027$) und einen signifikanten Haupteffekt der Testart ($F=245,701$; $p<0,001$; $\text{Eta}^2= 0,618$). Nach Cohen (1988) liegt mit 61,8% ein starker Effekt der Testart vor. 61,8% der Gesamtvarianz wird durch die Testart erklärt. Zusätzlich illustrieren die Ergebnisse eine Wirkung der Gruppe, unabhängig von der Testart, auf die Testleistungen, d.h. es existieren Unterschiede in der Leistung, die abhängig von der Reihenfolge der Vorgabe sind.

Tab. 8 Ergebnisse Mixed ANOVA des Untertests „Angewandtes Rechnen“

	df	Mittel der Quadrate	F-Wert	Signifikanz	Partielles Eta ²
Testart * Gruppe	1	0,108	0,172	0,679	0,0061
Gruppe	1	35,642	4,205	0,042	0,027
Testart	1	154,064	245,701	<0,001	0,618

*signifikant $p < 0,05$; df=Freiheitsgrade; Eta^2 = Effektstärke

Wie schon beim Untertest „Alltagswissen“ zeigen sich auch hier die Einzeltestungen unabhängig davon, ob sie vor oder nach der Gruppentestung stattfanden, signifikant leichter. Darüber hinaus erzielte, wie es sich schon für den Untertest „Alltagswissen“ in der Abbildung 13 erkennen lies, auch hier jene Gruppe, in der zuerst im Einzelsetting und dann im Gruppensetting getestet wurde, zu beiden Bedingungen bessere Ergebnisse (Abb. 14 und Tab. 9).

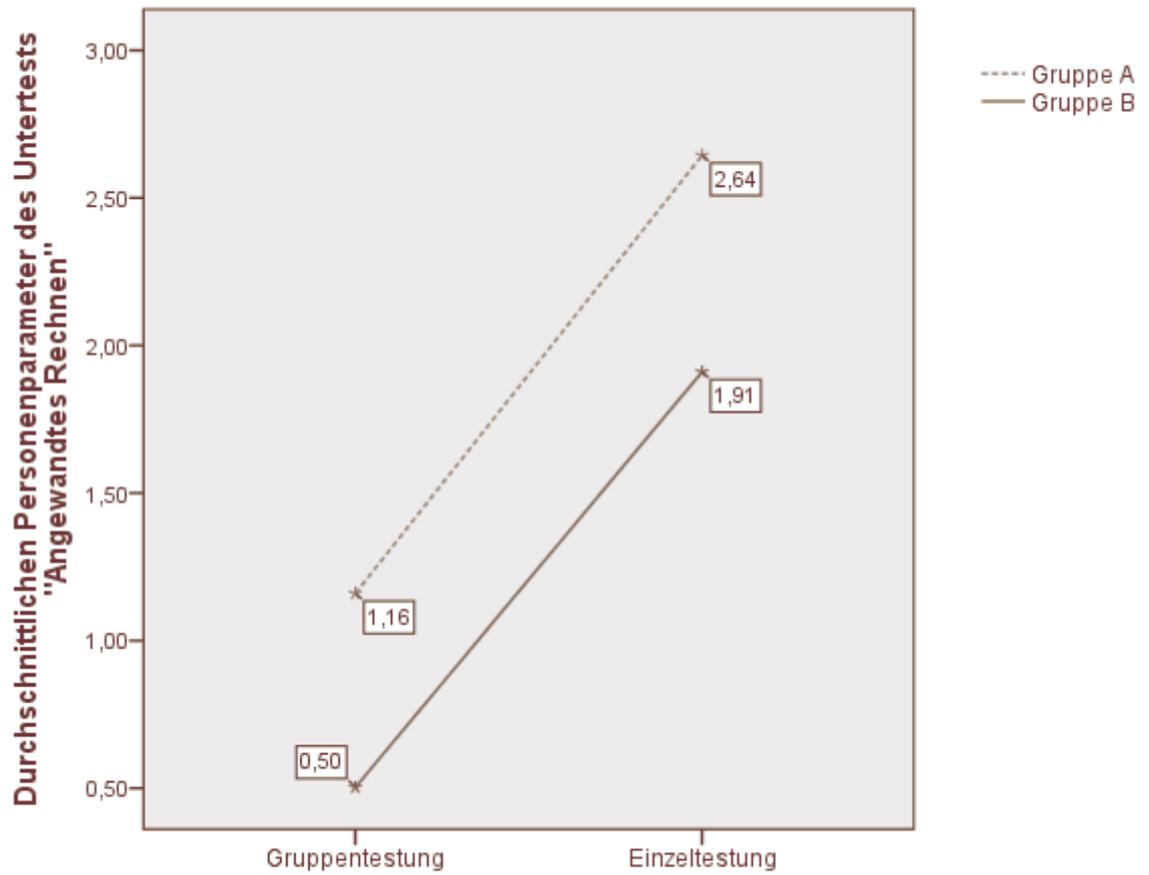


Abb. 14 Mittelwerte des Untertests „Angewandtes Rechnen“ Testart x Gruppe

Tab. 9 Darstellung der geschätzten Randmittel des Untertests „Angewandtes Rechnen“

		MW der Personenparameter	SE
Gruppe x Testart	Gruppe A Einzeltestung	2,644	0,282
	Gruppe B Gruppentestung	1,160	0,264
	Gruppe A Gruppentestung	0,503	0,229
	Gruppe B Einzeltestung	1,910	0,213

MW=Mittelwert; SE=Standardfehler

Kurze Zusammenfassung der Ergebnisse

H_{03} wird im Untertest „Angewandtes Rechnen“ beibehalten, da es zu einem nicht signifikanten Ergebnis kam. H_{01} und H_{02} werden, zwecks der signifikanten Ergebnisse, verworfen und H_{11} und H_{12} angenommen.

H_{11} nimmt an, dass die Testart, unabhängig von der Gruppe, einen Einfluss auf die Leistung hat. In der Gruppentestung kommt es zu schlechteren Ergebnissen als in der Einzeltestung.

H_{12} besagt, dass die beiden Gruppen, unabhängig von der Testart, einen Einfluss auf die Leistung haben. Gruppe A erreicht höhere Werte als Gruppe B.

Ergebnisse – Untertest „Synonyme Finden“

Der Untertest „Synonyme Finden“ erfasst den Wortschatz. Vorgegebene Begriffe sollen durch Begriffe mit gleicher Bedeutung ersetzt werden (Kubinger, 2009a).

HO₁: Die Testart hat keinen Einfluss auf die Testleistung im Untertest „Synonyme Finden“, unabhängig von der Gruppe.

H1₁: Die Testart hat einen Einfluss auf die Testleistung im Untertest „Synonyme Finden“, unabhängig von der Gruppe.

HO₂: Die Gruppe hat keinen Einfluss auf die Testleistung im Untertest „Synonyme Finden“, unabhängig von der Testart.

H1₂: Die Gruppe hat einen Einfluss auf die Testleistung im Untertest „Synonyme Finden“, unabhängig von der Testart.

HO₃: Es gibt keine Interaktionseffekte zwischen Testart und Gruppe im Untertest „Synonyme Finden“.

H1₃: Es gibt Interaktionseffekte zwischen Testart und Gruppe im Untertest „Synonyme Finden“.

Bei der Berechnung der Mixed ANOVA (Tab. 10) kann ein nicht signifikanter Interaktionseffekt ($F=1,684$; $p=0,196$) angenommen werden. Der Hauptfaktor „Testart“ ($F=76,644$; $p<0,001$; $\eta^2=0,335$) hat einen signifikanten Effekt auf die Testleistung.

Tab. 10 Ergebnisse Mixed ANOVA des Untertests „Synonyme Finden“

	df	Mittel der Quadrate	F-Wert	Signifikanz	Partielles Eta ²
Testart * Gruppe	1	1,126	1,684	0,196	0,011
Gruppe	1	8,912	1,162	0,283	0,008
Testart	1	51,260	76,644	<0,001	0,335

*signifikant $p < 0,05$; df=Freiheitsgrade; Eta²= Effektstärke

Die Grafik (Abb. 15) verbildlicht den signifikanten Effekt der Testart. Beide Gruppen schneiden in der Einzeltestung wieder besser ab als in der Gruppentestung.

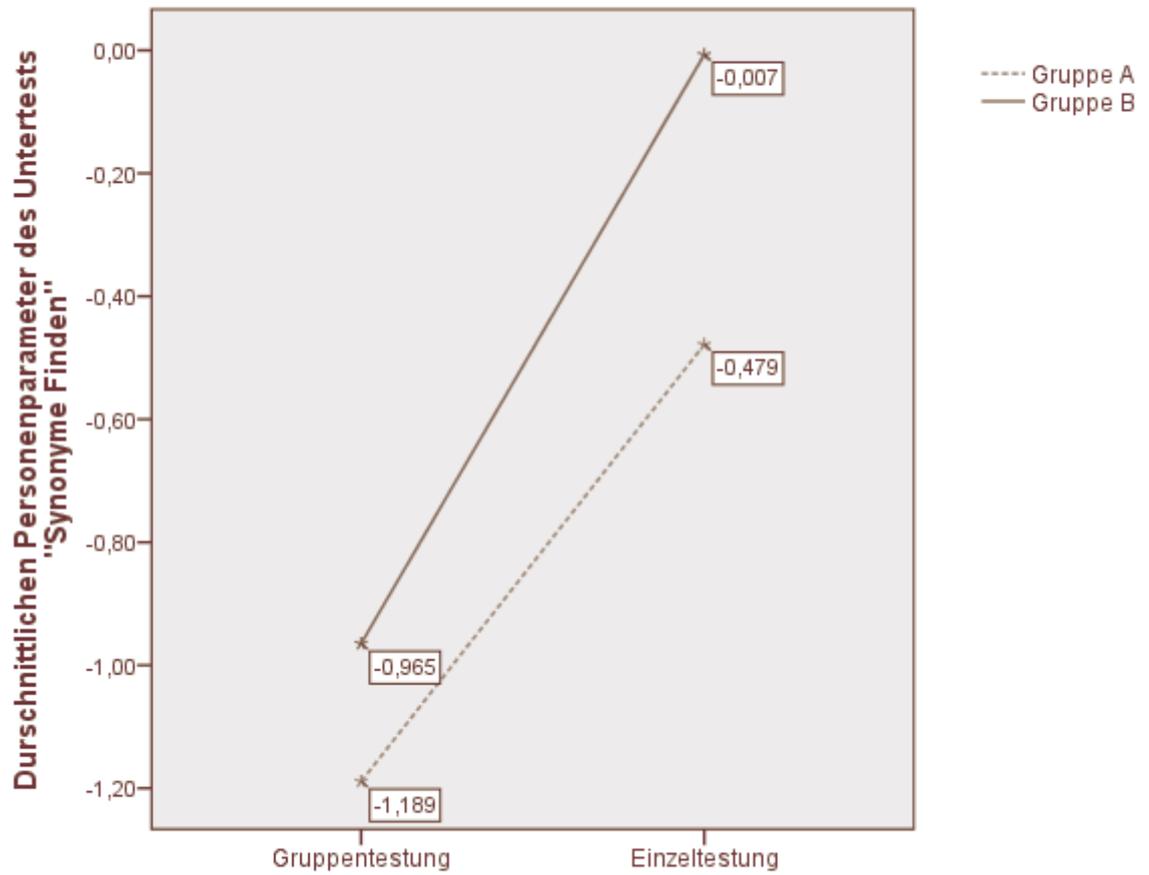


Abb. 15 Mittelwerte des Untertests „Synonyme Finden“ Testart x Gruppe

Tabelle 11 dokumentiert nochmals die Mittelwerte der beiden Faktoren und des Interaktionseffektes. Den Werten der Testart ist zu entnehmen, dass die Leistungen in den Einzeltestungen besser sind als in den Gruppentestungen.

Tab. 11 Darstellung der geschätzten Randmittel des Untertests „Synonyme Finden“

		MW der Personenparameter	SE
Gruppe x Testart	Gruppe A Einzeltestung	-0,479	0,267
	Gruppe B Gruppentestung	-1,189	0,256
	Gruppe A Gruppentestung	-0,965	0,216
	Gruppe B Einzeltestung	-0,007	0,207

MW=Mittelwert; SE=Standardfehler

Kurze Zusammenfassung der Ergebnisse

Im Untertest „Synonyme Finden“ werden H_{02} und H_{03} beibehalten. H_{01} wird, bedingt durch das signifikante Ergebnis, verworfen und H_{11} angenommen. Die Versuchspersonen erreichen in der Einzeltestung bessere Ergebnisse als in der Gruppentestung.

Ergebnisse – Untertest „Funktionen Abstrahieren“

Der Untertest „Funktionen Abstrahieren“ misst die Fähigkeit, durch Abstraktionen zu einer Begriffsbildung zu gelangen (Kubinger, 2009a).

HO₁: Die Testart hat keinen Einfluss auf die Testleistung im Untertest „Funktionen Abstrahieren“, unabhängig von der Gruppe.

H1₁: Die Testart hat einen Einfluss auf die Testleistung im Untertest „Funktionen Abstrahieren“, unabhängig von der Gruppe.

HO₂: Die Gruppe hat keinen Einfluss auf die Testleistung im Untertest „Funktionen Abstrahieren“, unabhängig von der Testart.

H1₂: Die Gruppe hat einen Einfluss auf die Testleistung im Untertest „Funktionen Abstrahieren“, unabhängig von der Testart.

HO₃: Es gibt keine Interaktionseffekte zwischen Testart und Gruppe im Untertest „Funktionen Abstrahieren“.

H1₃: Es gibt Interaktionseffekte zwischen Testart und Gruppe im Untertest „Funktionen Abstrahieren“.

Die Berechnung der Mixed ANOVA (Tab. 12) zeigt einen signifikanten Interaktionseffekt ($F=4,698$; $p=0,032$; $\text{Eta}^2=0,030$) und einen signifikanten Haupteffekt bei der Testart ($F=230,908$; $p<0,001$; $\text{Eta}^2=0,603$). Mit 60,3% liegt hier wieder ein starker Effekt der Testart vor. Dies wird durch Abbildung 16 nochmal verdeutlicht. Man erkennt wieder eindeutig, dass die Testart „Einzelsetting“ zu signifikant besseren Leistungen als die Testart „Gruppentestung“ führt.

Zusätzlich ist aber auch eine wechselseitige Beeinflussung der Variablen „Testart“ und „Gruppe“ ersichtlich, während nämlich die Leistungen im Gruppensetting nahezu gleich sind, können Testpersonen, die zuvor die Aufgaben schon einmal in der Gruppe

bearbeitet haben, diese dann im Einzelsetting noch besser bearbeiten, als jene Testpersonen, die gleich mit der Einzeltestung begonnen haben.

Tab. 12 Ergebnisse Mixed ANOVA des Untertests „Funktionen Abstrahieren“

	df	Mittel der Quadrate	F-Wert	Signifikanz	Partielles Eta ²
Testart * Gruppe	1	5,480	4,698	0,032	0,030
Gruppe	1	4,637	4,637	0,304	0,007
Testart	1	269,351	230,908	<0,001	0,603

*signifikant $p < 0,05$; df=Freiheitsgrade; Eta²= Effektstärke

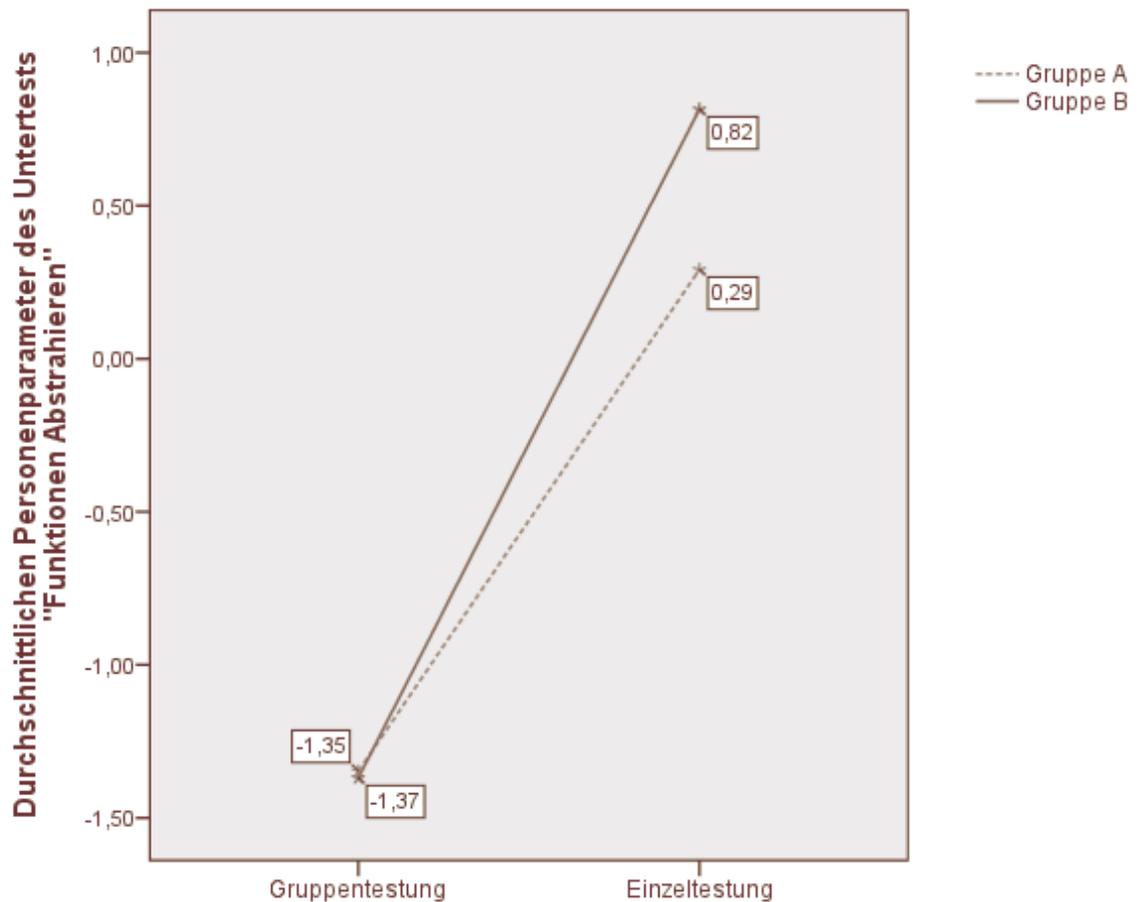


Abb. 16 Mittelwerte des Untertests „Funktionen Abstrahieren“ Testart x Gruppe

Die Angabe der Mittelwerte (Tab. 13) stellt klar, dass die Bearbeitung der Items im Einzelsetting leichter war als in der Gruppentestung. Die Interaktion besagt, dass es in der Gruppentestung keinen Unterschied in der Leistung gibt. In der Einzeltestung sind allerdings Unterschiede erkennbar. Die Personen aus Gruppe B erlangen bei der Einzeltestung bessere Ergebnisse als jene aus der Gruppe A.

Tab. 13 Darstellung der geschätzten Randmittel des Untertests „Funktionen Abstrahieren“

		MW der Personenparameter	SE
Gruppe x Testart	Gruppe A Einzeltestung	0,291	0,208
	Gruppe B Gruppentestung	-1,348	0,218
	Gruppe A Gruppentestung	-1,370	0,168
	Gruppe B Einzeltestung	0,815	0,176

MW=Mittelwert; SE=Standardfehler

Kurze Zusammenfassung der Ergebnisse

Im Untertest „Funktionen Abstrahieren“ gab es signifikante Ergebnisse im Faktor „Testart“ und in der Interaktion „Gruppe x Testart“. In Folge dieses Trends werden die beiden $H1_1$ und $H1_3$ angenommen. $H0_2$ wird aufgrund des nicht signifikanten Ergebnisses beibehalten.

$H1_1$ nimmt an, dass die Testart im Untertest „Funktionen Abstrahieren“, unabhängig von der Gruppe, einen Einfluss auf die Leistung hat. Wie bei den vorherigen Untertests sind in der Einzeltestung bessere Leistungen zu erkennen.

$H1_3$ nimmt an, dass die Testart und die Gruppe zusammen einen Einfluss auf die Leistung haben. Im Gegensatz zur Gruppentestung, sind in der Einzeltestung Unterschiede vorzufinden, nämlich, dass Testpersonen die zuerst schon in der Gruppe getestet wurden, die Aufgaben dann im Einzelsetting noch besser lösen konnten.

Ergebnisse – Untertest „ Soziales Erfassen und Sachliches Reflektieren“

Der Untertest „Soziales Erfassen und Sachliches Reflektieren“ überprüft, inwieweit das Kind über sozial angepasste Verhaltensweisen und gesellschaftliche Bedingungen informiert ist (Kubinger, 2009a).

HO₁: Die Testart hat keinen Einfluss auf die Testleistung im Untertest „Soziales Erfassen und Sachliches Reflektieren“, unabhängig von der Gruppe.

H1₁: Die Testart hat einen Einfluss auf die Testleistung im Untertest „Soziales Erfassen und Sachliches Reflektieren“, unabhängig von der Gruppe.

HO₂: Die Gruppe hat keinen Einfluss auf die Testleistung im Untertest „Soziales Erfassen und Sachliches Reflektieren“, unabhängig von der Testart.

H1₂: Die Gruppe hat einen Einfluss auf die Testleistung im Untertest „Soziales Erfassen und Sachliches Reflektieren“, unabhängig von der Testart.

HO₃: Es gibt keine Interaktionseffekte zwischen Testart und Gruppe im Untertest „Soziales Erfassen und Sachliches Reflektieren“.

H1₃: Es gibt Interaktionseffekte zwischen Testart und Gruppe im Untertest „Soziales Erfassen und Sachliches Reflektieren“.

Die Ergebnisse der Mixed ANOVA (Tab. 14) zeigen einen starken Trend beim Hauptfaktor „Testart“ ($F=229,056$; $p<0,001$, $\text{Eta}^2=0,601$). 60,1 % der Gesamtvarianz wird durch diesen erklärt. Die Interaktion „Testart x Gruppe“ ($F=0,688$; $p=0,408$) ist nicht signifikant und hat daher keinen Einfluss auf die Leistung.

Tab. 14 Ergebnisse Mixed ANOVA des Untertests „Soziales Erfassen und Sachliches Reflektieren“

	df	Mittel der Quadrate	F-Wert	Signifikanz	Partielles Eta^2
Testart * Gruppe	1	0,439	0,688	0,408	0,005
Gruppe	1	2,860	0,545	0,462	0,004
Testart	1	146,018	229,056	<0,001	0,601

*signifikant $p < 0,05$; df=Freiheitsgrade; Eta^2 = Effektstärke

In Tabelle 15 sind kaum Unterschiede in den Mittelwerten beim Faktor „Gruppe“ und der Interaktion zu erkennen. Der Effekt der Testart ist an der Streuung der Mittelwerte ersichtlich. Die Bearbeitung der Items fällt den Testpersonen in der Einzeltestung sichtlich leichter. Die Abbildung 17 illustriert wie bei den Untertests „Synonyme Finden“ und „Funktionen Abstrahieren“, dass die Leistungen in der Gruppe B generell besser sind als in Gruppe A.

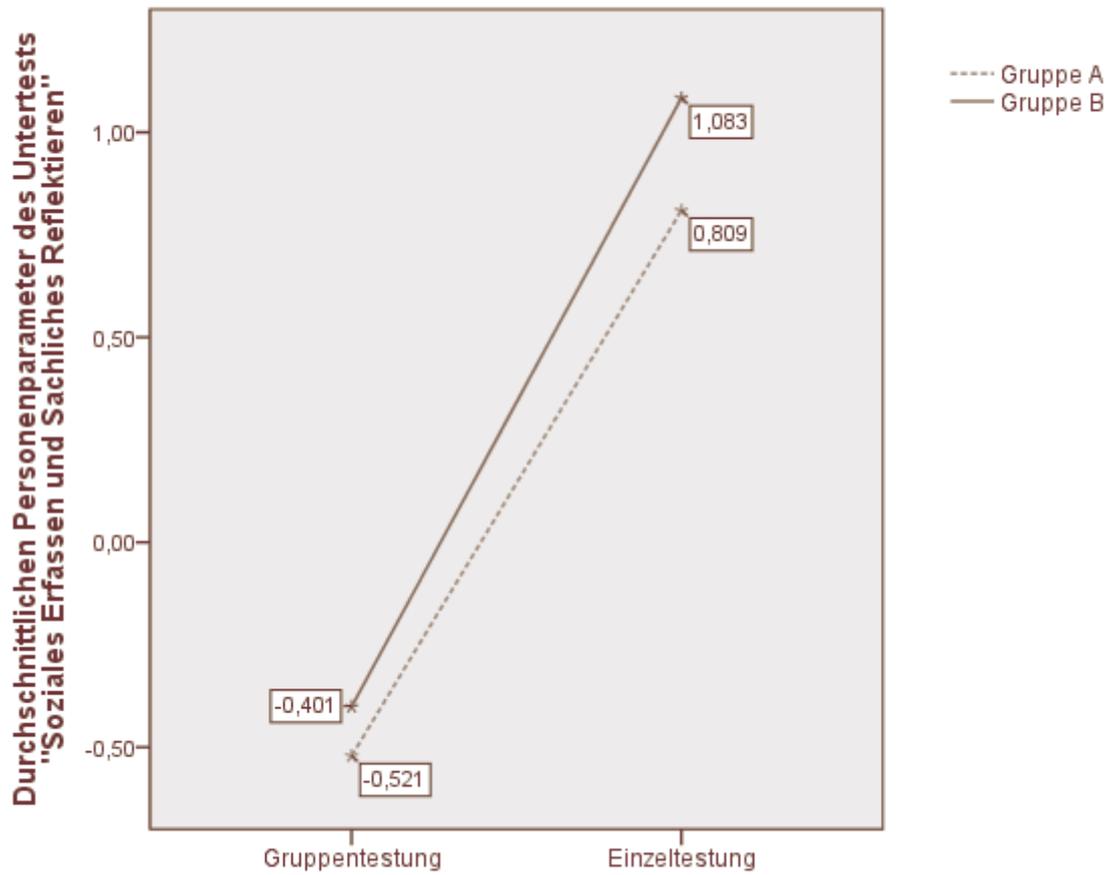


Abb. 17 Mittelwerte des Untertests „Soziales Erfassen und Sachliches Reflektieren“
Testart x Gruppe

Tab. 15 Darstellung der geschätzten Randmittel des Untertests „Soziales Erfassen und Sachliches Reflektieren“

		MW der Personenparameter	SE
Gruppe x Testart	Gruppe A Einzeltestung	0,809	0,214
	Gruppe B Gruppentestung	-0,521	0,225
	Gruppe A Gruppentestung	-0,401	0,174
	Gruppe B Einzeltestung	1,083	0,182

MW=Mittelwert; SE=Standardfehler

Kurze Zusammenfassung der Ergebnisse

Im Untertest „Soziales Erfassen und Sachliches Reflektieren“ können die beiden H_{02} und H_{03} beibehalten werden. H_{01} muss aufgrund des signifikanten Ergebnisses verworfen werden und somit wird H_{11} angenommen.

H_{11} nimmt an, dass die Testart im Untertest „Soziales Erfassen und Sachliches Reflektieren“, unabhängig von der Gruppe, einen Einfluss auf die Leistung hat. Wie bei den vorherigen Untertests sind in der Einzeltestung bessere Leistungen zu erkennen.

7 Interpretation der Ergebnisse, Diskussion und Ausblick

7.1 Interpretation der Ergebnisse

Zusammenfassend kann die Forschungsfrage dahingehend beantwortet werden, dass es Unterschiede, zwischen Gruppentestung und Einzeltestung, hinsichtlich der Testleistungen der Testpersonen gibt.

Bei den Untertests „Alltagswissen“ und „Angewandtes Rechnen“ war Gruppe A besser als Gruppe B. Bei den Untertests „Synonyme Finden“, „Funktionen Abstrahieren“ und „Soziales Erfassen und Sachliches Reflektieren“ erzielte Gruppe B bessere Leistungen als Gruppe A. Generell gilt aber, egal welche Testung zuerst stattgefunden hat, Gruppentestung oder Einzeltestung, alle Versuchspersonen erzielten in der Einzelvorgabe signifikant bessere Leistungen als in der Gruppenvorgabe.

Insgesamt nahmen 154 Schüler an der Untersuchung teil. Das Geschlecht war annähernd gleich verteilt. Das Alter lag in Gruppe A bei durchschnittlich 11,73 Jahren. Bei Gruppe B waren es im Durchschnitt 11,12 Jahre und somit war das Alter ebenfalls gleich verteilt.

Der Schultyp war in den beiden Gruppen ungleich verteilt und könnte somit einen Einfluss auf das Ergebnis haben, besonders auf die besseren Leistungen der Gruppe B in den drei Untertests „Synonyme Finden“, „Funktionen Abstrahieren“ und „Soziales Erfassen und Sachliches Reflektieren“. Gruppe A war bei den VolksschülerInnen unterrepräsentiert. Von 65 teilnehmenden VolksschülerInnen, waren nur 19 SchülerInnen der Gruppe A zugeteilt.

Die Stichprobenszusammensetzung kann aber die deutlichen und über beide Gruppen sowie alle Untertests bestehenden signifikanten Leistungsunterschiede zwischen Einzel- und Gruppentestung nicht erklären. Da das Design ausbalanciert war, können auch Ermüdungserscheinungen ausgeschlossen werden. Auch gravierende Störungen während der Gruppentestung können, wie in der Verhaltensbeobachtung beschrieben, ausgeschlossen werden.

Wie aus der Theorie abzuleiten ist und sich auch in der Verhaltensbeobachtung deutlich zeigte, verstehen lernschwache Kinder die Instruktionen nicht oder sind

langsamer beim Lesen der Instruktion. Desweiteren kommen sie beim Bearbeiten der Aufgaben unter Zeitdruck und geben auf. Somit sind solche Kinder im Einzelsetting besser aufgehoben bzw. kann im Einzelverfahren auf die individuellen Bedürfnisse der Kinder eingegangen werden.

7.2 Diskussion

Untertest „Alltagswissen“

Im Untertest „Alltagswissen“ mussten die Kinder Fragen zu alltäglichen Sachverhalten beantworten. In der Gruppentestung diente eine Zeile, um die Antwort zu notieren. Das freie Antwortformat stellte bei diesem Untertest kein Problem für die Auswertung dar, wie es von Kubinger (2009b) und Seiwald (2003) in Kap.3 angeführt wird. Aufgrund der eindeutig richtigen Antworten war die Zuordnung zu „gelöst“ oder „nicht gelöst“ einfach.

Untertest „Angewandtes Rechnen“

Im Untertest „Angewandtes Rechnen“ wurden den Versuchspersonen Rechnungen vorgelesen, welche zu lösen waren. In der Gruppentestung wurden die Schüler und Schülerinnen, nur in der Instruktion, darauf hingewiesen, die richtige Lösung in das linke Kästchen einzutragen. Nur dann wird das korrekte Ergebnis als richtig gewertet.

Auffällig scheinen bei diesem Untertest die Unterschiede in der Bedingung „Gruppe“. Dieser Effekt konnte nur in diesem Untertest bestätigt werden. Hier schnitten die Testpersonen in Gruppe A viel besser ab als jene in Gruppe B. Argumente für diesen Effekt, könnten daher rühren, dass die Kinder in der Individualtestung dazu animiert werden, sich die Aufgaben bei falschen Antworten nochmals durchzulesen. Die Antworten beinhalten meistens nur eine Zahl die leicht gemerkt werden kann und bei der Gruppentestung schnell abrufbar ist.

Testpersonen aus der Gruppe B müssen sich bei der ersten Testung die Angaben selbst durchlesen und werden bei Flüchtigkeitsfehlern, die durch zu schnelles Rechnen und unaufmerksames Durchlesen entstehen können, nicht dazu motiviert, sich die Aufgaben nochmals durchzusehen. Das Rechenergebnis der Gruppentestung übernehmen sie dann in der Einzeltestung. Der Untertest „Angewandtes Rechnen“ unterlag in der Gruppentestung einem wichtigen Kriterium für ein gelöstes Item. Die Lösung musste in einem vorgedruckten Kästchen eingetragen werden. Stand die richtige Lösung daneben, war das Item trotzdem „nicht gelöst“. Bei einigen Testpersonen musste das richtige Ergebnis als „nicht gelöst“ gewertet werden, weil es nicht korrekt in das vorgesehene Kästchen eingetragen wurde. Die Schüler und

Schülerinnen wurden nur durch die Instruktion auf das Kästchen hingewiesen. Bei einigen Gruppentestungen stand das richtige Ergebnis neben der Rechnung. Vermutlich haben sie beim Rechnen vergessen, dass die Rechnung nur als richtig gewertet werden kann, wenn das Ergebnis im Kästchen steht.

Mögliche Erklärungen für bessere Ergebnisse der Gruppe A in den Untertests „Alltagswissen“ und „Angewandtes Rechnen“ könnten darauf zurückzuführen sein, dass die Versuchspersonen in der Einzeltestung darauf aufmerksam gemacht worden sind, sich die Angabe erneut durchzulesen. Das Ergebnis der Einzeltestung wurde ebenfalls in der Gruppentestung verwendet. Vor allem bei den Rechenaufgaben setzten die meisten Schüler und Schülerinnen das Ergebnis sofort ein, ohne eine Rechnung zu machen.

Desweiteren ist zu erwähnen, dass der Untertest „Alltagswissen“ und „Angewandtes Rechnen“ in ihrer Aufgabenstellung leicht verständlich waren. Ebenso bietet die Auswertung dem Testleiter wenig Freiraum bei der Interpretation der Ergebnisse. Bei beiden Untertests kann die Lösung nur als richtig oder falsch gewertet werden. Unter anderen Bedingungen, wie etwa eine längere Zeitvorgabe bei jüngeren Schülern, wäre es durchaus möglich, diese Tests als Gruppentestung durchzuführen.

Untertest „Synonyme Finden“

Im Untertest „Synonyme Finden“ müssen die Versuchspersonen für ein vorgegebenes Wort ein anderes Wort finden, welches die gleiche Bedeutung hat. Dieser Untertest ist durch eine schwierigere Aufgabenstellung, im Gegensatz zu den beiden vorherigen Untertests, gekennzeichnet. Die Beispielaufgabe in der Instruktion ist, im Vergleich zu den Items, sehr leicht gestaltet. Leistungsschwache Kinder haben in diesem Untertest Probleme, die Aufgaben korrekt zu beantworten. Bei den Gruppentestungen kam es vor, dass Schüler kaum Aufgaben richtig lösten, da sie die Aufgabenstellung nicht verstanden.

Untertest „Funktionen Abstrahieren“

Dieser Untertest ist eher schwieriger gestaltet. Es sollen die Gemeinsamkeiten von zwei vorgegebenen Begriffen gefunden werden. Auch hier zeigen sich die gleichen Probleme wie beim Untertest „Synonyme Finden“. Leistungsschwache Schüler waren durch die Vorgabe von mittelschweren Items im Nachteil. Einige Schüler, darunter auch jene mit Migrationshintergrund, hatten Schwierigkeiten, die Instruktion zu verstehen. In diesem Untertest kann ein signifikanter Interaktionseffekt beobachtet werden. Es zeigen sich Leistungsunterschiede in der Einzeltestung. Gruppe B ist hier besser als Gruppe A.

Ein Vorteil für die Auswertung sind genau definierte Gemeinsamkeiten, welche eine einfache Auswertung ermöglichen.

Untertest „Soziales Erfassen und Sachliches Reflektieren“

Beim letzten Untertest „Soziales Erfassen und Sachliches Reflektieren“ wurden den Kindern Fragen über alltägliche Sachverhalte gestellt. Wie in Kapitel 3 (Bühner, 2011) beschrieben, hat der Testleiter hier viel Interpretationsfreiraum zur Verfügung. Bei der Einzeltestung ist die Auswertung noch einfacher gestaltet, da der Testleiter das Kind genauer Befragen kann, wie das denn jetzt gemeint sei. In der Gruppentestung ist dies nicht der Fall. Bei den alltäglichen Sachverhalten spielt auch die familiäre Beziehung eine wichtige Rolle. Besonders bei Frage 2 „Was kannst du tun, wenn deine Mutter, beim Vorbereiten für das Geburtstagsfest, unter Zeitdruck kommt?“ Für manche Kinder ist es wichtig, der Mutter zu helfen, andere wiederum kümmern sich um die Geschwister, so dass die Mutter mehr Zeit hat. Hier kommt es eventuell bei beiden Testarten zu Interpretationsschwierigkeiten. In der Einzeltestung könnte der Testleiter genauer nachfragen.

Für die Versuchspersonen war dieser Untertest einfacher gestaltet. Die Instruktion bereitete kaum Verständnisschwierigkeiten. Aufgrund des freien Antwortformates, kombiniert mit der Speedkomponente, konnten viele Kinder nur wenige Items bearbeiten, da sie sich zu lange mit einer Aufgabe beschäftigten. Dies galt besonders für die jüngeren Kindern, die durch die langsamere Schreibweise gehandicapt waren, aber auch für jene Schüler, die besonders viel zu einer Aufgabe schreiben wollten. Wie schon im Literaturteil angeführt (vgl. Kap. 3, Kubinger, 2009b), können bei Speed-and-

Powertests schlechte Leistungen auf den Mangel an Schnelligkeit oder auf den Mangel an Fähigkeit zurückzuführen sein.

Eine Begründung für die besseren Ergebnisse der Gruppe B in den Untertests „Synonyme Finden“, „Funktionen Abstrahieren“ und „Soziales Erfassen und Sachliches Reflektieren“ könnten die schwierigeren Aufgabenstellungen sein. Möglicherweise haben Versuchspersonen der Gruppe A in der Einzeltestung Probleme, die gesprochenen Instruktionen und Aufgaben, durch den Testleiter, zu verstehen. Kürschner und Schnotz (2008) führen an, dass schwierigere Texte leichter zu verstehen sind, wenn diese geschrieben dargestellt werden. Versuchspersonen der Gruppe B kennen die Texte in der Einzeltestung bereits, da sie diese in der Gruppentestung durchgelesen haben.

In der Gruppentestung kann durchaus die Möglichkeit eingeräumt werden, dass einige Testpersonen alleine an der Instruktion gescheitert sind. Beobachtungen während den Gruppentestungen zeigten, dass ältere Testpersonen viel schneller im Lesen der Instruktion und Bearbeiten der Items waren als jüngere Testpersonen. Die beiden Autoren (Amelang & Schmidt-Atzert, 2006) führten auch den Nachteil der schwächeren Lesefertigkeit bei Gruppentestungen an. Es waren die jüngeren Testpersonen die sich länger als die anderen Personen mit dem Durchlesen der Instruktion beschäftigten. Leistungsschwache Kinder würden unter solchen Bedingungen eher demotiviert und enttäuscht werden. Es soll noch angemerkt werden, dass bei diesen Gruppentestungen ein hoher Zeitfaktor für die Auswertung mit einzubeziehen ist. Der Testleiter muss erstens überlegen, wie die Testperson zu dieser Antwort gelangt ist und wie sie zu verstehen ist, zweitens gibt es Schüler, die eine unleserliche Schrift haben. Kann die Antwort nicht gelesen werden, muss sie als „nicht gelöst“ gewertet werden. Die Verrechnungssicherheit bei freiem Antwortformat wäre in diesem Fall nicht gegeben. Im Einzelsetting kann der Psychologe, durch Verhaltensbeobachtung und direkte Kommunikation, das zu testende Kind sehr gut kennenlernen. Dem Erziehungsberechtigten können viel mehr Informationen übermittelt werden, was umgekehrt dem Erziehungsberechtigten einen kompetenten und vertrauensvollen Eindruck des Psychologen vermittelt.

An der Untersuchung nahmen nicht viele Schüler mit Migrationshintergrund teil, ihnen bereiteten das Lesen der Instruktion, sowie die Bearbeitung der Aufgaben in der Gruppentestung Schwierigkeiten. Eine Studie von Tiedemann und Billmann-Mahecha (2007) bestätigt die verminderte Leseverständnisleistung bei Migrationskindern.

Hinzu kommt noch, dass bei dieser Untersuchung eine adaptive Testung außer Acht gelassen werden musste. Es wurden allen Testpersonen die gleichen, ihrer Altersgruppe entsprechenden, mittelschweren Items vorgegeben. Schwache Schüler waren hier deutlich im Nachteil, was ebenfalls einen Einfluss auf die Ergebnisse ausübt.

Der Ansatz von Amelang und Schmidt-Atzert (2006) erwies sich als bestätigt, der besagt, dass Testpersonen unter der Bedingung einer Gruppentestung dazu neigen, beim Nachbar abzuschreiben. Dies traf bei den älteren Testpersonen zu.

Der angeführte Vorteil des individuellen Eingehens des Testleiters auf die Testperson im Individualverfahren von Herle (2003) spielt eine wesentliche Rolle. Die Unterschiede zwischen der Gruppentestung und der Einzeltestung können auf diesen Vorteil zurückzuführen sein. Die Schüler besprechen im Einzelsetting gemeinsam mit dem Testleiter die Instruktion und der Testleiter hat bei diesem Setting einen besseren Überblick, ob die Instruktion verstanden wurde oder nicht. Im Einzelverfahren werden die Testpersonen dazu angeregt, die Antwort nochmals genauer zu beschreiben bzw. werden sie gefragt, wie denn das jetzt gemeint ist. All dieses Eingehen auf die Testperson wird in der Gruppentestung unterlassen. Die Schüler sind hier quasi auf sich alleingestellt.

7.3 Ausblick

Die aufgezeigten Befunde decken sich mit der Literatur und weiteren Forschungsergebnissen (vgl. Reindl, 1998; Böck, 2010) und sprechen dagegen, dass man, bei den fürs Einzelsetting konzipierten Aufgaben des AID2, auf Normen von Gruppentestungen zurückgreifen könne.

Weitere Untersuchungen, die diese Ergebnisse nochmals bestätigen, wären für die psychologische Diagnostik von hoher Relevanz.

8 Zusammenfassung

Das AID 2.2 ist eines der am meisten eingesetzten Verfahren in der psychologischen Diagnostik (Kubinger, 2009a). Vor allem im schulpyschologischen Kontext findet das Adaptive Intelligenz Diagnostikum große Anwendung (Kubinger, 2009b). Die adaptive Vorgabe verhindert bei Kindern Frustration und führt zu einem Motivationsaufschub (Preusche & Leiss, 2003).

In der vorliegenden Arbeit wurden fünf Untertests aus der aktualisierten Version des Adaptiven Intelligenz Diagnostikums AID 3, welche sprachliche und rechnerische Fertigkeiten erfassen, sowohl in der Gruppe als auch im Einzelsetting vorgegeben. Für die Gruppentestung wurden, unter Berücksichtigung der drei unterschiedlichen Altersgruppen, Testhefte gestaltet. Zu Beginn der Testung wurden die Testpersonen durch den Testleiter kurz instruiert, danach erfolgte vor jedem Untertest eine schriftliche Instruktion. Die Bearbeitung jedes Untertests erfolgte unter einer bestimmten Zeitangabe. Die Testpersonen mussten die fünf Untertests einmal in der Gruppe bearbeiten und einmal im Einzelsetting, wozu die Hälfte der Stichprobe zuerst im Einzelsetting und dann im Gruppensetting und die andere Hälfte umgekehrt, dieselben Aufgaben bearbeiteten. Gemessen wurde, ob es Unterschiede zwischen diesen beiden Testarten hinsichtlich der Testleistungen gibt. Im Mittelpunkt des Interesses stand die Schwierigkeit der Items für die Testpersonen unter diesen beiden Bedingungen. 154 Volksschüler und Gymnasiasten nahmen an der Untersuchung teil. 61 Schüler wurden der Gruppe A zugeordnet. Sie bearbeiteten zuerst die Untertests im Einzelsetting und zwei bis drei Tage später in der Gruppe, welche meistens aus vier bis sechs Schülern bestand. 93 Schüler wurden Gruppe B zugeteilt. Diese mussten sich zuerst der Gruppentestung unterziehen und nach Einhaltung desselben Abstandes folgte das Einzelverfahren.

In allen Untertests gab es einen Effekt beim Hauptfaktor „Testart“. Die Testpersonen schnitten in den Einzeltestungen immer besser ab als in den Gruppentestungen. Desweiteren konnte ein Effekt im Untertest „Angewandtes Rechnen“ hinsichtlich des Faktors „Gruppe“, unabhängig von der Testart, gefunden werden. Hier konnten Unterschiede zwischen Gruppe A und Gruppe B beobachtet werden. Ein zusätzlicher Effekt zeigte sich im Untertest „Funktionen Abstrahieren“, betreffend der Interaktion „Testart x Gruppe“. Der Leistungsunterschied bei Gruppe A und B ist im Einzelsetting anders als in der Gruppentestung.

Worauf der signifikante und oftmals hohe Effekt der Testart zurückzuführen ist, kann aus der Theorie und Verhaltensbeobachtung abgeleitet werden. So sind besonders Instruktionsverständnisse, Zeitdruck, Probleme beim Leseverständnis und Motivationsprobleme theoriekonform als Ursache zu vermuten. Da die vorliegende Studie nur ein kleines Datensample erfasst, sind weitere Forschungen zu diesem Thema von hoher Relevanz.

9 Literaturverzeichnis

Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* (4. Auflage). Heidelberg: Springer.

Asendorpf, J.B. (2007). *Psychologie der Persönlichkeit*. (4. Auflage). Heidelberg: Springer Medizin.

Böck, J. (2010). *AID 2 als Gruppentestung? Eine Machbarkeitsstudie*. Unveröffentlichte Diplomarbeit, Universität Wien.

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. (6. Auflage). Heidelberg: Springer Medizin.

Brocke, B. & Beauducel, A. (2001). Intelligenz als Konstrukt. In Stern, E. & Guthke, J. (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 13-43). Lengerich: Pabst Science Publishers.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Auflage). München: Pearson Studium.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1-22.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Hillsdale, NJ: Erlbaum.

Eichhorn, T. (2011). *Systematische psychologisch-diagnostische Gesprächsführung und Verhaltensbeobachtung zur Erfassung leistungsrelevanter Arbeitshaltungen*. Unveröffentlichte Diplomarbeit, Universität Wien.

Floyd, R.G., Keith, T.Z., Taub, G.E. & McGrew, K.S. (2007). Cattell-Horn-Carroll Cognitive Abilities and Their Effects on Reading Decoding Skills: Has Indirect Effects, More Specific Abilities Than Direct Effects. *School Psychology Quarterly*, 22, 200-233.

Funke, J. & Vaterrodt-Plünnecke, B. (1998). *Was ist Intelligenz*. München: Beck.

Gardner, H. (1991). *Abschied vom IQ: Die Rahmentheorie der vielfachen Intelligenzen*. Stuttgart: Klett-Cotta.

- Guthke, J. (1996). *Intelligenz im Test. Wege der psychologischen Intelligenzdiagnostik*. Göttingen: Vandenhoeck & Ruprecht.
- Häcker, H.O. & Stapf, K-H. (2004). *Dorsch Psychologisches Wörterbuch* (14. Auflage). Bern: Huber.
- Haseloff, O.W. (1970). *Kleines Lehrbuch der Statistik*. Berlin: de Gruyter.
- Herle, M. (2003). Individualtestung. In K.D. Kubinger & R.S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 205-207). Weinheim: Beltz.
- Hofstätter, P.R. (1957). *Psychologie*. Frankfurt am Main: Fischer.
- Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik*. (Kompendien, Psychologische Diagnostik, Bd.6). Göttingen: Hogrefe.
- Horn, R. (1993). *Stichwort Intelligenz*. München: Heyne.
- Hornk, L.F. (2006). Testökonomie. In F. Petermann & M. Eid (Hrsg.). *Handbuch der Psychologischen Diagnostik* (S. 434-446). Band 4. Göttingen: Hogrefe.
- Jacobs, C., Heubrock, D. & Petermann, F. (2003). Testinformationen. *Diagnostica*, 49, 184–194.
- Jonkisz, E. & Moosbrugger, H. (2007). Planung und Entwicklung von psychologischen Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.). *Testtheorie und Fragebogenkonstruktion* (S. 27-72). Heidelberg: Springer.
- Kelava, A. & Moosbrugger, H. (2007). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.). *Testtheorie und Fragebogenkonstruktion* (S. 73-98). Heidelberg: Springer.
- Krohne H.W. & Hock, M. (2007). *Psychologische Diagnostik: Grundlagen und Anwendungsfelder*. Stuttgart: Kohlhammer.
- Kubinger, K.D. & Teichmann H. (1997). *Psychologische Diagnostik und Intervention in Fallbeispielen*. Weinheim: Beltz.
- Kubinger, K.D. & Wurst, E. (2000). *Adaptives Intelligenzdiagnostikum 2*. Beltz: Göttingen.

- Kubinger, K.D. (1989). *Moderne Testtheorie. Ein Abriss samt neuesten Beiträgen* (2. Auflage). Weinheim: Beltz.
- Kubinger, K.D. (2009a). *Adaptives Intelligenzdiagnostikum – Version 2.2 (AID 2)*. Göttingen: Beltz.
- Kubinger, K.D. (2009b). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens* (2. Auflage). Göttingen: Hogrefe.
- Kürschner, Ch. & Schnotz, W. (2008). Das Verhältnis gesprochener und geschriebener Sprache bei der Konstruktion mentaler Repräsentationen. *Psychologische Rundschau*, 59 (3), 139-149.
- Lamberti, G. (2006). Zur Entwicklung der Psychometrie nach James McKeen Cattell (1860-1944). In Lamberti, G. (Hrsg.), *Intelligenz auf dem Prüfstand-100 Jahre Psychometrie* (S. 9-22). Göttingen: Vandenhoeck & Ruprecht.
- Lienert, G.A. & Ratz, U. (1994). *Testaufbau und Testanalyse* (5. Auflage). Weinheim: Beltz.
- Myers, D.G. (2004). *Psychologie*. Heidelberg: Springer.
- Neubauer, A.C. (2005). Intelligenz. In H. Weber & T. Rammsayer (Hrsg.), *Handbuch der Persönlichkeitspsychologie und Differentiellen Psychologie* (S. 321-333). Göttingen: Hogrefe.
- Oerter, R. & Montada, L. (2008). *Entwicklungspsychologie*. (6. Auflage). Weinheim: Beltz.
- Phelps, L., McGrew, K.S., Knopik, S.N. & Ford, L. (2005). The General (g), Broad, and Narrow CHC Stratum Characteristics of the WJ III and WISC-III Tests: A Confirmatory Cross-Battery Investigation. *School Psychology Quarterly*, 29, 66-88.
- Pospeschill, M. & Spinath, F.M. (2009). *Psychologische Diagnostik*. München: Ernst Reinhardt.
- Preusche, I. & Leiss, U. (2003). Intelligenztests für Kinder. *Report Psychologie*, 28, 12-26.
- Rasch, D. & Kubinger, K.D. (2006). *Statistik für das Psychologiestudium*. Heidelberg: Spektrum-Akademischer Verlag.

- Reindl, A. (1998). *Einzel- vs. Gruppenvorgabe. Niveauunterschiede im Leistungsbereich*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Rost, D.H. (2009). *Intelligenz. Mythen und Fakten*. Weinheim: Beltz.
- Schweizer, K. (2005). *Leistung und Leistungsdiagnostik*. Heidelberg: Springer Medizin Verlag.
- Seiwald, B.B. (2003). Antwortformat. In K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 23-28). Weinheim: Beltz.
- Spearman, C. (1927). *The abilities of man*. London: Macmillan.
- Stemmler, G., Hagemann, D., Amelang, M. & Bartussek, D. (2011). *Differentielle Psychologie und Persönlichkeitsforschung* (7. Auflage). Stuttgart: Kohlhammer.
- Süß, M. (2003). Intelligenztheorien. In K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 217-224). Weinheim: Beltz.
- Taub, G.E., Floyd, R.G., Keith, T.Z. & McGrew, K.S. (2008). Effects of General and Broad Cognitives Abilities on Mathematics Achievement. *School Psychology Quarterly*, 23, 187-198.
- Tiedemann, J. & Billmann-Mahecha, E. (2007). Leseverständnis, Familiensprache und Freizeitsprache. Ergebnisse aus der Hannoverschen Grundschulstudie. *Zeitschrift für Pädagogische Psychologie*, 21, (1), 41-49.
- Vock, M. & Holling, H. (2006). Intelligenzdiagnostik. In F. Petermann & M. Eid (Hrsg.). *Handbuch der Psychologischen Diagnostik* (S. 494-501). Band 4. Göttingen: Hogrefe.
- Wagener, D. (2003). Leistungstest. In K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 264-267). Weinheim: Beltz.
- Wagner-Menghin, M.M. (2003). Computerdiagnostik. In Kubinger, K.D. Jäger, R.S. (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 68-82). Weinheim: Beltz.
- Wechsler, D. (1956). *Die Messung der Intelligenz Erwachsener*. Bern: Huber
- Woike, J.K. (2003). Testökonomie. In K.D. Kubinger & R.S. Jäger (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 405-407). Weinheim: Beltz.

Zimbardo, P.G. & Gerrig, R.J. (2004). *Psychologie* (16. Auflage). München: Pearson Studium.

10 Abbildungs- und Tabellenverzeichnis

Abbildungen

Abb. 1 Lösungswahrscheinlichkeit einer Binet-Aufgabe. 50% der Kinder im Alter von 7 Jahren lösen die Aufgabe richtig (Stemmler, Hagemann, Amelang und Bartussek, 2011, S.30)	6
Abb. 2 Einige Aufgaben aus dem Binet Test (Stemmler, Hagemann, Amelang und Bartussek, 2011, S.31) 6	
Abb. 3 Zwei-Faktoren-Theorie nach Spearman (nach Hofstätter, 1957, modifiziert).....	10
Abb. 4 Multiple-Faktoren-Theorie nach Thurstone (Hofstätter, 1957, modifiziert).....	12
Abb. 5 Schematische Darstellung der oberen und mittleren Ebenen von Carrolls Drei-Schichten-Modell (Schweizer, 2006, zitiert nach Carroll, 1993)	16
Abb. 6 Grafische Darstellung der Versuchsbedingungen.....	30
Abb. 7 Testvorgabe des Untertests Soziale und Sachliche Folgerichtigkeit und Analysieren und Synthetisieren-abstrakt (Kubinger & Wurst, 2000).....	36
Abb. 8 Testinstruktion „Alltagswissen“	39
Abb. 9 Testinstruktion „Angewandtes Rechnen“	39
Abb. 10 Testinstruktion des Untertests „Synonyme Finden“	39
Abb. 11 Testinstruktion „Funktionen Abstrahieren“	40
Abb. 12 Testinstruktion „Soziales Erfassen und Sachliches Reflektieren“.....	40
Abb. 13 Mittelwerte des Untertests „Alltagwissen“ Testart x Gruppe	54
Abb. 14 Mittelwerte des Untertests „Angewandtes Rechnen“ Testart x Gruppe	58
Abb. 15 Mittelwerte des Untertests „Synonyme Finden“ Testart x Gruppe	62
Abb. 16 Mittelwerte des Untertests „Funktionen Abstrahieren“ Testart x Gruppe	66
Abb. 17 Mittelwerte des Untertests „Soziales Erfassen und Sachliches Reflektieren“ Testart x Gruppe... 70	

Tabellen

Tab. 1 Verteilung des Geschlechts	45
Tab. 2 Verteilung des Schultyps	46
Tab. 3 Verteilung der Muttersprache	47
Tab. 4 Ergebnisse der Zwischensubjekteffekte der multivarianten Varianzanalyse	49
Tab. 5 Beschreibung der Mittelwerte in den Gruppen	51
Tab. 6 Ergebnisse Mixed ANOVA des Untertests „Alltagswissen“	53
Tab. 7 Darstellung der geschätzten Randmittel des Untertests „Alltagswissen“	55
Tab. 8 Ergebnisse Mixed ANOVA des Untertests „Angewandtes Rechnen“	57
Tab. 9 Darstellung der geschätzten Randmittel des Untertests „Angewandtes Rechnen“	59
Tab. 10 Ergebnisse Mixed ANOVA des Untertests „Synonyme Finden“	61
Tab. 11 Darstellung der geschätzten Randmittel des Untertests „Synonyme Finden“	63
Tab. 12 Ergebnisse Mixed ANOVA des Untertests „Funktionen Abstrahieren“	65
Tab. 13 Darstellung der geschätzten Randmittel des Untertests „Funktionen Abstrahieren“	67
Tab. 14 Ergebnisse Mixed ANOVA des Untertests „Soziales Erfassen und Sachliches Reflektieren“	69
Tab. 15 Darstellung der geschätzten Randmittel des Untertests „Soziales Erfassen und Sachliches Reflektieren“	71

11 Anhang

11.1 Elterninformationsbrief und Einverständniserklärung

Liebe Eltern!

Als ehemalige Absolventin dieser Schule ersuche ich Sie im Rahmen meines Psychologiestudiums um Mithilfe bei meiner Diplomarbeit. Bei meiner Arbeit wird der AID-2 Original Ihrem Kind einmal einzeln vorgegeben und ein zweites Mal in der Gruppe. Der AID-2 ist ein Adaptives Intelligenzdiagnostikum, der in der Schulpsychologie vorgegeben wird und zur Überprüfung von Teilleistungsstörungen dient. Er ist ein Einzelverfahren, was bedeutet, dass der Test jedem Kind einzeln vorgegeben wird. In meiner Diplomarbeit geht es darum, den Test auch in der Gruppe vorzugeben. Es wird nicht der gesamte AID-2 vorgegeben.

Folgende Untertests werden zum Einsatz kommen:

Untertest A: „Alltagswissen“: Über mündlich gestellte Fragen, soll die Fähigkeit überprüft werden, sich Sachkenntnisse und Inhalte anzueignen, die aus der heutigen Gesellschaft alltäglich sind.

Untertest E: „Angewandtes Rechnen“: Mittels alltäglicher Rechenbeispiele sollen numerisch einfach gehaltene Aufgaben gelöst werden. Hier wird die Fähigkeit zur Anwendung von Rechenoperationen untersucht.

Untertest F: „Synonyme finden“: Für einen vorgegeben Begriff soll ein anderes Wort mit derselben Bedeutung gefunden werden. In diesem Untertest wird der Wortschatz überprüft.

Untertest H: „Funktionen Abstrahieren“: Das Kind soll die gemeinsame Funktion von jeweils zwei genannten Objekten erkennen und beschreiben. Hier handelt es sich um die Fähigkeit, durch Abstraktionen zu einer Begriffsbildung zu gelangen.

Untertest I: „Soziales Erfassen und Sachliches Reflektieren“: Durch mündliche Fragen soll überprüft werden, inwieweit das Kind über gesellschaftliche Zusammenhänge und soziale Verhaltensweisen Bescheid weiß.

Dauer etwa 40 Minuten pro Test. Die Testung erfolgt an zwei Vormittagen in Absprache der Lehrer.

Die Anonymität über die Daten und Ergebnisse Ihres Kindes ist streng gewährleistet.

Wenn Sie mich bei meiner Diplomarbeit unterstützen wollen, würde ich Sie bitten, die Einverständniserklärung zu unterschreiben und Ihrem Kind in die Schule mitzugeben.

Ich möchte mich hiermit schon im Voraus bedanken und verbleibe

Mit freundlichen Grüßen

Martina Scheffl

Einverständniserklärung

Vorname des Kindes:

Name des Erziehungsberechtigten:

Tel.Nr* :..... oder E-Mail* :.....

Ich bin einverstanden, dass Sie meinem Kind den AID-2 Original und die Gruppenversion vorgeben.

Unterschrift.....

Ich hätte Interesse an der Rückmeldung der Ergebnisse ja nein

*Bitte angeben, wenn sie eine Rückmeldung der Ergebnisse erwünschen.

11. 2 Gesprächsleitfaden

Hallo, mein Name ist Martina Scheffl. Ich studiere Psychologie an der Universität Wien. Dort wurde ein neuer Intelligenztest entwickelt.

Eure Eltern haben mir die Erlaubnis gegeben, dass ihr bei diesem Test mitmachen dürft.

Jeder von euch bekommt jetzt ein Testheft von mir ausgeteilt. Ihr lasst das Heft vor euch liegen, bis ich euch sage, was zu tun ist. Wenn ihr vorblättert muss die Person den Test abbrechen und den Raum verlassen.

Testhefte werden ausgeteilt. Ein Mustertestheft in die Höhe halten und den Kindern zeigen, was sie ausfüllen müssen.

Zuerst füllt ihr die erste Seite aus. Ihr tragt euren Namen, euer Geburtsdatum und eure Muttersprache ein. Füllt es bitte jetzt aus.

Warten bis alle fertig sind.

Ich möchte euch jetzt noch erklären, dass der Test in fünf Teile gegliedert ist, in denen ihr verschiedene Aufgaben lösen müsst. Zu Beginn jedes Teiles steht genau beschrieben, was ihr zu tun habt. Für jeden Teil habt ihr nur eine bestimmte Zeit zur Verfügung.

Falls Kinder fragen, wie viel Zeit das ist, dann antworten: „Die Zeitbegrenzung ist von Aufgabe zu Aufgabe verschieden“.

Wenn ich Stopp rufe, ist die Zeit vorbei und ihr legt die Stifte auf den Tisch. Ihr dürft nicht mehr weiter schreiben. Auch wenn ihr nicht alle Aufgaben gelöst habt. Wenn ihr schon vor der Zeit fertig seid, dürft ihr nicht vorblättern, sonst müsst ihr den Test abbrechen und den Raum verlassen. Während ihr die Aufgaben bearbeitet, müsst ihr bitte leise sein und ihr dürft nicht miteinander sprechen.

Ihr könnt jetzt anfangen.

Mit der Stoppuhr wird die Zeit gestoppt. Wenn die Zeit abgelaufen ist:

„ Stopp! Alle legen die Stifte hin“

Schreibt noch jemand, dann nochmals ermahnen:

„Nicht mehr weiter schreiben.“

Bitte blättert jetzt um und beginnt mit dem 2. Teil...wiederholen bis zum vorletzten Untertest.

.....

„Stopp!! Stifte hinlegen und Testhefte zu machen, ich werde sie jetzt einsammeln.“

„Danke, dass ihr bei meiner Testung mitgemacht habt.“

Kinder dürfen sich Zuckerl....nehmen.

12 Lebenslauf

Martina Scheffl
Manhartstraße 53/4/7
2000 Stockerau

Tel.: 0680/1219143
e-mail: martina_scheffl@gmx.at

Persönliche Daten: Geboren am 27. April 1983 in Wien
Österreichische Staatsbürgerschaft, ledig

Bildungsgang: Volksschule in Stockerau
Bundes- und Bundesrealgymnasium in Stockerau
Erzbischöfliches Real- und Aufbaugymnasium in
Holltarunn. Berufsreifepfung 2001
Seit Wintersemester 2001 Studentin des Diplomstudiums
Psychologie an der Universität Wien

Berufliche Tätigkeiten: Seit 2001: als Lernbegleiterin am NÖ Hilfswerk in
Stockerau tätig.

August/September 2006: 6-wöchiges Pflichtpraktikum am
Weißen Hof in Klosterneuburg

Mai 2011: Ausbildung zur Legasthienetrainerin.