



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Die Stabilität des AID 2

Verfasserin

Claudia Parfuss

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, im März 2009

Studienkennzahl lt. Studienblatt:

A 298

Studienrichtung lt. Studienblatt:

Psychologie

Betreuer:

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

Danksagung

Das Erstellen dieser Diplomarbeit hat mir viel Freude bereitet, da mir die Arbeit mit den Kindern Spaß machte, ich das Thema sehr spannend finde und nicht zuletzt weil die Zusammenarbeit mit anderen Menschen gut funktionierte. Deshalb möchte ich allen Leuten, die mich unterstützt haben nochmals danken.

Als erstes danke ich Herrn **Univ.-Prof. Dr. Mag. Klaus D. Kubinger** für die gute Betreuung. Besser hätte ich mir diese nicht vorstellen können.

Herrn **Mag. Thomas Pletschko** danke ich für die Hilfe am Anfang meiner Diplomarbeit, Institutionen zu finden, die mich unterstützten.

Ich danke allen Institutionen, die mir anonymisierte Daten zur Verfügung stellten.

Frau **Mag. Stefana Holocher-Ertl** gilt ein großes Dankeschön für die zur Verfügung gestellten anonymisierten Testergebnisse der Test- und Beratungsstelle der Psychologischen Fakultät.

Vielen Dank an das Kompetenzzentrum für Erziehungs- und Schulfragen in Wolkersdorf, an deren Leiterin Frau **Dr. Renate Strasser** sowie an Frau **Petra Hickl** und alle Klinischen und Gesundheitspsychologinnen.

Auch Frau **Dr. Mathilde Zeman** danke ich stellvertretend für die Schulpsychologie Wien und deren MitarbeiterInnen.

Vielen Dank an Frau **Mag. Elke Durstberger**, an Frau **Mag. Theresia Herbst**, an Herrn **Mag. Christian Katzbeck** und an Herrn **Mag. Heinz Längle**, die mir anonymisierte Daten Ihrer Praxen zukommen ließen.

Ich danke Frau **Mag. Sandra Reisenhofer** für die Leitung der drei AID 2 Zertifizierungskurse, deren Teilnehmer für mich Kinder und Jugendliche testeten. Frau **Mag. Veronika Marek** danke ich für die Mithilfe bei der Organisation und Aufbewahrung der Testbögen.

Ein großes Dankeschön geht an alle Testleiter, die viele Mühen nicht gescheut haben, testeten, Rückmeldungen an die Eltern schrieben und stets gut mit mir zusammenarbeiteten: **Johanna Böck, Catrin Denk, Karin Fleischhacker, Tobias Glück, Iris Greifeneder, Brigitte Hansmann, Sandra Hofer, Claudia Kiesel, Mag. Sara Lampe, Simon Lehner, Gerda Neumann, Sandra Neussl, Nadine Rüdissler, Mag. Silvia Schubhart, Elisabeth Unterhofer** und **Andrea Witting**.

Vielen Dank auch an die Eltern und deren Kinder, die freiwillig an den Testungen teilnahmen und somit einen wesentlichen Beitrag zu dieser Diplomarbeit leisteten.

Auch meinen KollegInnen aus dem Forschungsseminar danke ich vielmals, die mich im Rahmen der Hilfestunden verlässlich und gut unterstützten: **Elke Eisenhauer, Mag. Sylvia Martinovsky, Barbara Strasser, Alexander Uitz, Katharina Vock, Gunnar Vogt, Andrea Wagner** und **Mag. Birgit Zuba**. Speziell danke ich nochmals **Claudia Kiesel** für das Korrekturlesen dieser Diplomarbeit.

Last but not least danke ich meinen Freundinnen und Freunden für den Beistand. Auch meiner Familie gilt mein Dank. Vielmals danke ich meinen Eltern, dass Sie mir diese gute Ausbildung ermöglichten. Meinem Vater, **Ing. Josef Parfuss** danke ich auch für das Korrekturlesen dieser Diplomarbeit.

Inhaltsverzeichnis

Danksagung	3
Tabellenverzeichnis	7
Abbildungsverzeichnis	9
Einleitung	11
1 Theoretischer Teil	13
1.1 <i>Intelligenz</i>	14
1.1.1 Definition von Intelligenz	14
1.1.2 Überblick über die Geschichte der Intelligenztests	16
1.1.3 Entwicklung der Intelligenz im Kindes- und Jugendalter	19
1.1.4 Stabilität und Inkonsistenz der Intelligenz	24
1.2 <i>Reliabilität</i>	30
1.2.1 Definition der Reliabilität	30
1.2.2 Messgenauigkeit im Sinne der Klassischen Testtheorie	31
1.2.2.1 Retest-Reliabilität	32
1.2.2.2 Paralleltest-Reliabilität	33
1.2.2.3 Innere Konsistenz	34
1.2.3 Messgenauigkeit im Sinne der Probabilistischen Testtheorie	35
1.2.4 Reliabilitätsanforderungen	37
1.2.5 Die Retest-Reliabilität in vergleichbaren Tests	39
1.2.5.1 Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK)-III und HAWIK-IV	40
1.2.5.2 Kaufmann-Assessment Battery for Children (K-ABC)	42
1.3 <i>Das Adaptive Intelligenz Diagnostikum 2 (AID 2)</i>	44
1.3.1 Allgemeine Beschreibung der Testbatterie	44
1.3.2 Die Reliabilität des AID und des AID 2	45
1.3.3 Die Subtests des AID 2	49
2 Empirischer Teil	55
2.1 <i>Ziel der Untersuchung</i>	56
2.2 <i>Untersuchungsplan</i>	57
2.2.1 Vorgabe des AID 2	57
2.2.2 Intendierte Stichprobe	57

2.2.3	Untersuchungszeitraum, Durchführung und Arbeitsaufwand der Untersuchung	58
2.3	<i>Durchführung der Untersuchung</i>	59
2.3.1	Testpersonenakquirierung	59
2.3.2	Abweichungen der Durchführung vom Untersuchungsplan	61
2.3.3	Stichprobenbeschreibung	62
2.3.4	Auswertungsverfahren	67
2.4	<i>Ergebnisse</i>	68
2.4.1	Deskriptive Ergebnisse der Testwerte	68
2.4.2	Retest-Reliabilität bei Vorgabe am selben Tag	73
2.4.3	Stabilität nach einem Intervall von ca. einem Monat	74
2.4.4	Stabilität nach einem Intervall von ein bis drei Monaten	76
2.4.5	Stabilität nach einem Intervall von 0.5 bis 1.92 Jahren	78
2.4.6	Stabilität nach einem Intervall von 2 bis 2.67 Jahren	80
2.4.7	Stabilität nach einem Intervall von 2.75 bis 6.42 Jahren	82
3	Diskussion	84
4	Zusammenfassung	93
	Literaturverzeichnis	95
5	Anhang	103
5.1	<i>Abstract</i>	104
5.2	<i>E-Mail an PsychologInnen des BÖP</i>	105
5.3	<i>Attachment des e-Mails an PsychologInnen</i>	106
5.4	<i>Elternbrief der Autorin</i>	109
5.5	<i>Elternbrief des Kompetenzzentrums Wolkersdorf</i>	110
5.6	<i>Elternbrief der Schulpsychologie Wien</i>	111
5.7	<i>Beispielrückmeldung</i>	113
5.8	<i>Deskriptive Statistik der Testwerte nach den unterschiedlichen Intervallen</i>	114
5.9	<i>Lebenslauf</i>	125

Tabellenverzeichnis

Tabelle 1: Retest-Reliabilität des HAWIK-IV für 3 Subtests	41
Tabelle 2: Testwiederholungsreliabilität des K-ABC für die Altersstufen 9;0 – 12;5, N = 24	43
Tabelle 3: Standardschätzfehler und split-half-Reliabilität des AID 2.....	45
Tabelle 4: Stabilität des AID	47
Tabelle 5: Retestreliabilität des AID 2 nach 4 Wochen; N=128	48
Tabelle 6: Häufigkeit der Vorgabe von Untertests in Kombination mit oder ohne Zusatztests	61
Tabelle 7: Grund der Vorstellung.....	63
Tabelle 8: Wohnort der Testpersonen zum zweiten Testzeitpunkt	63
Tabelle 9: Häufigkeiten und Geschlecht der Testpersonen über die Akquirierungsquellen.....	64
Tabelle 10: Verteilung der Geschlechter über das Alter in Jahren.....	64
Tabelle 11: Verteilung des Alters der Testpersonen über die verschiedenen Zeitintervalle	65
Tabelle 12: Häufigkeiten der Testpersonen über die Intervalle.....	66
Tabelle 13: Verteilung der Quelle der Testpersonenakquirierung über die Intervalle	67
Tabelle 14: Deskriptive Statistik der Testwerte – Testwerte der Ersttestung.....	69
Tabelle 15: Deskriptive Statistik – Testwerte der Zweittestung.....	70
Tabelle 16: Retest-Reliabilität bei Vorgabe am selben Tag	73
Tabelle 17: Stabilität nach ca. einem Monat	74
Tabelle 18: Stabilität nach ein bis drei Monaten	76
Tabelle 19: Stabilität nach 0.5 bis 1.92 Jahren	78
Tabelle 20: Stabilität nach 2 bis 2.67 Jahren.....	80
Tabelle 21: Stabilität nach 2.75 bis 6.42 Jahren.....	82

Tabelle 22: Vorgabe am selben Tag Ersttestung; Testwerte deskriptiv	114
Tabelle 23: Vorgabe am selben Tag;Zweittestung; Testwerte deskriptiv	114
Tabelle 24: Intervall ca. 1 Monat;Ersttestung; Testwerte deskriptiv.....	115
Tabelle 25: Intervall ca. 1 Monat; Zweittestung; Testwerte deskriptiv.....	116
Tabelle 26: Intervall 1-3 Monate; Ersttestung; Testwerte deskriptiv	117
Tabelle 27: Intervall 1-3 Monate; Zweittestung; Testwerte deskriptiv	118
Tabelle 28: Intervall 0.5 bis 1.92 Jahre; Ersttestung; Testwerte deskriptiv.....	119
Tabelle 29: Intervall 0.5 bis 1.92 Jahre; Zweittestung; Testwerte deskriptiv.....	120
Tabelle 30: Intervall 2 bis 2.67 Jahre; Ersttestung; Testwerte deskriptiv.....	121
Tabelle 31: Intervall 2 bis 2.67 Jahre; Zweittestung; Testwerte deskriptiv.....	122
Tabelle 32: Intervall 2.75 bis 6.42 Jahre; Ersttestung; Testwerte deskriptiv.....	123
Tabelle 33: Intervall 2.75 bis 6.42 Jahre; Zweittestung; Testwerte deskriptiv.....	124

Abbildungsverzeichnis

Abbildung 1: Vorgabeschema nach dem Branched-testing	19
Abbildung 2: Intelligenzanstieg und –abfall; Hamburger-Gesamtskala.....	21
Abbildung 3: Leistungen im Verbal- und Handlungsteil der Wechsler Tests	21
Abbildung 4: Erfassung allgemeiner, fluider und kristalliner Intelligenz	22
Abbildung 5: Korrelationen zwischen der Intelligenz in verschiedenen Lebensaltern und der Intelligenz mit 18 Jahren	25
Abbildung 6: Berechnung des Konfidenzintervalls.....	37
Abbildung 7: T-Werte und deren Konfidenzintervalle.....	38

Einleitung

Intelligenz und kognitive Fähigkeiten sind ein Gebiet der Psychologie, das bereits viel Interesse erweckte und die Themen sind immer noch sehr populär, sowohl in der Wissenschaft als auch im Alltag. In Fernsehsendungen wird z.B. über den IQ von Stars diskutiert, Eltern wollen den IQ ihrer Kinder wissen und auch viele Bücher erschienen, um „Laien“ in das Thema einzuführen (vgl. z.B.: Dietrich, 1972; Ehrle, 1962; Funke & Vaterrodt-Plünnecke, 1998).

Psychologische Diagnostik sollte jedoch nur bei gegebener Fragestellung erfolgen und neben intellektuellen Fähigkeiten auch emotionale Bedingungen berücksichtigen, um notwendige Maßnahmen festsetzen und entsprechende Förderungen durchführen zu können (Kubinger, 1997).

Vor allem bei Kindern in der Schul- und Bildungsberatung kommt es oft zu Fragestellungen, zu deren Beantwortung Intelligenztests eingesetzt werden (vgl. z.B. Heller, 2000), wie z.B.: Ist mein Kind schulreif? Welchen Ausbildungsweg soll es wählen? Für welchen Schulzweig ist mein Kind geeignet? Liegt eine Hochbegabung vor? Hat mein Kind Legasthenie?

Um diese Fragen zu beantworten, braucht man ein zuverlässiges Messinstrument, das auch über die Zeit stabile Werte liefert. Das Adaptive Intelligenz Diagnostikum 2 (der AID 2) von Kubinger und Wurst (2000) wird häufig zur Begabungsdiagnostik eingesetzt. Inwiefern die Messwerte, die damit erhoben werden, auch über die Zeit stabil sind, soll anhand dieser Diplomarbeit geklärt werden.

Es werden dafür zuerst wichtige Erkenntnisse der Geschichte der Intelligenztests bis zu den neuesten wissenschaftlich Erkenntnissen der Intelligenzmessung und der Reliabilität dargestellt und der AID 2 (Kubinger und Wurst, 2000) in seinen wichtigsten Grundzügen vorgestellt. Danach wird auf die durchgeführte Studie zur Retest-Reliabilität eingegangen. In dieser Studie wurde die Stabilität des AID 2 (Kubinger und Wurst, 2000) berechnet. Besonderes Augenmerk wird dabei auf die Stabilität nach kurzen Zeitintervallen gelegt, da hier die Ergebnisse nicht so sehr

durch Veränderungen der Entwicklung der Testpersonen verfälscht werden. Deshalb wurden Teile des Tests an einem Tag zweimal vorgegeben, und der gesamte Test wurde nach einem Monat und nach drei Monaten wiederholt. Auch die Korrelationskoeffizienten nach dem Intervall von ca. einem Jahr, nach ca. zwei Jahren und nach ca. drei bis sechs Jahren wurden berechnet, um die Langzeitstabilität zu berücksichtigen.

1 Theoretischer Teil

1.1 Intelligenz

Im Folgenden soll zuerst die Problematik der Intelligenzdefinition kurz dargestellt werden. Danach wird auf einige Gebiete der Intelligenzforschung eingegangen: auf den experimentalpsychologischen Ansatz, der die Gesetzmäßigkeiten intelligenten Verhaltens erforscht (vgl. Punkt 1.1.2), und der auch „Informationsverarbeitungsansatz“ genannt wird. Auch der entwicklungspsychologische Ansatz, wird beschrieben, der sich mit der Entwicklung der Intelligenz im Leben eines Menschen beschäftigt (vgl. Punkt 1.1.3). Die größte Bedeutung hat der korrelationsstatistische und differentialpsychologische Ansatz, der auch „psychometrischer Ansatz“ genannt wird (vgl. Punkt 1.1.2 und 1.1.4). Dabei werden mithilfe von Tests zur Messung individueller Unterschiede in Denkergebnissen Aussagen über den Aufbau der Intelligenzstruktur gemacht (vgl. z.B. Funke & Vaterrodt-Plünnecke, 1998; Guthke, 1991).

1.1.1 Definition von Intelligenz

Intelligenz ist ein wichtiges Merkmal, das einen positiven Zusammenhang mit dem Fortkommen in Gesellschaften westlicher Lebensart aufweist, wie z.B. mit der Ausbildung und der beruflichen Tätigkeit (Amelang, Bartussek, Stemmler & Hagemann, 2006). Die vielleicht schwierigste Herausforderung in der Intelligenzforschung ist herauszufinden, wie man Intelligenz definiert (Sternberg, 2002a). Es gibt keine einheitliche und verbindliche Terminologie (Kubinger, 2006). Viele unterschiedliche Definitionen konkurrieren miteinander (Guthke, 1991). Hier sollen einige Definitionen angeführt werden.

Binet und Simon (1904) definierten Intelligenz als gutes urteilen, verstehen und denken in der Bewältigung einer aktuellen Situation. Sie überprüften z.B. die Fähigkeit zur Auge-Hand Koordination, und die Gedächtniskapazität. Wechsler (1944, S.13, zitiert nach Amelang et al., 2006) bezeichnet Intelligenz als Fähigkeit eines Individuums „zweckvoll zu handeln, vernünftig zu denken und sich mit der Umwelt wirkungsvoll auseinander zu setzen“.

Stern (1950) betont die Neuartigkeit eines Problems. Er versteht unter Intelligenz die Fähigkeit sein Leben selbst umgestalten und sich auf neue Forderungen einstellen zu können (zitiert nach Amelang et al., 2006).

Einen wesentlichen Beitrag zur Erforschung des Konstruktes lieferten die Verfahren zur Messung von Intelligenz (Amelang et al., 2006). Von Boring (1923) stammt die häufig zitierte Aussage Intelligenz sei das, was der Intelligenztest misst (vgl. Amelang et al., 2006; Guthke, 1991).

Im Zusammenhang mit dem AID 2 (Kubinger & Wurst, 2000) wird Intelligenz in Anlehnung an Cattells Theorie „Wissen ist investierte Intelligenz“ definiert als

...das Bündel aller kognitiver Voraussetzungen, die notwendig sind, um Wissen zu erwerben und Handlungskompetenzen zu entwickeln – wobei „Kognition“ sich bezieht auf „jeden Prozess, durch den das Lebewesen Kenntnis von einem Objekt erhält oder sich seiner Umwelt bewusst wird...: Wahrnehmung, Erkennen, Vorstellen, Urteilen, Gedächtnis, Lernen, Denken, ... Sprache“ (Lexikon der Psychologie – Arnold, Eysenck & Meili, 1980, S. 1086)...(S.30)

Die Intelligenzanlage ist weitgehend unbekannt und genetisch bestimmt. Der Intelligenzstatus entsteht aus der Wechselbeziehung zwischen Anlage, Umwelt und Erziehung. Dieser Intelligenzstatus ist zu verschiedenen Zeitpunkten im Leben des Menschen durch Tests feststellbar, da diese mehr oder minder vollständige Abbildungen liefern. Die Intelligenzpotenz ist die Fähigkeit, Leistung zu steigern (Guthke, 1991).

Auch implizite Theorien von „Normalbürgern“ wurden bereits erfasst. So wurde z.B. festgestellt, dass Experten und Laien in genannten Verhaltensweisen, die für Intelligenz sprechen, recht hohe Übereinstimmungen zeigten. Als intelligente Verhaltensweisen wurden z.B. genannt „sieht alle Aspekte eines Problems“, „kennt sich in bestimmten Wissensgebieten gut aus“ oder „akzeptiert andere so wie sie sind“. Diese Aussagen konnten in die Kategorien „Praktische Problemlösefähigkeit“, „Verbale Fähigkeiten“ und „Soziale Kompetenz“ eingeordnet werden (Sternberg, Conway, Bernstein & Ketron, 1981, zitiert nach Amelang et al., 2006). In

wissenschaftlichen Modellen werden die soziale und die praktische Intelligenz eher vernachlässigt (Amelang et al., 2006; Sterberg, 2002b).

1.1.2 Überblick über die Geschichte der Intelligenztests

Tests als Eignungsprüfung gibt es schon lange. Bereits im alten China 221 vor Christus bis 220 nach Christus gab es z.B. Verfahren zur Eignungsauslese von Beamten. Psychologische Tests als wissenschaftliche Methode haben jedoch eine kurze Geschichte (Guthke, 1996).

Die Anfänge der heutigen Psychologischen Diagnostik entstammen der Experimentellen Psychologie, in der Sinnesfunktionen untersucht und Reaktionszeiten gemessen wurden. Vertreter dieser Richtung war z. B. Francis Galton (1869), der von Darwins Evolutionstheorie beeinflusst wurde (Brody & Brody, 1976). James McKeen Cattell verwendete als Erster den Begriff „mental test“ (1890, zitiert nach Brody & Brody, 1976). Emil Kraepelin stellte den Testpersonen später umfassendere Aufgaben, wie z. B. Rechnen und Problemlösen (Kubinger, 2006).

Simon und Binet (1904) entwickelten einen Test, um feststellen zu können, welche Kinder wegen Minderbegabung in Spezialschulen unterrichtet werden sollten.

Nach Eintritt der USA in den ersten Weltkrieg wurde von amerikanischen Psychologen unter der Leitung von Yerkes 1917 der erste Gruppenintelligenztest entwickelt. Für Analphabeten und Lesekundige gab es zwei verschiedene Versionen: Army Alpha und Army Beta (Guthke, 1991).

Spearman (1927, zitiert nach Guthke, 1991), der Begründer der Faktorenanalyse und der systematischen Intelligenzforschung beobachtete positive Korrelationen zwischen verschiedenen Intelligenztests und schloss daraus, dass der gemeinsame Generalfaktor (g-Faktor) „mental energy“ dahinter stecken müsse. Laut Spearman basiert jeder Test, der Intelligenz misst auf einem Generalfaktor und auf einem Spezialfaktor. Dies wird auch als Zwei-Faktoren-Theorie bezeichnet (Brody & Brody, 1976). In zahlreichen Arbeiten wurde bereits versucht den Generalfaktor zu bestätigen oder zu widerlegen (vgl. z.B. Sternberg & Gardner, 1982).

Die „Primary mental Abilities“ von Thurstone aus dem Jahr 1931 sind auch heute noch eine Grundlage für viele Intelligenztests und spezielle Leistungstests. Die sieben unabhängigen Fähigkeiten „Verbal Comprehension“, „Word Fluency“, „Number“, „Space“, „Memory“, „Perceptual Speed“ und „Reasoning“ werden dabei als Determinanten intelligenten Verhaltens gesehen (zitiert nach Kubinger, 2006). Bekannte Intelligenztests, die auf diesem Modell basieren, sind die verschiedenen Versionen des Intelligenz-Struktur-Tests (I-S-T) von Amthauer (1957).

Durch die Intelligenztests nach David Wechsler, die nach der Wechsler-Bellevue Intelligence Scale von 1939 (zitiert nach Amelang & Schmidt-Atzert, 2006) entwickelt wurden, sollen so viele spezifische Fähigkeiten wie möglich erfasst werden, um die globale Fähigkeit einer Testperson bestimmen zu können. David Wechsler stellte einen Test aus vorhandenen Aufgaben zusammen. Die meisten Tests gliederten die „Allgemeinintelligenz“ in einen Handlungsteil und einen Verbalteil (Amelang & Schmidt-Atzert, 2006; Kubinger, 2006). Es wird neben der Berechnung der Allgemeinintelligenz auch eine Profilauswertung vorgenommen, die auf Rossolimo (1860-1928) zurückgeht (Guthke, 1991). Im neuen Hamburg Wechsler Intelligenztest für Kinder (HAWIK-IV) wird auf eine Teilung in Verbal- und Handlungsteil verzichtet, dafür wird ein Schwerpunkt auf fluides logisches Denken, die Kapazität des Arbeitsgedächtnisses und die Geschwindigkeit von Informationsverarbeitungsprozessen gelegt (Petermann, 2006).

Mithilfe der Progressiven Matrizen-Tests nach Raven (zitiert nach Guthke, 1991) sollte 1956 die kulturunabhängige Grundintelligenz, der g-Faktor von Spearman bzw. die „flüssige Intelligenz“ nach Catell (1963) gemessen werden. Die Annahme, dass die „flüssige Intelligenz“ die angeborene Komponente darstellt, bleibt unbewiesen (Guthke, 1991). Nach Catell (1963) sollten dagegen kenntnisabhängige Verbaltests die bildungsbedingte Komponente bzw. die „kristallisierte Intelligenz“ darstellen.

Im „Structure of Intellect“-Modell von Guilford (1956) ergeben sich durch die Kombination von vier Inhalten (z.B. figural, symbolisch) mit fünf Operationen (z.B. Kognition, Gedächtnis) und sechs Produkten (z.B. Einheiten, Klassen) 120 Faktoren.

Jäger (1984) vereinigte Elemente aus den Kategorisierungssystemen von Spearman, Thurstone und Guilford im Berliner Intelligenzstrukturmodell.

Durch Lerntestkonzepte wurde 1970 die reine Statusdiagnostik überwunden. Mithilfe der Prozessdiagnostik kann das Intelligenzpotential und die Lernfähigkeit gemessen werden (Guthke, 1990; Guthke, 1991; Kubinger, 2006). Diese Lerntests erfassen jedoch nur einen speziellen Aspekt der Intelligenz (Kubinger, 2006).

Die triarchische Theorie von Sternberg (1985, zitiert nach Petermann, 2006) berücksichtigt bei der Einschätzung der Intelligenz sowohl kognitive Verarbeitungsprozesse als auch Faktoren der Erfahrung und des Kontextes.

Ein Schwerpunkt neuer Forschungsarbeiten liegt beim adaptiven Testen, das auf der Probabilistischen Testtheorie beruht. Die Itemauswahl soll dabei dem Leistungsvermögen der Testperson angepasst sein. Eine Art des adaptiven Testens ist das „tailored-testing“, das durch Computertestungen möglich wurde. Je nachdem ob die Testperson ein Item löst oder nicht, bekommt sie als nächstes ein schwierigeres oder ein leichteres vorgegeben. Das Item, das als nächstes vorgegeben werden soll, muss mit seinem Schwierigkeitsparameter σ_i die Bedingung von Formel 1 erfüllen (Kubinger, 2006).

Formel 1: Tailored Testing, Itemvorgabe

$$\frac{e^{\hat{\xi}_v - \sigma_i}}{1 + e^{\hat{\xi}_v - \sigma_i}} \cdot \frac{1}{1 + e^{\hat{\xi}_v - \sigma_i}} = \text{Max!}$$

Dabei steht $\hat{\xi}_v$ für die aktuelle Schätzung des Personenparameters ξ_v der Person v. Das Maximum von 0.25 wird erfüllt, wenn die Wahrscheinlichkeit, dass die Person das Item löst, 50% beträgt. Mit diesem Item kann der Personenparameter bestmöglich geschätzt werden (Kubinger, 2006).

Beim „branched-testing“, das auch zum adaptiven Testen gehört, werden Items in Gruppen zusammengefasst. Der Vorteil ist, dass diese Variante auch ohne Computer vorgegeben werden kann. Im AID 2 wird dieses Vorgabeschema für viele Untertests verwendet (Kubinger & Wurst, 2000). Je nach Alter wird mit einer anderen Gruppe begonnen. In Abhängigkeit von der Anzahl der gelösten Items der Testperson, werden schwierigere oder leichtere Gruppen vorgegeben (vgl. Abbildung 1).

Das adaptive Testen liefert einige Vorteile: eine höhere Messgenauigkeit bei derselben Testlänge, in allen Fähigkeitsbereichen wird gleich genau gemessen und die Motivation der Testperson bleibt länger aufrechterhalten.

1.1.3 Entwicklung der Intelligenz im Kindes- und Jugendalter

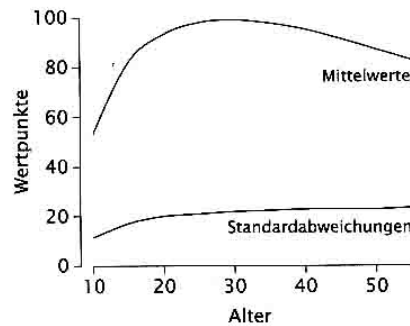
19

wird. Viele Studien beschäftigen sich auch mit den Auswirkungen von Ernährung auf die intellektuelle Entwicklung. Einen guten Überblick liefert Grigorenko (2003). Vor allem Proteinmangel und Unterernährung können kognitive Entwicklungen beeinträchtigen. Nach McCall (1981) können in den ersten zwei Lebensjahren nur sehr schwerwiegende Umwelteinflüsse die Entwicklung des Kindes negativ beeinflussen. Danach führen sowohl Umwelteinflüsse als auch die genetische Veranlagung zu individuellen Unterschieden.

Es gibt sechs weitere Perspektiven nach deren Anschauung man die intellektuelle Entwicklung während des Lebens betrachten kann: die psychometrische Sichtweise, die Sichtweise nach Piaget, die Neo-Piaget Sichtweise, die Sicht der Informationsverarbeitung, des Lernens und des Kontextes. Die Vorteile, die Entwicklung aus verschiedenen Perspektiven zu sehen sind zahlreich. Es kann ein genaues Bild von den Entwicklungen der Altersstufen aufgezeigt werden. Jede Perspektive zeigt eine neue Facette (Berg, 2002). Auf die psychometrische Perspektive soll hier etwas genauer eingegangen werden, da sie die Grundlage dieser Arbeit darstellt, aber auch die anderen Perspektiven werden kurz vorgestellt.

Manche Menschen assoziieren mit dem Wort Intelligenz nur Intelligenztestungen. Die psychometrische Herangehensweise ist jedoch bloß eine von vielen, um intellektuelle Entwicklungen zu erforschen. Diese konzentriert sich verstärkt auf Ergebnisse intellektueller Leistungen, um die intelligenten Entwicklungen zu verschiedenen Zeitpunkten während der Entwicklung zu charakterisieren. Die psychometrische Perspektive weist, wie auch in Punkt 1.1.2 beschrieben, die längste Tradition im Forschungsfeld der Intelligenzentwicklung auf (Berg, 1992). Mit der interessanten Frage, ab wann intraindividuelle quantitative Veränderungen bedeutsam sind, beschäftigt sich z.B. Alsaker (1992). Aus psychometrischer Sichtweise wurde beispielsweise die Standardisierungsstichprobe der Wechsler Tests untersucht (Berg, 1992). Die Testpersonen erzielten die in Abbildung 2 zu sehenden Ergebnisse. Die Wertpunkte sind eine direkte Funktion der gelösten Aufgaben. Der steile Anstieg verlangsamt sich im Jugendalter und fällt nach dem Alter von 30 Jahren allmählich ab (Amelang et al., 2006).

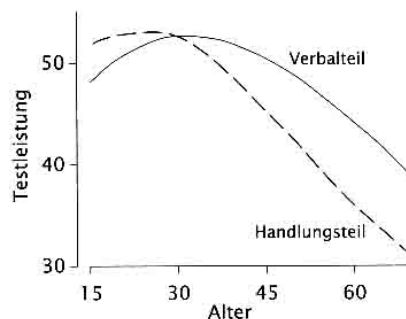
Abbildung 2: Intelligenzanstieg und –abfall; Hamburger-Gesamtskala



Anmerkung: Wechsler (1964, zitiert nach Amelang et al., 2006)
 Aus *Differentielle Psychologie und Persönlichkeitsforschung* (6. Auflage) (S.173)
 von M. Amelang, D. Bartussek, G. Stemmler & D. Hagemann, 2006, Stuttgart:
 Kohlhammer.

In Abbildung 3 werden die Unterschiede zwischen Handlungs- und Verbalteil dargestellt. Der Leistungsabfall des Handlungsteiles ist im Alter steiler. Hier handelt es sich jedoch um Querschnittsuntersuchungen (Amelang et al., 2006).

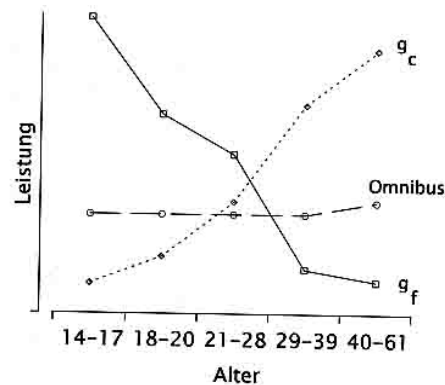
Abbildung 3: Leistungen im Verbal- und Handlungsteil der Wechsler Tests



Anmerkung: Aus *Differentielle Psychologie und Persönlichkeitsforschung*
 (6. Auflage) (S.173) von M. Amelang, D. Bartussek, G. Stemmler & D. Hagemann,
 2006, Göttingen: Hogrefe. Stuttgart: Kohlhammer.

Von Horn und Cattell (1967, zitiert nach Amelang et al., 2006) wurde ebenfalls eine Querschnittsuntersuchung durchgeführt. Dabei wurde festgestellt, dass die Entwicklung der fluiden Intelligenz (g_f), die z.B. induktives Denken und assoziatives Denken misst, zwischen dem 14. und 15. Lebensjahr abgeschlossen ist. Kristallisierte Intelligenz (g_c), wie z.B. verbales Verständnis erreicht ihre volle Entwicklung vier bis fünf Jahre später. Die Entwicklung wie in Abbildung 4 ist noch nicht bestätigt worden, da noch keine Längsschnittstudien vorliegen.

Abbildung 4: Erfassung allgemeiner, fluider und kristalliner Intelligenz



Anmerkung: Horn und Cattell (1967, zitiert nach Amelang et al., 2006)
 Aus *Differentielle Psychologie und Persönlichkeitsforschung* (6. Auflage) (S.185)
 von M. Amelang, D. Bartussek, G. Stemmler & D. Hagemann, 2006, Göttingen:
 Hogrefe. Stuttgart: Kohlhammer.

Auf die psychometrische Perspektive der intellektuellen Entwicklung in Kindheit und Jugend gehen auch Gardner und Clark näher ein (1992). McCall, Hogarty und Hurlburt (1972) untersuchten wie gut der IQ im Erwachsenenalter durch den IQ in der Kindheit vorhergesagt werden kann. Sie kommen zu dem Schluss, dass es keinen konstanten Generalfaktor in der Entwicklung gibt. Sie unterscheiden verschiedene Kompetenzen während unterschiedlicher Abschnitte in der Kindheit, die für die geringen Korrelationskoeffizienten verantwortlich sind.

Aus Sicht der anderen Perspektiven ist der Prozess, durch den intellektuelle Leistungen konstruiert werden, interessant. Die Perspektive von Piaget konzentriert sich auf einzelne Verhaltensweisen, die die meisten Kinder zu einem bestimmten Entwicklungszeitpunkt aufweisen, während andere Perspektiven sich auf Unterschiede zwischen Individuen spezialisieren. Piagets Perspektive entstand 1919 während seiner frühen psychologischen Untersuchungen in Binets psychometrischem Labor. Die Kodierung der Antworten in „richtig“ und „falsch“ waren für ihn unbefriedigend. Er untersuchte die dahinter stehenden Schlussfolgerungen der Kinder. Diese Untersuchungen führten zu dem Ergebnis, dass die Art des Schlussfolgerns von Kindern in verschiedenen Altersstufen qualitative Unterschiede in der Art des Denkens aufweisen (Berg, 1992). Piaget (1952, zitiert nach Berg, 1992) beschrieb die intellektuelle Entwicklung als Prozess, der Wissen aus der Interaktion mit der Umwelt entstehen lässt. Intelligenz sah er als Instrument an, um

ein Gleichgewicht zwischen kognitiven Strukturen und der Umwelt zu schaffen. Dabei unterschied er vier Perioden: die sensumotorische Periode (0-2 Jahre), die Periode des voroperativen Denkens (2-7 Jahre), die Periode des konkret-operativen Denkens (7-11 Jahre) und die des formal-operativen Denkens (11 Jahre-Erwachsenenalter) als letzte Periode (Funke & Vaterrodt-Plünnecke, 1998). Piaget vertrat die Ansicht, dass Entwicklung zum Erwerb allgemeiner kognitiver Strukturen führt, die das Lernen erst ermöglichen (Ginsburg & Opper, 1988/1998; Aebli, 1963).

Die Neo-Piaget Perspektive entwickelte sich 1970 nachdem viel an Piagets Theorie kritisiert wurde. Die Vertreter interessierten sich hauptsächlich dafür, wie kognitive Strukturen innerhalb einer Domäne angelegt sind. Außerdem stellten sie individuelle Unterschiede in kognitiven Strukturen fest und charakterisierten den Zusammenhang zwischen kognitiven, emotionalen und affektiven Strukturen (vgl. Berg, 1992; Case, 1992).

Die Anfänge der Informationsverarbeitungstheorie waren Ende des 19. Jahrhunderts. 1960 lebte diese wieder auf, nachdem in der klassischen Lerntheorie geistige Prozesse nicht beachtet worden waren, die intelligentes Verhalten auslösen. Eine zentrale Frage ist, wie die Prozesse charakterisiert werden, die bei Individuen zu intelligentem Handeln führen. Es wird untersucht, was zwischen dem Input von Information und dem Output von Reaktion liegt und wie sich das im Laufe der Entwicklung verändert. Das Denken des Menschen wird mit der Informationsverarbeitung von Computern verglichen (Berg, 1992). Kail und Bisanz (1992) beschreiben die Steigerungen einer Vielzahl von Fähigkeiten während der Entwicklung eines Kindes. Diese werden durch Veränderungen der Strategien erklärt, die Kinder verwenden, wenn sie an Aufgaben herangehen und durch eine gesteigerte Anstrengung, die sie für die Lösung von Aufgaben aufbringen. Eine kognitive Theorie über Intelligenz und deren Entwicklung wird von Anderson (1992) beschrieben.

Die Wurzeln der „klassischen“ Lerntheorie liegen vor allem in den Werken Thorndikes Anfang 1900 (Berg, 1992). Die neuen Lerntheorien, wie sie z.B. von Canfield und Ceci (1992) beschrieben werden, beinhalten sowohl Aspekte der traditionellen Lerntheorie als auch kognitive Konstrukte der

Informationsverarbeitungsperspektive. Eine zentrale Frage dieser neuen Lerntheorien ist, wie die Menge und die Organisation von Wissen die intellektuelle Entwicklung beeinflussen. Unter Entwicklung wird verstanden, dass man immer mehr Experte in Wissensgebieten wird. Erfahrung garantiert jedoch nicht, dass man sich bei bestimmten Aufgaben optimal verhält.

Die kontextuelle Perspektive ist die neueste im Forschungsfeld der intellektuellen Entwicklung (Berg, 1992). Die Perspektive von Vygotsky (Wertsch & Kanner, 1992) stammt aus den 1920er und den 1930er Jahren. Sie wurde in den USA jedoch erst ab 1970 aufgegriffen. Diese besagt, dass das Denken des Individuums durch Interaktionen mit anderen geprägt wird. Kulturen liefern Richtlinien zum Problemlösen.

1.1.4 Stabilität und Inkonsistenz der Intelligenz

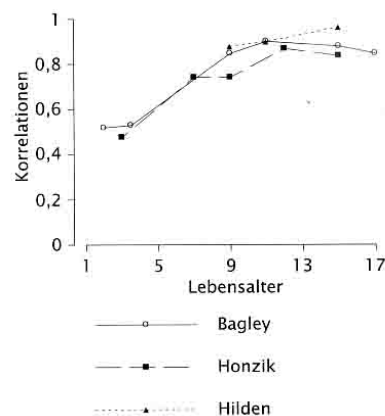
Die Inkonsistenz von Intelligenzmessungen bezieht sich auf intraindividuelle Veränderungen im Leistungsbereich. Solche Schwankungen resultieren nicht nur aus Messfehlern, sondern auch durch kurzzeitige Veränderungen die durch verschiedene schwankende (z.B.: Blutdruck, Blutzuckerspiegel, Müdigkeit, Motivation, Schlaf-Wach-Rhythmus, Stimmungen) oder stabile (z.B.: Ernährung, Gesundheit, soziale Stressoren) Einflüsse hervorgerufen werden (Dixon, 2003).

Die meisten Stabilitätskoeffizienten von Intelligenztests weisen eine Höhe von mindestens $r_{tt}=0.9$ auf. Um die Konstanz zu bestimmen werden wiederholt Testungen mit identischen Testformen durchgeführt. Dabei wird sowohl die Messgenauigkeit des Tests als auch die Schwankung des Merkmals erfasst. Umso jünger die Probanden sind, umso mehr sinken die Stabilitätskoeffizienten. Die Umgebungsbedingungen sind schwieriger zu standardisieren, die Kinder schwieriger zu motivieren und Entwicklungsschübe führen zu raschen Veränderungen (Amelang et al., 2006).

In einer Untersuchung mit 140 Kindern wurde eine Stabilität vom 3. zum 4. Lebensjahr des Stanford-Binet IQs von $r_{tt}=0.83$ ermittelt. Mit länger werdendem Intervall nahmen die Korrelationen ab. Der Korrelationskoeffizient vom 3. zum 12. Lebensjahr wies eine Höhe von $r_{tt}=0.46$ auf (Sontag, Barker & Nelson, 1958).

In Abbildung 5 sind die Ergebnisse dreier Untersuchungen dargestellt, die die Korrelationskoeffizienten ca. vom zweiten Lebensjahr bis zum 15. bzw. 17. Lebensjahr erhoben hatten. McCall erzielte ähnliche Ergebnisse (1977, zitiert nach Amelang, et al. 2006).

Abbildung 5: Korrelationen zwischen der Intelligenz in verschiedenen Lebensaltern und der Intelligenz mit 18 Jahren



Anmerkung: : Aus *Differentielle Psychologie und Persönlichkeitsforschung* (6. Auflage) (S.197) von M. Amelang, D. Bartussek, G. Stemmler & D. Hagemann, 2006, Göttingen: Hogrefe. Stuttgart: Kohlhammer.

Wie in diesen Studien ersichtlich ist, führen kürzere Intervalle zu höheren Koeffizienten. Bradway, Thompson und Cravens testeten Kinder der Eichstichprobe des Stanford-Binet Tests nach 10 Jahren ein zweites Mal und wieder 15 Jahre später ein drittes Mal. Vom 4. bis 14. Lebensjahr konnte eine Stabilität von $r_{tt}=0.65$ errechnet werden. Für das um fünf Jahre längere Intervall vom 14. bis zum 29. Lebensjahr ergab sich ein Wert von $r_{tt}=0.85$. Trotz längerem Zeitintervall resultierte im höheren Alter ein höherer Stabilitätskoeffizient (1958, zitiert nach Amelang, et al. 2006).

Das „Überlappungs-Modell“ von Anderson (1939) besagt, dass die gemessenen Leistungen aus der Summe der zu früheren Zeitpunkten gemessenen Leistungen bestehen, plus dem in der Zwischenzeit erfolgten Neuerwerb. Mit wachsendem Lebensalter muss demnach auch die Zuverlässigkeit ansteigen, weil die Zuwächse durch Neuerwerb geringer werden.

Pinneau (1961, zitiert nach Amelang, et al. 2006) ergänzte durch seine Befunde das Überlappungsmodell. Er stellte bei besseren Probanden einen größeren längsschnittlichen Zugewinn fest, bei schlechteren jedoch nur einen geringen Fortschritt.

Bei Vorgabe desselben Verfahrens oder zweier Paralleltests in kurzem Abstand werden Zuwächse von einem Drittel bis zu einem Viertel einer Standardabweichung beobachtet (Amthauer, 1957). Solche Veränderungen werden als Trainingseffekte gewertet (Amelang, 2006).

Auch nach 30 Jahren wurden Zuwächse der Testwerte beobachtet (Owens, 1963; Amelang & Hoppensack, zitiert nach Amelang, et al. 2006). Bei größeren Stichproben könnten ebenfalls ähnliche Effekte festgestellt werden, wie z.B. bei der Tennessee-Studie, bei der 3000 Schüler von Gebirgsgemeinden einen Zuwachs des mittleren IQs von 82 auf 93 aufwiesen (Wheeler, 1942). Ein Prozentrang von 83 der US-Rekruten des Zweiten Weltkrieges entsprach dem Prozentrang von 50 der US-Rekruten im Ersten Weltkrieg (Tuddenham, 1978). Dies waren wahrscheinlich Auswirkungen des ökonomischen, kulturellen und bildungsmäßigen Fortschritts (Amelang, et al. 2006). Garfinkel und Thorndike (1976) zeigten auf, dass die meisten Items des Binet-Tests von der Standardisierung des Jahres 1930 zu der Standardisierung im Jahr 1970 einfacher wurden.

Flynn entdeckte 1984 den nach ihm benannten „Flynn-Effekt“ in der amerikanischen Bevölkerung. In der Zeit von 1932 bis 1978 stieg der IQ um 13.8 Punkte. Dies wurde an den Standardisierungsstichproben der Stanford-Binet und der Wechsler-Tests abgelesen. Mittlerweile wurde dieser Effekt auch schon in vielen anderen Ländern anhand anderer Tests beobachtet. Der durchschnittliche IQ steigt seit ungefähr 1920 alle zehn Jahre um ca. fünf IQ Punkte an (Amelang, et al. 2006). Beim Flynn Effekt handelt es sich jedoch lediglich um einen Vergleich zwischen zwei Tests, der auch durch kulturelle Veränderungen erklärt werden könnte, z.B. weil Inhalte immer mehr zum Allgemeinwissen zählen. Deshalb sollten Intelligenztests alle zehn Jahre neu normiert werden. Dieser Effekt beruht auf querschnittlich erhobenen Daten (Amelang, et al. 2006). Rodgers (1998) stellte fest, dass die Zunahme der Mittelwerte eventuell durch eine geringere Streuung der neueren Tests ausgelöst

wurde. Wichters et al. (2004) brachten den Einwand, dass Testwerte wie der IQ sich in deren Zusammensetzung aus den verschiedenen Untertests verändern, was auch ein Grund für den Anstieg sein könnte.

Colom, Lluís-Font und Andrés-Pueyo (2005) vertraten die Hypothese, dass Intelligenzzuwächse hauptsächlich durch verbesserte Ernährung und kognitive Stimulation hervorgerufen werden. Sie testeten in den Jahren 1970 und 2000 zwei äquivalente Stichproben von Kindern. Bei den Testpersonen mit niedrigen Testwerten kam es ebenfalls zu einer Verminderung der Streuung. Eine bessere Ernährung kann nicht als einziger Grund für die Intelligenzsteigerung angesehen werden, da Effekte in afrikanischen Ländern ähnlich ausgefallen sind wie in den USA (Amelang et al. 2006). Lee (1951) beobachtete einen Anstieg des IQs von schwarzen Kindern, die von den Südstaaten in den Norden umzogen und begründete dies durch eine anregendere Umwelt. Wolf zeigte, dass diese Unterschiede darauf zurückzuführen sind, dass höher Leistungsmotivierte eher auswandern (1979, zitiert nach Amelang, et al. 2006).

Howard (1999) zeigte auf, dass die Zahl der Patente in den USA gestiegen ist und herausragende Schachspieler immer jünger werden. Diese Tatsachen sprechen dafür, dass die „wahre“ Intelligenz angestiegen ist und nicht nur die Testwerte der Messungen. Mingroni (2004) sieht „Heterosis“, das üppige Wachstum der Tochtergeneration als Grund für den Flynn-Effekt. Auch andere physikalische und psychologische Eigenschaften unterliefen ähnlichen Veränderungen wie der IQ. Dazu gehören z.B. die Körpergröße, Asthma und Autismus. Zuwächse in den Testwerten, die kristallisierte Intelligenz messen sind meistens geringer, als anhand der Hypothese für verbesserte Anregungsbedingungen vermutet wird (Amelang, 2006). Eine Ausnahme stellen Teasdale und Owen (2000) dar, die Verbesserungen sowohl in fluiden als auch in kristallisierter Intelligenz aufwiesen. Nettlebeck und Wilson (2004) untersuchten die Informationsverarbeitungsgeschwindigkeit (Inspection Time, IT) und die Intelligenz anhand des Peabody Picture Vocabulary Test (PPVT) zweimal an Kindern mit 6 bis 13 Jahren mit einem Abstand von 20 Jahren. Während der Mittelwert der IQs um durchschnittlich 5 Punkte anstieg, gab es keine Verbesserungen in der Geschwindigkeit der Informationsverarbeitung. Der

Flynn-Effekt kann also nicht durch die Geschwindigkeit der Informationsverarbeitung erklärt werden.

Sundet, Barlaug und Torjussen (2004) untersuchten einberufene Wehrpflichtige in Norwegen von 1950 bis 2002. Diese wiesen in Sprachtests, mathematischen Tests, Raven Matrizen Tests und in einem generellen Fähigkeitsscore bis in die Anfänge der 1970er Jahre Verbesserungen auf, danach flachte der Anstieg bis zum Jahr 1990 ab. Nach 1990 blieben die Werte konstant. Auch Teasdale und Owen (2000) berichten von Anstiegen ab 1950, die in den letzten zehn Jahren abflachten.

Leistungseinbußen resultieren zumeist aus Reizabschirmung (Dennis & Najarian, 1957) und als Folgen von Hirnschädigungen (vgl. z.B. Weinstein & Teuber, 1957). Bei klinischen Stichproben kann es jedoch auch durch Ermüdung, Ärger und Frustration eher zu Veränderungen der Testleistungen kommen (Rist & Dirksmeier, 2001).

Außerdem konnten Zusammenhänge zwischen Verarbeitungsmechanismen und Leistungsveränderungen festgestellt werden. Haan (1963, zitiert nach Amelang et al., 2006) entdeckte, dass Personen, die nach 25 Jahren die höchsten Intelligenzsteigerungen aufwiesen, ihre Probleme objektiv und realistisch bewältigten, während bei den anderen Personen Verdrängung und Rationalisierung häufiger auftrat.

Kausale Zusammenhänge, um Veränderungen in der Intelligenz hervorzurufen können mit den hier aufgezeigten Untersuchungen jedoch nicht getroffen werden (Amelang et al., 2006). Dafür wurden spezielle Programme für die Verbesserung der Intelligenz entwickelt. Ein berühmtes Programm ist das „Project Head Start“, das jedoch auch kritisiert wurde (Sommer & Sommer, 1983).

Nach einem Überblick über Ergebnisse von Längsschnittstudien kamen Weinert und Hany (2003) zu der Schlussfolgerung, dass interindividuelle Unterschiede bei mehr als der Hälfte der Testpersonen erhalten bleiben. Dies resultiert z.B. daraus, dass diejenigen, die bessere kognitive Fähigkeiten haben mehr von der Stimulation der Umwelt profitieren (vgl. auch Schröder, 2002). Für ein Drittel der Stichproben waren

intraindividuelle und interindividuelle Unterschiede in der kognitiven Entwicklung feststellbar (Hany, 2003). Diese Veränderungen traten in allen verschiedenen Intelligenzbereichen auf. Warum charakterisieren Stabilität und Veränderung das Leben des Menschen? Kognitive Entwicklung ist von einer Reihe an kognitiven, motivationalen Faktoren sowie Umweltfaktoren abhängig. Neuere Sichtweisen (Bidell & Fischer, 1997, zitiert nach Weinert & Hany, 2003) betrachten Genetik und Umwelt nicht als Faktoren, die unabhängig von der Entwicklung sind, sondern als aktive, selbst organisierende Systeme, die an anderen selbst organisierten Systemen teilhaben und die die Entwicklung der Intelligenz bestimmen. In diesem integrativen System ist die Entwicklung ein sich selbst organisierender Prozess.

1.2 Reliabilität

Hier soll der Begriff Reliabilität definiert werden. Danach wird auf die Messgenauigkeit im Sinn der Klassischen und der Probabilistischen Testtheorie eingegangen. Anschließend werden Retest-Reliabilitäten von Tests vorgestellt, die mit dem AID 2 vergleichbar sind.

1.2.1 Definition der Reliabilität

Die Reliabilität ist der statistische Kennwert, der den Grad der Messgenauigkeit und der Präzision des geprüften Merkmals bestimmt (Bortz und Döring, 2005). Sie beschreibt die formale Exaktheit der Merkmalserfassung und somit die Zuverlässigkeit von Testergebnissen (Bortz, 2005; Kubinger 2006; Wagner-Menghin, 2003). Das Maß der Zuverlässigkeit heißt Reliabilitätskoeffizient (Fröhlich, 2002). Die Genauigkeit der Erfassung der Merkmalsdimension vernachlässigt den Umstand, ob es sich dabei um die Merkmalsdimension handelt, die mit dem Test erfasst werden soll (Amelang & Schmidt-Atzert, 2006).

Der klassische Realitätsbegriff, auf den hier eingegangen wird, ist nur eine Art der Definition der Messgenauigkeit von Tests. Für Tests mit einer kategorialen Personenvariablen beispielsweise existiert keine einheitliche Definition der Messgenauigkeit (Rost, 2004).

Nach dem klassischen Realitätsbegriff besteht jeder Testwert (X) aus dem wahren Wert (T) und einem Fehler (E). Die Axiome der klassischen Testtheorie beschreiben die Eigenschaften eines Messfehlers (vgl. z.B. Gulliksen 1950, zitiert nach Rost, 2004). Die Reliabilität lässt sich wie in Formel 2 als ein Teil der wahren Varianz (s_T^2) an der beobachteten Varianz (s_X^2) definieren. Umso größer der Anteil der wahren Varianz ist, umso kleiner ist der Anteil der Fehlervarianz (s_E^2) der Testwerte. Wenn der Reliabilitätskoeffizient den größten Wert 1 erreicht, entspricht der Messwert dem wahren Wert ($X=T$). Wenn er den kleinsten Wert hat und demnach 0 beträgt, besteht der Messwert nur aus Fehlern ($X=E$). (Bortz & Döring, 2005; Steyer & Eid, 1993). Je kleiner der zu einem Messwert gehörende Fehleranteil ist, umso höher ist die Reliabilität. Ein Test mit perfekter Reliabilität müsste den wahren Wert ohne jeden

Fehler erfassen (Bortz & Döring, 2005). Unter gleichen Bedingungen mit demselben Test gewonnene Testwerte derselben Testperson sollten miteinander übereinstimmen (Kubinger 2006). In der Praxis treten jedoch Fehlereinflüsse z.B. durch Müdigkeit oder Raten auf (Bortz & Döring, 2005).

Um die Reliabilität für einen Test zu berechnen, ist für die empirisch ermittelbare Varianz der Testwerte (s_X^2) auch die wahre Varianz (s_T^2) zu schätzen (Bortz & Döring, 2005). Zur Schätzung der wahren Varianz gibt es in der klassischen Testtheorie drei Methoden, auf die in Punkt 1.2.2 eingegangen wird.

Formel 2: Allgemeine Berechnung der Reliabilität

$$REL = \frac{s_T^2}{s_X^2} = \frac{s_T^2}{s_T^2 + s_E^2}$$

Wie in Formel 3 kann nach der Berechnung der Reliabilität der Standardmessfehler ermittelt werden (Wagner-Menghin, 2003; Kubinger, 2006).

Formel 3: Standardmessfehler in der klassischen Testtheorie

$$SMF = \sqrt{s^2 \cdot (1 - r_{tt})}$$

Um die Breite des Konfidenzintervalls angeben zu können, in dem mit einer Wahrscheinlichkeit von 95% der „wahre“ Testwert einer Person liegt, der keine Messfehler enthält, wird wie in Formel 4 die Irrtumswahrscheinlichkeit einkalkuliert. Für den Fehler 1. Art mit einer Wahrscheinlichkeit von 5 Prozent ($\alpha = 0.05$) ist der kritische Wert in der Standard-normalverteilung $z_\alpha = 1.96$. Der wahrscheinliche Fehler ist demnach gerundet zweimal so groß wie der Standardmessfehler (Wagner-Menghin, 2003).

Formel 4: Konfidenzintervall für den wahren Testwert

$$T_v^{1,2} = x_v \pm z_\alpha \cdot SMF$$

1.2.2 Messgenauigkeit im Sinne der Klassischen Testtheorie

In der ursprünglichen Testkonstruktions-Lehre sind die korrelationsbasierten Ansätze stichprobenabhängig. Ein Nachteil der Klassischen Testtheorie ist, dass andere Personen die davor getestet wurden, die Genauigkeit des Testergebnisses mitbestimmen (Kubinger, 1996 a; Kubinger & Wurst, 2000). Die Reliabilität aus der Sicht der Klassischen Testtheorie ist dennoch für die Praxis relevant. Auch die Retest-Reliabilität und die Paralleltest-Reliabilität, die für Berechnung des

empirischen Teils dieser Diplomarbeit herangezogen wurden, gehören zu den Methoden der Klassischen Testtheorie. Deshalb sollen hier die in der klassischen Testtheorie verwendeten Methoden zur Bestimmung der Zuverlässigkeit eines Tests vorgestellt werden.

1.2.2.1 Retest-Reliabilität

Um die Retest-Reliabilität zu bestimmen, wird derselbe Test derselben Stichprobe zweimal vorgegeben. Das Intervall zwischen den beiden Messzeitpunkten (t_1 : erste Messung und t_2 : zweite Messung) kann dabei variiert werden. Die Retest-Reliabilität ist als Korrelation beider Messwertreihen definiert (siehe Formel 5). Die Korrelation mit 100 multipliziert, gibt an wie viel Prozent der Merkmalsvarianz auf „wahre“ Merkmalsunterschiede zurückzuführen sind. Die Differenz auf 100% gibt den Anteil der Fehlervarianz an (Bortz & Döring, 2005).

Formel 5 Berechnung der Retest Reliabilität

$$REL_{RETEST-Methode} = \frac{s_T^2}{s_X^2} = \frac{\text{cov}(t_1, t_2)}{s_{t_1} \cdot s_{t_2}} = r_{t_1 t_2}$$

Mittels Retest-Reliabilität können sowohl Aussagen über die Messgenauigkeit des Testinstruments und als auch über die zeitliche Stabilität der erhobenen Fähigkeiten gemacht werden. Beide Komponenten fließen untrennbar in den Kennwert des Koeffizienten ein, deshalb ist auch die Bezeichnung Stabilitätskoeffizient üblich (vgl. z.B. Rost, 2004; Bortz & Döring, 2005). Die Festlegung eines optimalen Zeitintervalls führt zu Schwierigkeiten (Amelang & Schmidt-Atzert, 2006). Während bei physikalischen Messungen eine kurzfristige Wiederholung möglich ist, kann es bei psychologischen Tests dabei zu Störungen durch Übungs- und Erinnerungseffekten kommen (Kubinger, 2006). Die Wahrscheinlichkeit von Erinnerungseffekten nimmt ab, wenn der zeitliche Abstand zwischen den Testvorgaben größer wird. Bei einem großen Zeitintervall können jedoch auch reale Schwankungen des erfassten Merkmals auftreten (Amelang & Schmidt-Atzert, 2006; Bortz & Döring, 2005). Über einen längeren Zeitraum wird die Langzeitstabilität gemessen, also inwiefern die Fähigkeit ein situationsübergreifendes Merkmal darstellt, das auch über die Zeit stabil ist (Kubinger, 1996a; Kubinger, 2006; Kubinger & Wurst, 2000).

Bei zweimaliger Vorgabe eines Intelligenztests im Abstand von einem Monat kommt es zu systematischen Mittelwertszunahmen von ca. der Hälfte einer Standardabweichung (Amthauer, 1957; Catron, 1978). Diese Übungsgewinne fallen jedoch nach Leistungsstand, Lernfähigkeit und Gedächtnis interindividuell verschieden aus (Amelang & Schmidt-Atzert, 2006).

Wenig brauchbar ist diese Methode für Tests, die instabile Merkmale, wie z.B. Stimmungen messen (Bortz & Döring, 2005). Je geringer die zeitliche Stabilität ist, desto kleiner ist auch die Korrelation zwischen beiden Messungen. Mit bestimmten Strukturgleichungs-Modellen, die latent-state-latent-trait Modell genannt werden, ist eine Separierung von Stabilität und Reliabilität jedoch möglich (Rost, 2004). Angemessen ist die Berechnung der Retest-Reliabilität bei reinen Speedtests und bei Persönlichkeitsfragebögen. Störungen fallen bei Intelligenztests ins Gewicht, wenn Skalen wenige Items beinhalten, wenn der Inhalt der Aufgaben eigentümlich ist und wenn das Retestintervall sehr kurz ist (Amelang & Schmidt-Atzert, 2006). Ein Nachteil der Retest-Reliabilität ist der hohe zeitliche und untersuchungstechnische Aufwand dieselben Testpersonen erneut zu kontaktieren und zur Teilnahme zu motivieren. Da mit großen Ausfallzahlen zu rechnen ist, muss bei der Untersuchungsplanung eine große „Start-Stichprobe“ gezogen werden. Das Problem der „Drop Outs“ ist jedoch auch, dass überwiegend Testpersonen mit schlechten Testergebnissen ausfallen. Dadurch wird die repräsentative Stichprobe verzerrt. Dieses Problem kann nicht gelöst werden (Bortz & Döring, 2005).

1.2.2.2 Paralleltest-Reliabilität

Um die Paralleltest-Reliabilität (Äquivalenz) zu bestimmen ist ebenfalls ein untersuchungstechnischer Aufwand notwendig. Es müssen zwei Testversionen entwickelt werden, die beide Operationalisierungen desselben Konstruktes, also gleichwertige Nachahmungen des Tests darstellen. Die Untersuchungsteilnehmer bearbeiten die beiden Tests in derselben Sitzung kurz hintereinander (Bortz & Döring, 2005). Manchmal wird auch ein kurzes Intervall von einigen Tagen für die Vorgabe gewählt. Die Hälfte der Probanden sollte Test A vor B, die andere B vor A bearbeiten (Amelang & Schmidt-Atzert, 2006). Je ähnlicher die beiden Testergebnisse sind, umso reliabler ist das Messinstrument. Die wahre Varianz wird

als Kovarianz zwischen den Testwerten einer Stichprobe in den zwei Paralleltests geschätzt (Bortz & Döring, 2005).

Dabei wird jedoch nicht nur die Konstanz der Messung an sich geprüft, sondern auch die Konstanz der Messung über die Zeit. Ein weiteres Problem ist die Konstruktion von zwei völlig gleichwertigen Paralleltests (Kubinger, 2006). Dafür muss zuerst ein Itempool erstellt werden. Danach wird dieser zur Itemanalyse einer Stichprobe vorgegeben. Wenn die Schwierigkeitsindizes und Trennschärfekoeffizienten berechnet wurden, können Item-Zwillinge zusammengestellt werden, die miteinander vergleichbare Schwierigkeiten und Trennschärfen aufweisen. Zur Erstellung der Paralleltests werden die Zwillinge zufällig der einen oder der anderen Testform zugeordnet. Wenn einer neuen Stichprobe beide Paralleltests A und B vorgegeben werden, kann die Reliabilität wie in Formel 6 berechnet werden (Bortz & Döring, 2005).

Formel 6: Paralleltest-Reliabilität

$$REL_{Paralleltest-Methode} = \frac{s_T^2}{s_X^2} = \frac{cov(t_A, t_B)}{s_{t_A} \cdot s_{t_B}} = r_{t_A} r_{t_B}$$

Durch die beschriebene Herangehensweise werden zwei so genannte „nominell parallele“ Tests erstellt. In „ τ -äquivalenten Tests“ entspricht der wahre Wert einer Person i im Test A (τ_{iA}) dem wahren Wert der Person i im Test B (τ_{iB}) (Bortz & Döring, 2005).

In der Probabilistischen Testtheorie ist die Konstruktion von Paralleltests eindeutig kontrollierbar. Wie in Punkt 1.2.3 ersichtlich werden in der Probabilistischen Testtheorie im Vergleich zur Klassischen Testtheorie jedoch methodisch völlig andere Wege zur Bestimmung der Zuverlässigkeit eines Testwerts gewählt (Kubinger, 2006).

1.2.2.3 Innere Konsistenz

Die Messgenauigkeit kann auch danach beurteilt werden wie sehr einzelne Teile des Tests dasselbe messen. Zur Feststellung der inneren Konsistenz gibt es verschiedene Methoden (Kubinger, 2006). Hier muss der Test einer Stichprobe nur einmalig vorgegeben werden (Bortz & Döring, 2005).

Bei der Testhalbierungsmethode oder (auch „Split-half“-Methode) wird ein Test in zwei Teile geteilt. Die resultierenden Testwerte beider Teile werden dann miteinander korreliert. Dieses Ergebnis entspricht der Paralleltest-Reliabilität eines Tests mit nur halber Testlänge und wird deshalb wie in Formel 7 nach Spearman und Brown aufgewertet (Lienert & Raatz, 1998).

Formel 7: Aufwertung der Split-Half Reliabilität nach Spearman Brown

$$r_{tt} = \frac{2r}{1+r}$$

Eine Abschätzung der unteren Grenze der Reliabilität liefert Cronbach-alpha, das bei Homogenität der Varianzen berechnet werden kann. Bei der Konsistenzanalyse wird ein Test in so viele Teile geteilt, wie er Items besitzt. Zur Berechnung kann dafür die Kuder-Richardson-Formel 20 herangezogen werden, die jedoch nur für dichotom zu verrechnende Items gilt (Lienert & Raatz, 1998).

1.2.3 Messgenauigkeit im Sinne der Probabilistischen Testtheorie

In der Probabilistische Testtheorie kann die Messgenauigkeit unabhängig von anderen zuvor getesteten Personen angegeben werden (Kubinger & Wurst, 2000). Die Messgenauigkeit ist in Abhängigkeit von der Größe des erzielten Testwertes unterschiedlich, da der Schätzfehler von der Fähigkeit der Person abhängt und vom Schwierigkeitsgrad der Items, die die Person bearbeitet hat. Deshalb sollte in einem nach dem Rasch-Modell konstruierten Test für jeden Testwert der Standardschätzfehler tabelliert sein (Wagner-Menghin, 2003). Der AID 2 wurde nach der Probabilistischen Testtheorie konstruiert und für die meisten Untertests gilt das dichotome logistische Modell von Rasch (Kubinger & Wurst, 2000). Deshalb soll hier kurz darauf eingegangen werden.

Wenn das dichotome logistische Testmodell von Rasch gilt, kann mit Formel 8 die Wahrscheinlichkeit berechnet werden, dass eine Person v_i mit der Fähigkeit ξ_{v_i} , Aufgabe i , mit der Schwierigkeit σ_i löst (Kubinger & Wurst, 2000; Wagner-Menghin, 2003).

Formel 8: Wahrscheinlichkeit für die Lösung eines Items

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

Wie in Formel 9 dargestellt kann der Standardschätzfehler der Fähigkeit für jede Person geschätzt werden, wenn ein Testergebnis gegeben ist. Dieses Testergebnis dient als Schätzung der Fähigkeit $\hat{\xi}_v$ der Person v. Die Funktion f beinhaltet, dass dies auch möglich ist, wenn pro Person k_v verschiedene Aufgaben vorgegeben wurden (Kubinger & Wurst, 2000; Kubinger, 1996b).

Formel 9: Standardschätzfehler

$$S(\xi_v) = \sqrt{\frac{1}{1(v)}} = \left[\sum_{j=f_1(v)}^{f_{k_v}(v)} \frac{e^{\xi_v - \sigma_j}}{1 + e^{\xi_v - \sigma_j}} \cdot \frac{1}{1 + e^{\xi_v - \sigma_j}} \right]^{\frac{1}{2}}$$

Mithilfe des Standardschätzfehlers kann wie in der klassischen Testtheorie ein Konfidenzintervall berechnet werden, das für die Zuverlässigkeit des Messwerts den Fehler 1. Art (α Fehler) berücksichtigt. In Formel 10 wird dargestellt, wie das Konfidenzintervall und somit der Messfehler berechnet werden kann (Kubinger & Wurst, 2000; Wagner-Menghin, 2003).

Formel 10: Berechnung des Konfidenzintervalls

$$\xi_v^{1,2} = \hat{\xi}_v \pm z_\alpha \cdot S\left(\hat{\xi}_v\right)$$

Wenn ein globaler Wert der Messgenauigkeit eines Tests benötigt wird, zum Beispiel für den Vergleich mit der Reliabilität eines Tests der Klassischen Testtheorie, kann der „index of person separation“ (IPS; Andrich, 1988, zitiert nach Wagner-Menghin, 2003) herangezogen werden. Durch diesen Index kann die innere Konsistenz interpretiert werden. Es wird der klassische Ansatz: $x = \tau + \varepsilon$ umgeformt zu $\hat{\xi}_v = \xi_v + \text{Schätzfehler}$, woraus Formel 11 folgt. Der Standardschätzfehler, der hier als SEM abgekürzt wird, entspricht dem $S(\xi_v)$ in Formel 9. Dieser Index ist jedoch stichprobenabhängig, da er mit der Varianz der Personenparameterschätzungen steigt (Wagner-Menghin, 2003).

Formel 11: index of person separation (IPS)

$$REL(\xi) = 1 - \frac{\frac{\sum_{i=1}^n SEM^2}{n}}{Var\left(\hat{\xi}\right)}$$

1.2.4 Reliabilitätsanforderungen

Wie hoch sollte die Reliabilität eines Testes sein, damit die Messgenauigkeit ausreicht, um in der Praxis Aussagen treffen zu können? Dazu sind in Abbildung 6 kurze Rechenbeispiele angeführt. Für einen Test mit der Reliabilität von $r_{tt}=0.80$ erstreckt sich das Konfidenzintervall über etwas mehr als 26 IQ-Punkte. Für einen Test mit der Reliabilität von 0.95 reduziert sich die Breite des Konfidenzintervalls jedoch bereits um ca. die Hälfte (Wagner-Menghin, 2003).

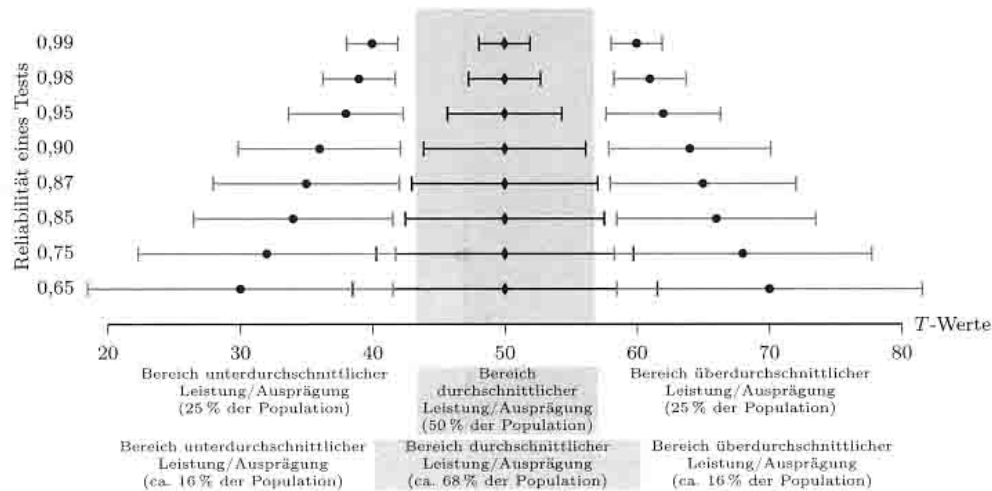
Abbildung 6: Berechnung des Konfidenzintervalls

$$\begin{aligned} SMF &= \sqrt{15^2 \cdot (1 - .80)} = 6.7 \\ T_v^{1,2} &= IQ_v \pm 1.96 \cdot 6.7 = IQ_v \pm 13.1 \\ SMF &= \sqrt{15^2 \cdot (1 - .95)} = 3.35 \\ T_v^{1,2} &= IQ_v \pm 1.96 \cdot 3.35 = 1Q_v \pm 6.6 \end{aligned}$$

Wenn eine Klassifizierung der Testleistung in „unter-“, „über-“ und „durchschnittlich“ ausreicht, kann berechnet werden, welche Höhe die Reliabilität des Tests haben muss, damit die Klassifizierung mit einer Wahrscheinlichkeit von 95% zutrifft (Wagner-Menghin, 2003).

In der Diagnostik werden meistens die mittleren 50% der Verteilung der Testwerte der Referenzpopulation als durchschnittlich angenommen. Wie auch in Abbildung 7 sichtbar, muss für diesen Fall zumindest eine Reliabilität von 0.87 gegeben sein, damit ein T-Wert von 50 mit 95% Sicherheit als durchschnittlich klassifiziert werden kann (Wagner-Menghin, 2003).

Abbildung 7: T-Werte und deren Konfidenzintervalle



Anmerkung: Aus *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens* (S.72) von K. D. Kubinger, 2006, Göttingen: Hogrefe.

In manchen Fällen wird als Durchschnittsbereich der Mittelwert $\pm$ der Standardabweichung herangezogen. In diesem Bereich liegen die mittleren 68% der Verteilung der Testwerte der Referenzpopulation. Für T-Werte ist dies z.B. der Bereich von 40 bis zu 60 T-Werten (vgl. z.B. Kubinger, 2006). Ausgehend von dieser Annahme ist bereits eine Reliabilität ab 0.75 ausreichend, damit ein T-Wert von 50 mit 95% Sicherheit als durchschnittlich klassifiziert werden kann (Wagner-Menghin, 2003).

Ab einer Reliabilität von 0.90 können Interpretationen getroffen werden ohne das Konfidenzintervall zu berücksichtigen. Hier kann vorausgesetzt werden, dass z.B. bei einem T-Wert von 50 auch der wahre Testwert mit einer Wahrscheinlichkeit von 95% im Durchschnitt liegt (Wagner-Menghin, 2003). Nach Horn (1986) sollte auf einen Test, der niedrigere Reliabilitätswerte als 0.9 hat, verzichtet werden (zitiert nach Kubinger, 2006). Weise (1975) stellt niedrigere Ansprüche. Nach ihm sollte ein Test eine Reliabilität von über 0.80 aufweisen, 0.8-0.9 seien als mittelmäßige Reliabilitätswerte anzusehen und Reliabilitäten ab 0.90 als hoch. Außerdem ist auf die Art der Reliabilitätsbestimmung zu achten. Hohe innere Konsistenz ist von jedem Test zu fordern. Bei hoher zeitlicher Variabilität eines zu erfassenden Merkmals ist auf die Paralltest-Reliabilität zu achten. Wenn stabile Merkmalsausprägungen

gemessen werden sollen, sollte großer Wert auf die Retest-Reliabilität gelegt werden (Bortz und Döring, 2005).

Für den Zusammenhang mit anderen Gütekriterien ist zu erwähnen, dass eine hohe Reliabilität eine Voraussetzung für eine hohe Validität eines Tests darstellt (Guthke, 1990). Die Reliabilität wird jedoch wiederum durch die Objektivität beeinflusst. Da Diskrepanzen zwischen Testanwendern Fehlervarianz erzeugen, beeinträchtigt eine geringe Objektivität die Reliabilität. Die Reliabilität kann demnach nur so hoch sein wie die Objektivität (Bortz und Döring, 2005).

Beim Adaptiven Testen steht die Testlänge im optimalen Verhältnis zur Messgenauigkeit. Es sollte im besten Fall so lange getestet werden, bis die gewünschte Messgenauigkeit erreicht ist. Dazu müssen genug Items vorhanden sein, die den Zuwachs der Messgenauigkeit ermöglichen. Dafür kann die Höhe des tolerierbaren Standardschätzfehlers aus dem IPS abgeleitet werden. Bei einer angenommenen Varianz der Personenparameter $Var(\hat{\xi})$ von 1.5 kann bei einem Standardschätzfehler von 0.47 eine relativ angemessene Reliabilität von 0.85 erreicht werden (Wagner-Menghin, 2003, S.285).

In der Praxis sollte immer das Konfidenzintervall für den beobachteten Testwert angegeben werden, da es andernfalls zu falschen Annahmen kommen kann. Vor allem bei Verlaufsstudien, Unterschieden von Bewerbungskandidaten und entscheidungsrelevanten Profildifferenzen ist die Messgenauigkeit der Testwerte besonders zu beachten (Wagner-Menghin, 2003).

1.2.5 Die Retest-Reliabilität in vergleichbaren Tests

Es scheint angemessen kurz darzustellen, welche Studien zur Retest-Reliabilität bzw. zur Stabilität andere Tests in deren Manualen anführen, die in dem gleichen Altersbereich bei Kindern und Jugendlichen eingesetzt werden, wie der AID 2. Hagmann-von Arx et al. (2008) liefen einen genauen Vergleich der Reliabilitäten der Intelligenztests AID 2, Hamburg Wechsler Intelligenztest für Kinder IV (HAWIK-IV), Intelligence and Development Scale (IDS), Kaufmann Assessment Battery for Children (K-ABC) und der nonverbalen Intelligenztests: Grundintelligenztest (CFT 1), Coloured Progressive Matrices (CPM), Snijders-Oomen non-verbaler

Intelligenztest (SON-R 2.5-7) und SON-R 5.5-17. Die Retest-Reliabilitäten liegen zwischen $r_{tt}=0.57$ und $r_{tt}=0.97$.

Hier soll genauer auf die Retest-Reliabilitäten des K-ABCs, des HAWIK-III und des HAWIK-IV eingegangen werden, da sie zum aktuellen Zeitpunkt die am häufigsten eingesetzten deutschsprachigen Verfahren darstellen, die mit dem AID 2 im Umfang und im umfassten Altersbereich der Testbatterie vergleichbar sind.

1.2.5.1 Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK)-III und HAWIK-IV

Mit dem Hamburg-Wechsler-Intelligenztest für Kinder – Dritte Auflage (HAWIK-III) kann die kognitive Entwicklung von Kindern und Jugendlichen in einem Altersbereich von 6 Jahren (6;0) bis 16 Jahren und 11 Monaten (6;11) untersucht werden (Tewes, Rossmann & Schallberger, 1999). Der HAWIK-IV ist die Neuauflage des HAWIK-III und ist in demselben Altersbereich einsetzbar (Petermann & Petermann, 2007). Der HAWIK-IV besteht aus 15 Untertests. Zehn Untertests werden als Kerntests bezeichnet, fünf als Optionaltests. Als Testkennwerte werden ein Gesamt-IQ und vier Indizes berechnet (Daseking, Petermann & Petermann, 2007; Petermann & Petermann, 2008).

Im Manual zur deutschsprachigen Version des HAWIK-III liegen keine Studien über die Retest-Reliabilität der deutschen Version vor. Es wird lediglich über eine Untersuchung zur englischsprachigen Version berichtet. Die Wechsler Intelligenz Scale for Children – Third Edition von David Wechsler (WISC III) wurde den Testpersonen nach zwei bis drei Wochen erneut vorgegeben. Die Korrelationen der Untertestwerte lagen im Bereich von $r_{tt}=0.60$ bis $r_{tt}=0.82$. Für den Gesamt-IQ betrug die Korrelation $r_{tt}=0.9$. Für einen längeren Zeitraum liegen keine Ergebnisse vor. Dieser Mangel an Untersuchungen wird dadurch begründet, dass die Werte von Retest-Messungen nach einer geringen Zeitspanne durch Lerneffekte verfälscht würden und bei einem längeren Intervall durch Entwicklungen der Testpersonen beeinflusst werden (Tewes, et al. 1999, S.91).

Im Manual des HAWIK-IV (Petermann & Petermann, 2007, S.31) wird darauf hingewiesen, dass in der Praxis häufig dasselbe Kind öfters getestet wird. Es kann jedoch keine Zeitspanne für den HAWIK-IV angegeben werden, nach der

Lerneffekte bei der Zweittestung ausgeschlossen werden können. Untersuchungen, die mit früheren Versionen der Wechsler Skalen durchgeführt wurden, belegen, dass nach ein bis zwei Jahren fast keine Lerneffekte mehr zu verzeichnen sind. Bei sprachlichen Untertests ist dies nach einem Jahr der Fall (Canivez & Watkins, 1998, 2001, zitiert nach Petermann & Petermann, 2007; McCaffrey, Duff & Westerwelt, 2000, zitiert nach Petermann & Petermann, 2007; Rapport, Brines, Axelrod & Theisen, 1997, zitiert nach Petermann & Petermann, 2007). Vor allem bei den Untertests der Indizes „Arbeitsgedächtnis“ und „Wahrnehmungsgebundenes Logisches Denken“ waren Lerneffekte feststellbar. Es wird deshalb im Manual dazu geraten, nach einer kurzen Zeitspanne optionale Untertests einzusetzen. Außerdem sollten diese Einflussfaktoren bei einer Wiederholungstestung berücksichtigt werden (Petermann & Petermann, 2007).

Für die Untertests Zahlen-Symbol-Test, Symbol-Suche und Durchstreich-Test konnte der Split-Half-Koeffizient die Reliabilität nicht adäquat schätzen, da dies reine Geschwindigkeitstests sind (Petermann & Petermann, 2007). Für diese Untertests wurden daher Stabilitätskoeffizienten herangezogen (Allen & Yen, 1979, zitiert nach Petermann & Petermann, 2007; Magnusson, 1967, zitiert nach Petermann & Petermann, 2007). Es wurden 103 Kinder im Alter von 6 bis 16 Jahren mit einem durchschnittlichen Abstand von 18 Tagen zweimal getestet (Petermann & Petermann, 2007, S. 114). Die Stabilitätskoeffizienten der drei Untertests werden in Tabelle 1 dargestellt. Sie liegen zwischen .76 und .85 (Petermann & Petermann, 2007, S. 115; Hagmann-von Arx et. al., S. 237

Tabelle 1: Retest-Reliabilität des HAWIK-IV für 3 Subtests

Subtests	Korrelationskoeffizienten
Zahlen-Symbol Test	.85
Symbol-Suche	.79
Durchstreich-Test	.76

Anmerkung: Stichprobe: N=103; Alter 6-16 Jahre

Daraus kann geschlossen werden, dass die meisten Untertests des HAWIK-IV nicht für Wiederholungstestungen geeignet sind. Petermann und Petermann (2007) schlagen vor, die Einflussfaktoren zu berücksichtigen. Es ist jedoch unmöglich für die Testleiter die Lerneffekte zu berücksichtigen, wenn diese nicht genau berechnet

wurden. Kritisch ist auch anzumerken, dass die Studien über die Retest-Reliabilität der drei Untertests bereits 30 bzw. 40 Jahre zurückliegen. Die Sinnhaftigkeit einen Test vorzugeben, für den die Retest-Reliabilitäten nicht bekannt sind, ist anzuzweifeln. Es ist schwer vertretbar, Entscheidungen für die Zukunft eines Kindes zu treffen, wenn die Stabilitäten der Testwerte nicht ausgiebig untersucht wurden. Die Entwicklung einer Parallelfarm wäre sinnvoll und wünschenswert, um zumindest die Stabilitäten nach kurzen Zeitabständen bestimmen zu können.

1.2.5.2 Kaufmann-Assessment Battery for Children (K-ABC)

Die Kaufmann-Assessment Battery for Children (K-ABC) ist für Kinder im Alter von 2 Jahren und 6 Monaten (2;6) bis zu einem Alter von 12 Jahren und 5 Monaten (12;5) normiert (Melchers & Preuß, 1994). Sie umfasst also ebenso wie der AID 2 eine Altersspanne von 10 Jahren, ist jedoch für jüngere Kinder entwickelt worden (Petermann & Petermann, 2008).

Madest (1989, zitiert nach Melchers & Preuß, 1994) führte eine Untersuchung der Retest-Reliabilität an 24 Kindern im Alter von 9;0 bis 12;5 Jahren durch. Zwischen den beiden Testungen lagen durchschnittlich 19 Tage. Genauere Werte können Tabelle 2 entnommen werden.

Melchers und Preuß (1994) weisen darauf hin, dass die Ergebnisse von Madest auf einer kleinen Stichprobe beruhen. Sie erwähnen deshalb noch eine andere Untersuchung von Kaufmann und Kaufmann (1983, zitiert nach Melchers und Preuß, 1994) mit einer größeren Stichprobe von 246 Kindern, von denen 70 Kinder aus der Altersstufe von neun Jahren (9;0) bis zwölf Jahren und fünf Monaten (12;5) waren. Im Rahmen dieser Studie wurden vergleichbare Korrelationskoeffizienten berechnet.

Hier ist kritisch anzumerken, dass keine aktuellen Studien vorliegen und dass die Langzeitstabilität nicht untersucht wurde. Melchers und Preuß (1994) maßen sich nicht an mit der Kaufmann Assessment Battery for Children die genetisch vorhandenen Fähigkeiten eines Menschen messen zu können, da die Intelligenzentwicklung aus einem Zusammenspiel der genetischen Veranlagung und der von außen einwirkenden Umwelt entsteht. Sie sind der Meinung, dass man diese beiden Einflussgrößen nicht auseinander halten kann, insbesondere wenn die

Wechselwirkungen berücksichtigt werden sollten. Die Skala intellektueller Fähigkeiten wird nur als gegenwärtiges Maß gesehen, das durch Umwelteinflüsse, wie z.B. Fördermaßnahmen verändert werden kann.

Tabelle 2: Testwiederholungsreliabilität des K-ABC für die Altersstufen 9;0 – 12;5, N = 24

Untertests und Skalen	Korrelationskoeffizient
3. Handbewegungen	.57
4. Gestaltschließen	.57
5. Zahlennachsprechen	.64
6. Dreiecke	.63
7. Wortreihe	.68
8. Bildhaftes Ergänzen	.69
9. Rätselhaftes Gedächtnis	.70
10. Fotoserie	.61
12. Gesichter und Orte	.96
13. Rechnen	.90
14. Rätsel	.82
16. Lesen/Verstehen	.95
Skala einzelheitliches Denken	.84
Skala ganzheitliches Denken	.87
Skala intellektueller Fähigkeiten	.88
Fertigkeitenskala	.97

Anmerkung: Madest (1989, zitiert nach Melchers & Preuß, 1994)
aus *Kaufmann-Assessment Battery for Children (K-ABC)*, Deutsche Fassung (2. korrigierte und ergänzte Aufl.) (S.94) von P. Melchers & U. Preuß, Frankfurt am Main: Swets & Zeitlinger.

1.3 Das Adaptive Intelligenz Diagnostikum 2 (AID 2)

Im Folgenden werden die wesentlichen Grundzüge des AID 2 (Kubinger & Wurst, 2000) vorgestellt. Die Gütekriterien, insbesondere die Reliabilität, werden genauer beleuchtet und die Subtests werden einzeln vorgestellt.

1.3.1 Allgemeine Beschreibung der Testbatterie

Der AID 2 (Kubinger & Wurst, 2000) ist eine Testbatterie für Kinder und Jugendliche zur Erfassung komplexer und grundlegender Kognitionen. Er ist eine Weiterentwicklung des AID (Kubinger & Wurst, 1991). Dieser Individualtest kann Kindern ab einem Alter von 6 Jahren (6;0) und Jugendlichen bis zu einem Alter von 15 Jahren und 11 Monaten (15;11) und unter Umständen auch in einem höheren Altersbereich vorgegeben werden (Antretter, 2001). Inhaltlich ist der AID 2 am Konzept von David Wechsler orientiert (vgl. z.B. Petermann & Petermann, 2007). Die Untertests des AID 2 sind dem des HAWIK thematisch ähnlich (Kubinger & Wurst, 2000).

Im AID 2 wird das adaptive Testen der probabilistischen Testtheorie angewandt (Kubinger, 1989, 1996b). Durch die Technik des branched testing ist der AID 2 dem HAWIK-IV testökonomisch überlegen. Die Dauer der Durchführung ist geringer und die Messgenauigkeit ist höher und durch die Anpassung an das Leistungsniveau kann die Motivation der Testperson gut aufrechterhalten werden (Jacobs, Heubrock & Petermann 2003).

Vom Testleiter werden das Arbeits- und das Kontaktverhalten der Kinder qualitativ beurteilt. Die Testkennwerte der elf Untertests und der drei Zusatztests werden in T-Werten angegeben. Für die acht Untertests: **1, 3, 4, 6, 8, 9, 10** und **11** und den Zusatztest **10a** werden zusätzlich altersunabhängige Fähigkeitsparameter ξ_v angegeben. Außerdem werden die Testkennwerte untere Grenze der Intelligenzqualität, zweitniedrigste Untertestleistung und Range der „Intelligenz“ in Prozenträngen (PR) angegeben. Als Range wird die Differenz zwischen dem maximalen und minimalen T-Wert bezeichnet (Kubinger & Wurst, 2000).

Durch die Testvorgabe kann ein Screening in Bezug auf ausgewählte Teilleistungsstörungen vollzogen werden (vgl. auch Leiss, 2003). Das Testkonzept ermöglicht eine Profilinterpretation vieler verschiedener Fähigkeiten. Die Einsatzmöglichkeiten des AID 2 sind vielfältig. Er wird in der Schulpsychologie, in der Berufs- und Bildungsberatung, in der Klinischen Psychologie und auch bei fremdsprachigen Testpersonen eingesetzt. Zur Abklärung von Hochbegabung ist der AID 2 ebenfalls gut geeignet (vgl. z.B.: Holocher-Ertl, Kubinger & Hohensinn, 2008; Kubinger, 1997; Schlagheck & Petermann, 2006)

1.3.2 Die Reliabilität des AID und des AID 2

Gütekriterien garantieren die Qualität psychologisch-diagnostischer Verfahren (Kubinger, 2003b, S. 195). Genaue Informationen zu den Gütekriterien Objektivität, Validität, Skalierung, Normierung, Ökonomie, Zumutbarkeit, Fairness, Unverfälschbarkeit und Nützlichkeit des AID 2 können vielfach nachgelesen werden (vgl. z.B.: Jacobs et al., 2003; Kubinger 2003a; Kubinger, 2006; Kubinger & Wurst, 2000). Auf das Gütekriterium der Reliabilität soll hier genauer eingegangen werden. Insbesondere werden Studien zur Stabilität des AID 2 vorgestellt.

Tabelle 3: Standardschätzfehler und split-half-Reliabilität des AID 2

	Form	AID 2			AID split- half- Rel.
		Standardschätzfehler minimal/maximal		Mess- fehler	
		Kurz	Parallel	Standard	
1 Alltagswissen		0.68/1.48	0.61/1.93	0.55/1.47	.95
2 Realitätssicherheit				0.85/1.25	.70
3 Angewandtes Rechnen		0.71/1.16	0.62/1.74	0.57/1.60	.95
4 Soziale und Sachliche Folgerichtigkeit			1.18/1.38	0.89/1.58	.91
6 Synonyme Finden		0.68/1.18	0.60/1.36	0.55/1.14	.94
8 Antizipieren und Kombinieren ^{-figural}				1.03/1.80	.81
9 Funktionen Abstrahieren		0.67/1.12	0.55/1.14	0.54/1.10	.93
10 Analysieren und Synthetisieren ^{-abstrakt}			0.90/1.33	0.83/1.26	.95
11 Soziales Erfassen und Sachliches Reflektieren		0.65/1.18	0.57/1.15	0.54/1.16	.94
5a Unmittelbares Reproduzieren ^{-figural/abstrakt}				0.85/1.24	
5b Merken und Einprägen				0.71/1.09	
10a Strukturieren ^{-visumotorisch}				0.67/2.86	

Anmerkung: aus *Adaptives Intelligenz Diagnostikum 2 (AID 2)* (S.22-23) von K. D. Kubinger & E. Wurst, 2000, Goettingen: Beltz Test.

Der minimale Messfehler wird für die einseitige Irrtumswahrscheinlichkeit von $\alpha=0.05$ angegeben: split-half Rel. steht für split-half Reliabilität

Für den AID 2 gilt zumeist das dichotome logistische Testmodell von Rasch. Auf die Berechnung des Standardschätzfehlers und des Messfehlers wird unter Punkt 1.2.3

näher eingegangen. In Tabelle 3 sind die Standardschätzfehler und die minimalen Messfehler für den AID 2 tabelliert und auch die Split-half-Reliabilität für den AID wird angeführt. Die Standardmessfehler sind in den Einheiten der Fähigkeitsparameter ξ_v angegeben. Die Fähigkeitsparameter können in T-Werte umgewandelt werden. Beispielsweise bedeutet der minimale Standardschätzfehler von 0.55 im Untertest **1 Alltagswissen**, dass sich das Konfidenzintervall, in dem der wahre Fähigkeitsparameter mit einer Wahrscheinlichkeit von 95 Prozent liegt über 3 bis 4 T-Werte erstreckt. Im Sinne der klassischen Testtheorie wurden für den AID auch Split-half-Reliabilitäten berechnet (Kubinger & Wurst, 2000, S.22).

Ebenfalls im Sinne der klassischen Testtheorie wurde für den AID die Retest-Reliabilität bestimmt. Die Korrelationskoeffizienten sind Tabelle 4 zu entnehmen. Zur Bestimmung der Retest-Reliabilität des AID nach vier Wochen wurden 148 Testpersonen der Normierungsstichprobe ein zweites Mal getestet. Dabei konnte die Stabilität sowohl für die Standardform als auch für die Parallelforn berechnet werden. Auffallend ist, dass die Korrelationskoeffizienten für die standardmäßige adaptive Vorgabe durchwegs höher sind außer für Untertest **9 Funktionen Abstrahieren**, bei dem für die Parallelforn höhere Ergebnisse erzielt wurden und für Untertest **11 Soziales Erfassen und Sachliches Reflektieren** sind die Korrelationskoeffizienten der Standardform und der Parallelforn von gleicher Höhe. Diese Unterschiede wurden jedoch nicht auf Signifikanz überprüft (Kubinger & Wurst, 2000, S.23).

Knab (1996) untersuchte die Langzeitstabilität des AID. Dabei wurden 112 Testpersonen wiederholt getestet. Diese waren bereits zum einen an der Universitätsklinik an der Abteilung für Heilpädagogik und Psychosomatik und im Rahmen eines Forschungsprogramms für angeborene Stoffwechselanomalien, zum anderen am Institut für Psychologie der Universität Wien an der Abteilung für Entwicklungspsychologie und Pädagogische Psychologie getestet worden. Die Zeitspanne zwischen den beiden Testungen betrug dabei mindestens ein Jahr und im Durchschnitt drei Jahre. Die Korrelationskoeffizienten wurden nach unterschiedlichen Gesichtspunkten berechnet: für die Gesamtstichprobe, nach Institutionen getrennt, für Testpersonen mit einer Teilleistungsschwäche, für

Testpersonen mit kumulierenden Risikofaktoren, für jüngere und ältere Kinder und nach unterschiedlich langem Zeitintervall zwischen 1. und 2. Testung.

In Tabelle 4 sind die Korrelationskoeffizienten für die Gesamtstichprobe angegeben. Mittels Elternfragebogen wurde der Grad der Belastung in der Umwelt der Kinder und der Jugendlichen abgefragt. Ein wichtiges Ergebnis der Untersuchung war, dass für die Kinder und Jugendlichen, die keiner kumulierenden psychosozialen oder biosozialen Belastung ausgesetzt waren auch für große Zeitintervalle höhere Korrelationskoeffizienten berechnet wurden als für alle anderen Gruppen. Die Differenzen der Koeffizienten zwischen den Gruppen waren jedoch so gering, dass kein deutlicher Unterschied zwischen den unterschiedlichen Gruppen im Entwicklungsverlauf der Intelligenz festgestellt werden konnte.

Tabelle 4: Stabilität des AID

Testkennwerte	Korrelationskoeffizienten der T-Werte		
	nach 4 Wochen		nach mehr als einem Jahr
	Standard-form	Parallel-form	Standard-form
Subtests			
1 Alltagswissen	.95	.92	.7790
2 Realitätssicherheit	.75	.59	.3927
3 Angewandtes Rechnen	.94	.90	.7895
4 Soziale und Sachliche Folgerichtigkeit	.82	.59	.5629
5 Unmittelbares Reproduzieren- numerisch „vorwärts“	.71		.6223
5 Unmittelbares Reproduzieren- numerisch „rückwärts“	.67		.5275
6 Synonyme Finden	.85	.84	.7616
7 Kodieren und Assoziieren: Kodiermenge	.83		.6618
7 Kodieren und Assoziieren: Assoziationen	.77		.5846
8 Antizipieren und Kombinieren- figural	.77		.6049
9 Funktionen Abstrahieren	.83	.86	.7061
10 Analysieren und Synthetisieren	.91	.85	.7045
11 Soziales Erfassen und Sachliches Reflektieren	.90	.90	.8000
Gesamt IQ			.8736

Anmerkung:

aus *Langzeitstabilität des AID* (S.50) von C. Knab, 1996, unveröffentlichte Diplomarbeit, Wien.

aus *Adaptives Intelligenz Diagnostikum 2 (AID 2)* (S.23) von K. D. Kubinger & E. Wurst, 2000, Goettingen: Beltz Test.

Studien zur Langzeitstabilität des AID 2 liegen noch keine vor. Die Stabilität des AID 2 nach vier Wochen wurde jedoch bereits für einige Subtests überprüft (Milanovic, 1998). Die Korrelationskoeffizienten können Tabelle 5 entnommen werden. Es wurden 128 Kinder und Jugendliche aus Österreich nach einem Intervall

von ungefähr 4 Wochen ein zweites Mal getestet. In Tabelle 5 sind jedoch schon die neuen Bezeichnungen wiedergegeben. Die Retest-Reliabilitäten der Untertests **5 Unmittelbares Reproduzieren-_{numerisch}**, **7 Kodieren und Assoziieren** und **10a Strukturieren-_{visumotorisch}** weisen Korrelationskoeffizienten über $r_{tt}=0.6$ auf. Der Untertest **5a Unmittelbares Reproduzieren-_{figural/abstrakt}** zeigt jedoch als einziger eine niedrigere Stabilität auf.

Die Korrelationskoeffizienten der meisten Untertests sind zwar auch nach einem Jahr noch relativ hoch. Aussagen, die eine ferne Zukunft betreffen, müssen jedoch mit Vorsicht getroffen werden (Kubinger, 2006).

Tabelle 5: Retestreliabilität des AID 2 nach 4 Wochen; N=128

Testkennwerte		Korrelationskoeffizienten
Subtests		Rohwerte
5 Unmittelbares Reproduzieren vorwärts*		.813
5 Unmittelbares Reproduzieren rückwärts*		.736
7 Kodieren	10 Sekunden	.671
	1 Minute	.845
	2 Minuten	.886
7 Assoziationen		.636
5a Unmittelbares Reproduzieren- _{figural/abstrakt}	Gesamttest	.775
	Konkrete Bilder	.626
	Abstrakte Bilder	.607
5b Merken und Einprägen	Item 1	.491
	Item 2	.573
10a Strukturieren- _{visumotorisch}		.788

Anmerkung: aus *Erfassung der Teilkomponente Serialität im AID 2* (S.68) von M. Milanovic, 1998, unveröffentlichte Diplomarbeit, Wien.

Zusammenfassend können die Testwerte des AID und des AID 2 vor allem nach einem Intervall von ungefähr vier Wochen in der standardmäßigen adaptiven Vorgabe durchaus als stabil bezeichnet werden. Einige Untertests weisen mit Korrelationskoeffizienten zwischen $r_{tt}=0.7$ und $r_{tt}=0.8$ auch eine erstaunlich hohe Langzeitstabilität nach einem Intervall von durchschnittlich drei Jahren auf (Kubinger & Wurst, 2000). Vergleichsmöglichkeiten mit anderen Testbatterien sind spärlich. Die wenigen Ergebnisse des K-ABC und des HAWIK-IV, auf die unter Punkt 1.2.5 eingegangen wird, weisen jedoch ähnlich hohe Korrelationskoeffizienten auf. Ein großer Vorteil bei der Berechnung der Retest-Reliabilität des AID 2 ergibt sich durch das adaptive Testen. Dadurch wird gewährleistet, dass Testpersonen nach einem längeren Zeitintervall bei der Zweittestung andere Items zu lösen haben, als

bei der Ersttestung. Dies ermöglicht die Berechnung der Stabilitätskoeffizienten mit weitgehend ausgeschalteten Übungs- und Erinnerungseffekten. Ein weiterer Vorteil gegenüber vergleichbaren Intelligenztestbatterien ist die Möglichkeit Parallelfornen vorzugeben. So können auch nach kurzen Zeitabständen Stabilitäten berechnet werden.

1.3.3 Die Subtests des AID 2

Der AID 2 besteht aus 14 Subtests. Die Subtests **1, 3, 5, 6, 9** und **11** erfassen verbal-akustische und die Subtests **2, 4, 7, 8** und **10** manuell-visuelle Fähigkeiten. Die drei Subtests **5a, 5b** und **10a** werden als „Zusatztests“ bezeichnet und können bei spezifischen Fragestellungen vorgegeben werden. Der AID 2 hat für die Untertests **1, 3, 4, 6, 8, 9, 10** und **11** eine „adaptive“ Aufgabenvorgabe. Dabei erfolgt das Vorgabeschema nach dem „branched-testing“. Bei den restlichen sechs Untertests erfolgt die Vorgabe der Items konventionell. Für die fünf Untertests **2, 4, 7, 8, 10** und die zwei Zusatztests **5a** und **10a** gibt es auch zusätzlich eine sprachfreie Instruktion. Parallelfornen gibt es für die Untertests **1, 3, 4, 6, 9, 10** und **11** (Kubinger & Wurst, 2000).

Die Untertests **1, 3, 6, 9** und **11** messen eindimensional und es sind 15 der 60 Aufgaben vorzugeben. Die Testleistungen werden als „richtig“ oder „falsch“ verrechnet, teilrichtige Antworten werden nicht berücksichtigt (Kubinger & Wurst, 2000).

Nachfolgend werden die elf Subtests und drei Zusatztests des AID 2 vorgestellt. Es wird jeweils die zu messende Fähigkeit genannt, Besonderheiten werden hervorgehoben, die zu bearbeitenden Aufgaben beschrieben und nach Möglichkeit wird ein Itembeispiel genannt (Kubinger & Wurst, 2000).

- 1 Alltagswissen:** In diesem Untertest soll die Fähigkeit geprüft werden, sich Sachkenntnisse über heute alltägliche Inhalte anzueignen. Es handelt sich dabei um unterschiedliche, repräsentative Wissensgebiete. Mündlich gestellte Fragen sollen mündlich beantwortet werden.

Item Beispiel: Manche Kriege werden als Weltkriege bezeichnet. Wie viele

Weltkriege hat es gegeben?

richtige Antwort: Zwei.

- 2 **Realitätssicherheit:** Es soll das Verständnis um die Wirklichkeit der Dinge im Alltag gemessen werden. Die zwölf eindimensional messenden Aufgaben werden in der Reihe ihrer Schwierigkeit vorgegeben. Die erste Aufgabe dient zur Aufwärmung. Die Verrechnung erfolgt zweikategoriell. Die Testperson soll auf Bildern ein fehlendes Detail erkennen und benennen oder darauf zeigen, z.B. sind an einem Haus die fehlenden Fenster zu entdecken.

- 3 **Angewandtes Rechnen:** Es soll möglichst unabhängig von schulischen Rechenfertigkeiten geprüft werden, wie gut Personen bei der Problemstellung alltäglicher Aufgaben durch das Ziehen von Schlüssen passende Rechenoperationen anwenden. Textbeispiele werden vorgelesen und bei den schwierigeren Aufgaben wird es der Testperson freigestellt in einem Textheft mitzulesen.

Item Beispiel: Lotte hat 10 Bücher, Karl hat 11. Wie viele hat Karl mehr als Lotte?

richtige Antwort: 1.

- 4 **Soziale und Sachliche Folgerichtigkeit:** Erfasst werden soll die Fähigkeit, soziales Geschehen bzw. alltägliche Sachverhalte in ihrer Abfolge zu verstehen und zu kontrollieren. Es werden sechs bis sieben von acht eindimensionalen Aufgaben vorgegeben. Die Verrechnung der Leistungen erfolgt zu den Testwerten zweikategoriell. Die Testperson soll Bilderfolgen von verschiedenen Geschichten logisch richtig ordnen. Eine Geschichte zeigt z.B.: wie ein Bub einen Turm aus Bauklötzen baut.

- 5 **Unmittelbares Reproduzieren**-numerisch: Gemessen werden soll die Kapazität der Informationsverarbeitung von Reihen im verbal-akustischen Bereich. Die Testauswertung erfolgt separat mehrkategoriell je nach Leistungsgüte. Dieser Untertest besteht aus Zahlenreihen von zwei bis zu neun Ziffern. Für jede Länge gibt es je drei Zahlenreihen. Diese werden nacheinander von der kürzesten bis zur längsten vorgesagt und die Reihe soll komplett

wiedergegeben werden. Nach drei gescheiterten Versuchen wird die Vorgabe abgebrochen. Zuerst wird ein Aufgabenset „vorwärts“ vorgegeben, danach ein anderes „rückwärts“.

Item Beispiel: 8-1-5

- 6 Synonyme Finden:** Es soll sowohl das elementare Sprachverständnis geprüft werden, als auch die Erfassung der Bedeutung sprachgebundener Begriffe und der Wortschatz. Wörter werden mündlich dargeboten. Für diese Wörter sind alternative Begriffe anzugeben. Auch umgangssprachliche Antworten sind als richtige Lösungen zu bewerten, sofern die Bedeutung dieselbe ist.

Item Beispiel: klopfen

richtige Antworten: pochen, hämmern, pumpern

- 7 Kodieren und Assoziieren:** Es soll sowohl die Informationsverarbeitungsgeschwindigkeit als auch die Fähigkeit zum beiläufigen Lernen gemessen werden. Zwölf Objekte werden in zufälliger Reihenfolge auf einem zweiseitigen Arbeitsblatt angeführt. Von der Testperson sind nach einer Vorlage Symbole zu den Objekten passend abzuzeichnen, danach ohne Vorlage zu zeichnen. Die Testauswertung erfolgt nach Anzahl der kodierten Symbole in der Zeiteinheit und nach der Anzahl der aus dem Gedächtnis kodierten Symbole.

Beispiel für ein Symbol: Eine Birne wird z.B. mit dem Symbol $\perp$ kodiert

- 8 Antizipieren und Kombinieren:** Schlussfolgerndes Denken soll geprüft werden, dadurch dass Teile eines Ganzen zu erkennen sind und das Ganze gestaltet werden soll. Der Testperson werden fünf bzw. sieben von elf Aufgaben vorgegeben. Der Test misst eindimensional. Die Testauswertung ist bei allen Aufgaben, außer bei zweien dreikategoriell. Die drei Kategorien sind „richtig und schnell“, „richtig, aber langsam“ und „falsch“. Die Testperson soll Teile einer Figur, ähnlich wie bei einem Puzzle zusammensetzen. Dabei wird auf einen „Anker“ hingewiesen, der in der richtigen Position vorgelegt wird und an den alle anderen Teile angrenzen.

Beispiel für eine Figur: Teddybär

- 9 Funktionen Abstrahieren:** Es soll gemessen werden wie gut durch Verallgemeinerungen zu einer Begriffsbildung gelangt werden kann. Die Testperson soll gemeinsame wesentliche Funktionen von je zwei Objekten nennen. Dabei sind auch umgangssprachliche Formulierungen erlaubt.

Item Beispiel: Korb - Tasche

Antwort: tragen; hineingeben

- 10 Analysieren und Synthetisieren**-abstrakt: Es soll geprüft werden wie gut Gestalten durch geeignete Gliederung wiedergegeben werden können. Die Testperson soll mit Würfeln, deren Seiten unterschiedliche Muster zeigen ein großes Muster nachlegen. Die blauen Seiten sind für die Lösung nicht zu berücksichtigen. Die anderen Seiten sind rot, weiß, diagonal und horizontal geteilt weiß und rot und eine Seite ist weiß mit einem roten Quadrat. Es soll z.B. ein Quadrat aus vier roten Seiten gelegt werden.

- 11 Soziales Erfassen und Sachliches Reflektieren:** Es soll gemessen werden inwiefern eine Person in die Gesellschaft eingeordnet ist und über sozial angepasstes Verhalten bescheid weiß. Die Testperson hat mündlich gestellte Fragen zu beantworten.

Item Beispiel: Warum ist es gut Sonnencreme zu benutzen?

Antwort: Zum Schutz vor Sonnenbrand

- 5a Unmittelbares Reproduzieren:** Es soll das Aufnahmevermögen der Informationsverarbeitung von Reihen im visumotorischen Bereich gemessen werden. Der Test misst eindimensional. Die Verrechnung der Testleistung zu den Testwerten erfolgt zweikategoriell. Auf einer Tafel befinden sich 49 Bilder von Figuren und Objekten. Es werden Bilder auf dieser Tafel in einer gewissen Reihenfolge angetippt. Dies soll die Testperson in der selben Folge nachmachen. Die Folgen bestehen aus drei bis zu neun Bildern. Für jede Länge gibt es zwei Folgen: eine mit anschaulichen Objekten und eine mit abstrakten Figuren. Wenn die Testperson bei beiden Folgen derselben Länge scheitert und auch bei beiden darauf folgenden, wird die Vorgabe abgebrochen.

Bei einer Folge wird z.B. auf den Apfel, das Telefon und die Gießkanne getippt.

5b Merken und Einprägen: Es soll die Fassungskraft gemessen werden, die nach einer Reizdarbietung und nach einer einmaligen Wiederholung erreichbar ist. Die Testperson soll Worte von sinnfreien Silben nachsprechen. Neun Silben sind zweimal in unterschiedlicher Reihenfolge in Listen angeordnet. Verrechnet wird die Testleistung nach Vorgabe der zweiten Wortliste zweikategoriell. Ein Beispiel für eine sinnlose Silbe ist „nor“.

10a Strukturieren^{-visumotorisch}: Es soll die Fähigkeit erfasst werden, wie gut zusammengesetzte Gestalten in einzelne Bestandteile zerlegen werden können. Der Test misst eindimensional. Nach der Beispielaufgabe können elf Items von der Testperson in beliebiger Reihenfolge bearbeitet werden. Die Verrechnung der Testleistungen erfolgt zweikategoriell. Der Test ist nur für Personen ab acht Jahren vorgesehen. Die Testperson soll Muster durch Linien in die verschiedenen Seiten des Würfels einteilen. Die Würfel sind dieselben, die im Untertest 10 verwendet werden. Bei einer Figur ist z.B. aus drei weißen Flächen und einer roten ein Quadrat zu legen.

2 Empirischer Teil

2.1 Ziel der Untersuchung

Diese Arbeit soll die folgende Frage beantworten:

Wie hoch sind die Stabilitäts-/Retest-Reliabilitäts-Koeffizienten des AID 2 (Kubinger & Wurst, 2000), die mittels Retest-Methode nach unterschiedlichen Untersuchungsintervallen berechnet werden?

Das Untersuchungsintervall enthält unterschiedliche Zeitspannen. Dadurch wird die Berechnung der Stabilität nach unterschiedlichen Zeitintervallen ermöglicht. Durch zweimalige Vorgaben von einigen Untertests in derselben Testsituation, kann die Stabilität für denselben Tag berechnet werden. Auch die Korrelationskoeffizienten für die gesamte Testbatterie nach einem Monat, einem Jahr, zwei Jahren, drei Jahren und mehr als vier Jahren sollen berechnet werden.

Die Ermittlung der Reliabilität auf diesem Weg ist für Untertest 5 und 7 besonders wichtig, da sie die Berechnung der Konfidenzintervalle von Testergebnissen dieser Untertests ermöglicht. Für diese Untertests ist es nicht möglich auf Grund der Probabilistischen Testtheorie den Standardschätzfehler zu berechnen. Auch mittels Split-half-Reliabilität können diese aus testtheoretischen Gründen nicht berechnet werden.

Die Erfassung der Stabilität der Kennwerte ist für die Interpretation der Testwerte und für die Beantwortung von Fragestellungen von Interesse, vor allem wenn es um Entscheidungen für die Zukunft geht.

2.2 Untersuchungsplan

Im Folgenden wird dargestellt, wie die Untersuchung geplant war. Es wird zuerst auf die genaue Art und Weise der geplanten Vorgabe des AID 2 eingegangen, danach werden die geplante Stichprobe und der geplante Aufwand beschrieben.

2.2.1 Vorgabe des AID 2

Die Untertests **5 Unmittelbares Reproduzieren**-numerisch **vorwärts** und **rückwärts**, **1 Alltagswissen**, **6 Synonyme finden** und **9 Funktionen Abstrahieren** sollten unmittelbar im Anschluss an einige Retestungen ein zweites Mal vorgegeben werden. Der Untertest **5 Unmittelbares Reproduzieren**-numerisch sollte das zweite Mal in der Standardform vorgegeben werden, da nicht von Erinnerungseffekten ausgegangen wurde und keine Parallellform vorhanden ist. Die Untertests **1 Alltagswissen**, **6 Synonyme finden** und **9 Funktionen Abstrahieren** sollten jedoch in der Parallellform vorgegeben werden, da sonst Übungseffekte vermutet wurden. Diese vier Subtests wurden ausgewählt, da die Vorgabe nicht viel Zeit beansprucht und deshalb die zweifache Vorgabe an ein Kind in derselben Sitzung als zumutbar eingeschätzt wurde.

Außerdem sollte in weiteren Fällen die gesamte Testbatterie im Intervall von einem Monat, einem Jahr, zwei Jahren, drei Jahren und mehr als vier Jahren wiederholt werden. Dabei sollte nach einem Monat nach Möglichkeit die Parallellform gewählt werden. Bei längeren Zeitintervallen sollte die Standardform vorgegeben werden, da auf Grund des adaptiven Testens bei unterschiedlichen Altersstufen zum Großteil andere Items bearbeitet werden und somit Erinnerungseffekte ausgeschlossen werden können.

2.2.2 Intendierte Stichprobe

Es war geplant, die Stichprobe so zu wählen, dass sich das Geschlecht über die Altersstufen der Testpersonen zwischen 6 und 15 Jahren gleich verteilt.

Zur Akquirierung der Testpersonen war es geplant, Kontrollgruppen aus Diplomarbeiten heranzuziehen. Außerdem sollten anonymisierte Daten von Testpersonen aus Beratungszentren erhoben werden, die bereits zweimal mit dem

AID 2 getestet wurden. Weiters war es geplant praktizierende PsychologInnen zu ersuchen, Elternbriefe an Eltern auszusenden, deren Kinder bereits einmal mit dem AID 2 getestet wurden, um sie zu einer nochmaligen Testung einzuladen. Außerdem sollten Teilnehmer des AID 2 Zertifizierungskurses¹ Kinder aus ihrem Bekanntenkreis im Intervall von einem Monat zweimal testen.

Auf eine klinische Stichprobe von chronisch kranken Kindern sollte verzichtet werden, da hier erwartet wurde, dass unkontrollierbare Effekte das Ergebnis vermehrt verfälschen (vgl. z.B. Knab, 1996).

2.2.3 Untersuchungszeitraum, Durchführung und Arbeitsaufwand der Untersuchung

Der Beginn der Untersuchung war im März 2008 geplant. Alle Retestungen sollten im November 2008 abgeschlossen sein.

Die Elternbriefe sollten ab März ausgesandt werden. Es war geplant dass die Kinder, die bereits einmal von praktizierenden PsychologInnen getestet wurden, von Diplomandinnen aus dem Bereich der Psychologischen Diagnostik im Rahmen der Hilfestunden² und von TeilnehmerInnen des AID 2 Zertifizierungskurses ein zweites Mal getestet werden sollten. Die Testungen sollten je nach Bedürfnis der Klienten entweder bei den Testpersonen zu Hause, in deren Schule, an den Institutionen oder an Samstagen an der Test- und Beratungsstelle der Universität Wien stattfinden. Es sollte eine kurze schriftliche Rückmeldung an die Eltern über die Testergebnisse erfolgen. Diese sollte von den AID 2 Zertifizierungskursteilnehmern geschrieben werden.

¹ Der Zertifizierungskurs wird regelmäßig für PsychologiestudentInnen angeboten. An drei Tagen wird die Vorgabe, Auswertung, Interpretation und Rückmeldung des AID 2 erklärt und geübt. Anschließend werden von den Teilnehmern 30 AID 2 Testungen durchgeführt um ein Zertifikat zum AID 2 Testleiter zu erwerben.

² Jeder Diplomand im Bereich Psychologische Diagnostik leistet seinen Kollegen insgesamt 30 Hilfestunden.

Die Ergebnisse sollten mittels des AIDScore Programms (Version 1.02) von Maryschka (2000) und mit SPSS (Version 15.0) ausgewertet werden.

2.3 Durchführung der Untersuchung

Unter den folgenden Punkten wird auf die Durchführung der Untersuchung eingegangen. Zuerst wird auf die Testpersonenakquirierung eingegangen, danach werden Abweichungen vom Untersuchungsplan dargestellt, die Stichprobe wird beschrieben und die Auswertungsverfahren werden angeführt.

2.3.1 Testpersonenakquirierung

Die Testpersonenakquirierung verlief wie geplant. Eine Übersicht zu den Häufigkeiten der akquirierten Testpersonen über die verschiedenen Institutionen kann Tabelle 6 entnommen werden.

Es wurden sämtliche PsychologInnen des BÖP,³ die in Wien praktizieren und mit Kindern und Jugendlichen arbeiten per e-Mail angeschrieben. Im e-Mail wurde kurz die geplante Diplomarbeit vorgestellt und als Attachment wurden ein Ansuchen um Unterstützung der Diplomarbeit, eine Verpflichtungserklärung und der Untersuchungsplan angefügt. Der Text des e-Mails und die Attachments befinden sich im Anhang. Von drei Psychologen in freier Praxis konnten insgesamt neun Retestungen zur Verfügung gestellt werden. Zwei Klienten einer Psychologin konnten ein zweites Mal von der Autorin getestet werden.

Frau HR Dr. Mathilde Zeman der Schulpsychologie-Bildungsberatung im Stadtschulrat für Wien erklärte sich bereit 37 Eltern von Kindern, die seit dem Schuljahr 2005/06 mit dem AID 2 getestet wurden, anzuschreiben. Ihrem Brief wurde ein Elternbrief der Autorin beigelegt, der zu einer zweiten Testung einlud. Der Inhalt der Elternbriefe kann im Anhang nachgelesen werden. Wenn sich die Eltern bei der Autorin telefonisch oder per e-Mail meldeten und einverstanden waren,

³ Über das *PSYCHNET* können im Internet auf der Seite <http://www.boep.or.at/psychnet/> Psychologen nach verschiedenen Kriterien, wie Klienten-Altersgruppen und Bundesland gesucht werden (Stand Jänner 2009).

wurde der anonymisierte Testbogen der Ersttestung per Post an die Autorin gesandt und eine Zweittestung durchgeführt. Dies war bei zwei Kindern der Fall.

Frau Dr. Renate Strasser erklärte sich bereit mit Ihren Mitarbeiterinnen 570 Eltern von Kindern anzuschreiben, die seit dem Schuljahr 2004/05 am Kompetenzzentrum für Erziehungs- und Schulfragen in Wolkersdorf mit dem AID 2 getestet worden waren. Sie verfasste einen Elternbrief, dem ebenfalls der Elternbrief der Autorin beigelegt wurde. Auch dieser Elternbrief befindet sich im Anhang. Es wurden 154 Kinder und Jugendliche ein zweites Mal getestet. Außerdem wurden anonymisierte Daten von vier Retestungen zur Verfügung gestellt, die am Zentrum durchgeführt worden waren. Nachdem sich die Eltern bei der Autorin gemeldet hatten und damit einverstanden waren, ihr Kind ein zweites Mal testen zu lassen wurde eine zweite Testung durchgeführt. Da nicht alle Kinder von der Autorin getestet werden konnten, wurde in einigen Fällen mit Erlaubnis der Eltern deren Telefonnummer an eine andere Testleiterin oder an einen anderen Testleiter weitergegeben. Danach wurde ein Termin vereinbart.

Im Rahmen des Zertifizierungskurses testeten einige der Teilnehmer Kinder aus ihrem Bekanntenkreis oder über Lerninstitute akquirierte Kinder zweimal. Der Autorin wurden auf diesem Weg anonymisierte Retestungen von 47 Testpersonen zur Verfügung gestellt.

Frau Mag. Stefana Holocher-Ertl stellte anonymisierte Daten von 44 AID 2 Retestungen aus der Test- und Beratungsstelle der Universität Wien (Fakultät für Psychologie), die am Arbeitsbereich Psychologische Diagnostik durchgeführt worden waren, zur Verfügung. Diese wurden im Rahmen einer Studie zur Hochbegabung erhoben (Schubhart, 2008).

Eine Kontrollgruppe zweier Diplomarbeiten stützte sich auf AID 2 Retestungen und konnte somit ebenfalls in die Stabilitätsuntersuchung miteinbezogen werden (Gebetsberger, 2004; Pamperl, 2004).

Die Wahl des Ortes und der Zeit der Testungen verlief wie geplant. Die Zweittestungen konnten bis Ende November bei den Testpersonen oder bei den

TestleiterInnen zu Hause, an den Schulen der Testpersonen oder an Samstagen an der Test- und Beratungsstelle der Universität Wien durchgeführt werden. Dabei wurde stets auf ein ruhiges Setting geachtet.

Tabelle 6: Häufigkeit der Vorgabe von Untertests in Kombination mit oder ohne Zusatztests

Datenakquirierung über Institution	Mit Zusatztests	Ohne Zusatztests Wiederholung 4 Subtests	Ohne Zusatztests	gesamt
Kompetenzzentrum in Wolkersdorf	0	148*	7	155
AID 2 Zertifizierungskurs	47	0	0	47
Test- und Beratungsstelle	0	0	44	44
Kontrollgruppe Diplomarbeit	27	0	0	27
Psychologen des BÖP	0	1	10	11
Schulpsychologie Wien	0	2	0	2
Gesamtstichprobe	74	151	61	286

Anmerkung: * bei einer Testperson fehlt die Ersttestung

Es gab vier verschiedene Varianten der Vorgabe des AID 2. In Tabelle 6 werden die Häufigkeiten der verschiedenen Vorgaben dargestellt. Wenn bei der Ersttestung alle Untertests und alle Zusatztests durchgeführt wurden, wurden diese bei der zweiten Testung nach Möglichkeit alle wiederholt. In einigen Fällen wurden bei der Ersttestung keine Zusatztests vorgegeben. In diesen Fällen wurden diese auch bei der Zweittestung weggelassen. Dafür wurden nach Möglichkeit die vier Subtests **5 Unmittelbares Reproduzieren_{-numerisch}, 1 Alltagswissen, 6 Synonyme finden und 9 Funktionen Abstrahieren** unmittelbar im Anschluss an die Testung ein zweites Mal vorgegeben. In einigen Fällen wurden bei der Zweittestung weder die Zusatztests vorgegeben, noch die 4 Subtests wiederholt.

Es kam während der Vorgabe des AID 2 bei einzelnen Testpersonen vor, dass einige Untertests nicht durchgeführt werden konnten. Gründe dafür waren mangelnde Motivation der Testpersonen, Müdigkeit, vorzeitiger Abbruch der Testung aus diversen Gründen etc.

2.3.2 Abweichungen der Durchführung vom Untersuchungsplan

Leider konnten die AID 2 Zertifizierungskurse erst im Juni und Juli 2008 stattfinden. Somit war die Mehrzahl der TestleiterInnen im Mai und Juni 2008, als sich viele Eltern telefonisch meldeten noch nicht zum Testen bereit. Da die Autorin alleine nicht mehr als 63 Kinder in 2 Monaten testen konnte, mussten die Testungen in die Sommerferien verschoben werden. Durch die Urlaubszeit gab es

Terminfindungsschwierigkeiten. Die Terminvereinbarung war außerdem erschwert, da viele Familien in Niederösterreich wohnten, die meisten TestleiterInnen jedoch in Wien. Im September 2008 waren viele Testpersonen wieder durch die Schule verhindert und TestleiterInnen durch die Prüfungszeit an der Universität. Auch aus diversen anderen Gründen wurden einige Termine kurzfristig abgesagt. Somit erwies sich die Koordination zwischen TestleiterInnen und Testpersonen als schwierig. Einige Testpersonen konnten von den TestleiterInnen telefonisch nicht mehr erreicht werden, andere hatten im Oktober kein Interesse mehr. Deshalb konnten 31 Kinder, deren Eltern sich ursprünglich freiwillig zur Testung gemeldet hatten, leider nicht mehr getestet werden.

Es wurden alle Retestungen in der Standardform durchgeführt. Auch die Retestungen nach einem Monat erfolgten nicht wie geplant in der Parallelförm, sondern nochmals in der Standardform. Dies war einerseits der Fall, da die zur Verfügung gestellten Daten schon vor der Untersuchungsplanung erhoben worden waren (wie die Kontrollgruppe der Diplomarbeiten), andererseits notwendig weil es sonst zu einem erhöhten organisatorischen Aufwand geführt hätte, da die Ergebnisse der Ersttestung an die Testleiter der Zweittestung weitergegeben werden hätten müssen.

2.3.3 Stichprobenbeschreibung

Die Vorstellungsgründe in den Beratungsstellen waren vielfältig. In der Test- und Beratungsstelle wurden alle Testpersonen zur Abklärung von einer möglichen Hochbegabung vorgestellt. Die Testpersonen, die im Rahmen des Zertifizierungskurses akquiriert wurden, waren bekannte Kinder der Testleiter oder wurden über Lernzentren angeschrieben. Die Vorstellungsgründe der Testpersonen bei der Schulpsychologie Wien, im Kompetenzzentrum Wolkersdorf und bei den Psychologen des BÖP können Tabelle 7 entnommen werden. Sie reichen von einer gewünschten allgemeinen Begabungsabklärung, über Schulleistungs-, Aufmerksamkeitsprobleme bis hin zu spezifischeren Fragestellungen wie Legasthenie- und Hochbegabungsabklärung, Schullaufbahnberatungen und Abklärung der Raum- und Lageorientierung.

Tabelle 7: Grund der Vorstellung

Vorstellungsgrund	absolute Häufigkeiten
Legasthenieabklärung	86
Begabungsabklärung	29
Hochbegabung	20
Schulleistung	15
Aufmerksamkeit	8
unbekannt	6
Schulwahl	3
Raumlageorientierung	1
gesamt	168

Anmerkung: Testpersonen des Kompetenzzentrums, der Psychologen des BÖP und der Schulpsychologie

Die meisten Testpersonen wohnten zum zweiten Testzeitpunkt in Wien und Niederösterreich. Einige Testpersonen des Zertifizierungskurses stammten aus Oberösterreich und aus Südtirol. Die genauen Zahlen werden in Tabelle 8 dargestellt.

Tabelle 8: Wohnort der Testpersonen zum zweiten Testzeitpunkt

Wohnorte	absolute Häufigkeiten	relative Häufigkeiten
Niederösterreich	163	57.0
Wien	107	37.4
Südtirol	8	2.8
unbekannt	5	1.7
Oberösterreich	3	1.0
gesamt	286	100

Anmerkung: Testpersonen der Gesamtstichprobe

Die Häufigkeiten der beiden Geschlechter sind in dieser Untersuchung nicht gleichverteilt, wie ursprünglich geplant. Dies resultiert daraus, dass in den Beratungszentren mehr AID 2 Daten von männlichen Testpersonen vorlagen. Die Verteilung der Geschlechter über die verschiedenen Institutionen kann Tabelle 9 entnommen werden. In der Gesamtstichprobe befinden sich mehr als doppelt so viele männliche Testpersonen.

Tabelle 9: Häufigkeiten und Geschlecht der Testpersonen über die Akquirierungsquellen

Daten akquiriert über	absolute Häufigkeiten	relative Häufigkeiten	Geschlecht	
			männlich	weiblich
Kompetenzzentrum in Wolkersdorf	155	54.2	104	51
Zertifizierungskurs	47	16.4	27	20
Test- und Beratungsstelle Uni Wien	44	15.4	35	9
Kontrollgruppe	27	9.4	12	15
Diplomarbeit	11	3.8	8	3
Psychologen des BÖP	2	0.7	2	0
Schulpsychologie Wien				
Gesamtstichprobe	286	100	188	98

Die Häufigkeit der Testpersonen ist über die Altersstufen in Jahren nicht gleichverteilt. Das Alter der Testpersonen erstreckte sich bei der ersten Testung von fünf bis 15 Jahren, bei der zweiten Testung von sechs bis 17 Jahren. Die genauen Häufigkeiten des Alters der Testpersonen bei der ersten und der zweiten Testung können Tabelle 10 entnommen werden.

Tabelle 10: Verteilung der Geschlechter über das Alter in Jahren

Alter	1. Testung		2. Testung	
	männlich	weiblich	männlich	weiblich
5	1	0	0	0
6	23	14	5	2
7	38	17	3	7
8	35	20	22	14
9	28	15	31	19
10	28	11	33	11
11	12	6	33	12
12	11	4	22	11
13	8	3	16	5
14	4	3	13	9
15	0	5	4	6
16	0	0	4	2
17	0	0	2	0
Gesamt	188	98	188	98

Anmerkung: Gesamtstichprobe

In Tabelle 11 wird die Verteilung der Altersstufen über die verschiedenen Zeitintervalle dargestellt. Damit die Stichproben der Größe nach vergleichbar sind, wurde die Gesamtstichprobe in vier gleich große Teilstichproben aufgeteilt. Das erste Intervall umfasst max. 3 Monate, das zweite Intervall umfasst ein halbes Jahr bis zu

einem Jahr und elf Monaten, das dritte Intervall erstreckt sich von zwei Jahren bis zu zwei Jahren und acht Monaten und der längste Zeitabstand zwischen den beiden Testungen beträgt mindestens zwei Jahre und neun Monate bis maximal sechs Jahre und 5 Monate.

Gesondert wurden auch die Korrelationskoeffizienten der vier Subtests berechnet, die am selben Tag ein zweites Mal in der Parallelform vorgegeben wurden und die Korrelationskoeffizienten nach ein bis zwei Monaten, um die Stabilität nach kurzen Intervallen genau beleuchten zu können.

Tabelle 11: Verteilung des Alters der Testpersonen über die verschiedenen Zeitintervalle

Alter	Vorgabe am selben Tag		Intervall 1-2 Monate		Intervall 1-3 Monate		Intervall 0.5-1.92 Jahre		Intervall 2 bis 2.67 Jahre		Intervall 2.75 bis 6.42 Jahre	
	m	w	m	w	m	m	w	w	m	w	m	w
5	0	0	1	0	1	0	0	0	0	0	0	0
6	0	0	4	2	4	2	6	8	9	2	4	2
7	1	2	2	2	3	4	10	3	10	4	14	6
8	8	5	1	0	7	8	8	3	8	7	12	2
9	15	5	3	2	8	7	8	2	4	4	8	2
10	21	8	5	3	5	3	9	1	7	3	7	4
11	21	10	2	0	2	0	4	3	4	0	2	3
12	16	9	5	2	5	2	2	0	3	2	1	0
13	8	4	3	1	3	1	2	1	2	1	1	0
14	6	4	1	3	1	3	0	0	2	0	1	0
15	4	0	0	5	0	5	0	0	0	0	0	0
16	1	1	0	0	0	0	0	0	0	0	0	0
17	2	0	0	0	0	0	0	0	0	0	0	0
Gesamt	103	48	27	20	39	35	49	21	49	23	50	19
Gesamt	151		47		74		70		72		69	
Mw	11.04	10.85	9.74	11.40	9.31	10.03	8.69	7.95	8.73	8.65	8.42	8.47
Sd	2.06	2.01	11.4	3.27	2.18	2.96	1.90	2.16	2.32	1.85	1.76	1.74
Mw	10.98		10.45		9.65		8.47		8.71		8.43	
Sd	2.04		2.99		2.63		1.99		2.17		1.75	

Anmerkung: m steht für männlich, w für weiblich, Mw für Mittelwert, Sd für Standardabweichung; es wird das Alter der Testpersonen zum Zeitpunkt der Ersttestung angegeben

Eine genauere Verteilung der Testpersonen auf die verschiedenen Zeitintervalle kann

Tabelle 12 entnommen werden. Dieselben Testpersonen, die nach einem Zeitintervall von einem bis zwei Monaten das zweite Mal getestet wurden, wurden auch in die Berechnung des Korrelationskoeffizienten nach einem Zeitintervall von einem bis zu drei Monaten miteinbezogen.

Tabelle 12: Häufigkeiten der Testpersonen über die Intervalle

Intervall in Jahren	Intervall 1-2 Monate		Intervall 1-3 Monate		Intervall 0.5-1.92 Jahre		Intervall 2 bis 2.67 Jahre		Intervall 2.75 bis 6.42 Jahre	
	m	w	m	w	m	w	m	w	m	w
0.8	26	20	26	20						
0.17	1	0	1	0						
0.25			12	15						
0.5-1					6	0				
1.08-1.5					26	11				
1.58-1.92					17	10				
2.0-2.17							18	9		
2.25-2.42							15	7		
2.5-2.67							16	7		
2.75-3.25									21	10
3.33-3.67									22	5
4.0-6.42									7	4
Gesamt	27	20	39	35	49	21	49	23	50	19
Gesamt	47		74		70		72		69	
Mw	0.09	0.08	0.14	0.16	1.40	1.53	2.32	2.33	3.45	3.42
Sd	0.02	0.00	0.78	0.84	0.31	0.21	0.22	0.23	0.70	0.56
Mw	0.09		0.15		1.44		2.32		3.44	
Sd	0.01		0.08		0.29		0.22		0.66	

Anmerkung: Mw steht für den arithmetischen Mittelwert; Sd für Standardabweichung; m für männlich und w für weiblich

Tabelle 13 ist zu entnehmen über welche Quellen die Testpersonen in den unterschiedlichen Zeitintervallen akquiriert wurden. Auffallend ist hier, dass in den beiden kürzesten Intervallen die TestleiterInnen PsychologiestudentInnen waren, die einerseits im Rahmen des Zertifizierungskurses vor allem bekannte Kinder testeten, andererseits im Rahmen einer Diplomarbeit Kinder einer Schule in Niederösterreich testeten. Die Testpersonen der längeren drei Zeitintervalle wurden jedoch über diverse Beratungsstellen akquiriert, die meisten davon über das Kompetenzzentrum in Wolkersdorf. Hier wurden die Ersttestungen von diplomierten PsychologInnen durchgeführt, die meisten Zweittestungen wurden jedoch auch von PsychologiestudentInnen durchgeführt.

Tabelle 13: Verteilung der Quelle der Testpersonenakquirierung über die Intervalle

	Vorgabe am selben Tag	Intervall 1-2 Monate	Intervall 1-3 Monate	Intervall 0.5-1.92 Jahre	Intervall 2 bis 2.67 Jahre	Intervall 2.75 bis 6.42 Jahre
Kompetenzzentrum in Wolkersdorf	148			49	53	52
Zertifizierungskurs		47	47			
Test- und Beratungsstelle				14	17	13
Kontrollgruppe Diplomarbeit			27			
Psychologen des BÖP	1			6	1	4
Schulpsychologie Wien	2			1	1	
Gesamtstichprobe	151	47	74	70	72	69

2.3.4 Auswertungsverfahren

Die Auswertung der AID 2 Testergebnisse erfolgte wie geplant zum Großteil über das Computerprogramm AIDScore (Version 1.02) von Maryschka (2000). Einige Testergebnisse konnten von den Testleitern jedoch nur unvollständig übermittelt werden, da nur noch die T-Werte oder Prozentränge vorhanden waren, oder da nicht alle Untertests vorgegeben worden waren. Deshalb sind die Stichprobengrößen der Korrelationskoeffizienten für die verschiedenen Untertests unterschiedlich groß und die der Fähigkeitsparameter meist geringer als die der T-Werte. Die Deskriptivstatistiken und die Korrelationskoeffizienten wurden mittels SPSS (Version 15.0) berechnet.

2.4 Ergebnisse

Hier sollen zuerst die Deskriptivstatistiken der T-Werte und der Fähigkeitsparameter der Untertests und die Testkennwerte zum ersten und zum zweiten Testzeitpunkt dargestellt werden. Danach werden die Korrelationskoeffizienten für die verschiedenen Intervalle dargeboten. Dabei wird auf die Korrelationskoeffizienten der Gesamtstichprobe der unterschiedlichen Intervalle zum ersten und zum zweiten Testzeitpunkt eingegangen. Da die Stichproben nach Geschlechtern getrennt einen geringen Umfang aufweisen, werden die Korrelationskoeffizienten nur zwecks Vollständigkeit angegeben und es wird nicht näher darauf eingegangen.

Die Korrelationskoeffizienten der Fähigkeitsparameter stellen ein anderes Maß dar, als die der T-Werte. Die T-Werte sind für das entsprechende Alter der Testpersonen normierte Werte. Mittels der Korrelationskoeffizienten der T-Werte wird also überprüft, inwiefern die gemessene Fähigkeit der Testpersonen im Vergleich zu gleichaltrigen Testpersonen über die Zeit konstant bleibt. Da Fähigkeitsparameter ein alterunabhängiges Maß der Fähigkeit darstellen, wird mittels der Korrelationskoeffizienten über längere Zeiträume nicht die Stabilität der Fähigkeit im Vergleich zu Gleichaltrigen überprüft, sondern eher inwiefern die Testleistungen der Testpersonen innerhalb eines bestimmten Intervalls in gleicher Weise steigen. Bei kürzeren Zeitintervallen liefern die Korrelationskoeffizienten der Fähigkeitsparameter jedoch auch Informationen über die Stabilität des Merkmals.

2.4.1 Deskriptive Ergebnisse der Testwerte

Die genauen Verteilungen der Testwerte für die verschiedenen Intervalle befinden sich im Anhang. In Tabelle 14 werden die Testkennwerte für die Gesamtstichprobe der Testpersonen zum ersten Testzeitpunkt dargestellt unabhängig von der Länge des Intervalls. In Tabelle 15 werden die Testkennwerte für den zweiten Testzeitpunkt beschrieben. Die Tabellen mit der Deskriptivstatistik der Testkennwerte nach Intervallen getrennt befinden sich im Anhang.

Bei den Zusatztests gibt es nach Erst- und Zweittestung getrennt 78 bis 92 Testpersonen. Jedoch sind es im Intervall von ein bis drei Monaten nur 57 bis 74 Testpersonen, die bei der Erst- und bei der Zweittestung die Zusatztests bearbeitet

hatten. Die anderen Testpersonen verteilen sich vereinzelt auf die Intervalle, so dass für die längeren Intervalle keine Korrelationskoeffizienten für die Zusatztests berechnet wurden.

Tabelle 14: Deskriptive Statistik der Testwerte – Testwerte der Ersttestung

Untertests	Stichproben- umfang	Mittelwert	Median	Standard- abweichung	Perzentile		
					25	50	75
1 T-Werte	285	53.63	54.00	10.20	47.00	54.00	60.00
1 FP	257	-0.42	-0.70	2.50	-2.00	-0.70	1.20
2 T-Werte	285	50.81	51.00	8.49	47.00	51.00	55.00
3 T-Werte	185	55.44	56.00	13.18	47.00	56.00	65.00
3 FP	257	0.32	0.50	2.85	-1.55	0.50	2.40
4 T-Werte	282	55.07	55.00	11.66	45.00	55.00	65.00
4 FP	255	1.14	0.70	2.86	-0.70	0.70	3.4
5v T-Werte	285	56.16	54.00	7.68	51.00	54.00	61.00
5r T-Werte	282	55.64	55.00	9.64	50.00	55.00	60.00
6 T-Werte	282	51.51	52.00	10.28	44.75	52.00	58.00
6 FP	255	-1.65	-1.60	2.30	-3.2	-1.6	-0.1
7K T-Werte	283	54.16	54.00	9.36	47.00	54.00	61.00
7A T-Werte	283	53.24	53.00	8.70	48.00	53.00	60.00
8 T-Werte	284	50.27	50.00	9.29	46.00	50.00	56.75
8 FP	257	-0.55	-0.69	2.00	-2.2	-0.69	0.40
9 T-Werte	284	52.83	53.00	10.28	46.00	53.00	59.00
9 FP	257	1.11	1.20	1.79	0.10	1.20	2.20
10 T-Werte	284	54.23	53.00	10.41	47.00	53.00	61.00
10 FP	257	-0.09	-0.70	3.03	-2.30	-0.70	1.80
11 T-Werte	282	52.79	53.00	10.65	46.00	53.00	60.00
11 FP	255	0.33	0.30	1.92	-1.00	0.30	1.50
5a T-Werte	85	55.33	53.00	10.49	47.00	53.00	61.00
5b T-Werte	92	52.13	51.00	9.07	45.00	51.00	56.75
10a T-Werte	60	55.90	55.50	9.55	50.00	55.50	61.00
10a FP	58	0.59	0.40	1.87	-0.77	0.40	1.89
Testkennwerte							
max. T-Wert	248	67.94	67.00	7.64	62.00	67.00	73.75
min. T-Wert	247	39.39	40.00	7.67	36.00	40.00	45.00
PR min. T	248	67.01	72.50	26.58	50.00	72.50	90.50
2.niedr. T-Wert	248	43.49	44.00	7.06	40.00	44.00	48.00
PR 2.niedr. T	248	67.95	75.80	27.14	50.00	75.80	90.50
Range	248	28.71	28.00	8.06	22.00	28.00	34.00
PR Range	247	48.12	46.00	27.77	24.20	46.00	69.20

Anmerkung:

5v...Untertest 5 vorwärts; 5r...Untertest 5 rückwärts; 7A...Untertest 7 Assoziationen; 7K...Untertest 7 Kodieren; max. T-Wert...maximaler T-Wert; min. T-Wert...minimaler T-Wert; PR min. T... Prozentrang untere Grenze der Intelligenzquantität; 2.niedr.T...zweitniedrigster T-Wert; PR 2.niedr. T... Prozentrang zweitniedrigste Untertestleistung; Range...Range Intelligenzquantität; PR Range...Prozentrang Range

Tabelle 15: Deskriptive Statistik – Testwerte der Zweittestung

Untertests	Stichproben- umfang	Mittelwert	Median	Standard- abweichung	Perzentile		
					25	50	75
1 T-Werte	286	54.31	54.00	9.85	48.00	54.00	61.00
1 FP	278	1.29	1.30	2.36	-0.40	1.30	2.85
2 T-Werte	286	52.92	53.00	8.22	48.00	53.00	58.25
3 T-Werte	285	56.53	58.00	12.14	49.00	58.00	65.00
3 FP	276	2.36	2.70	2.41	1.30	2.70	3.70
4 T-Werte	285	57.22	56.00	11.31	50.00	56.00	65.00
4 FP	275	3.23	3.50	2.41	1.60	3.50	5.00
5v T-Werte	284	55.90	55.00	10.21	50.00	55.00	62.00
5r T-Werte	284	54.15	53.00	9.59	49.00	53.00	58.75
6 T-Werte	285	54.53	55.00	10.40	47.50	55.00	62.00
6 FP	278	-0.09	0.00	2.04	-1.2	0.00	1.12
7K T-Werte	286	54.86	55.00	10.08	43.00	55.00	61.00
7A T-Werte	236	53.97	55.00	9.15	48.00	55.00	60.00
8 T-Werte	285	54.69	54.00	9.85	49.00	54.00	60.00
8 FP	274	1.05	1.00	2.20	-0.10	1.00	2.20
9 T-Werte	285	55.44	55.00	9.82	50.00	55.00	62.00
9 FP	277	2.53	2.70	1.64	1.60	2.70	3.70
10 T-Werte	235	57.39	58.00	10.67	50.00	58.00	63.50
10 FP	274	2.23	2.50	2.87	0.15	2.50	4.10
11 T-Werte	284	55.83	55.50	10.85	48.00	55.50	62.75
11 FP	276	1.79	1.89	1.89	0.59	1.89	3.10
5a T-Werte	78	55.31	54.50	8.20	48.00	54.50	61.00
5b T-Werte	77	53.35	55.00	8.28	47.00	55.00	57.00
10a T-Werte	63	56.87	59.00	9.98	52.00	59.00	64.00
10a FP	62	0.77	0.90	1.92	0.05	0.90	1.50
Testkennwerte							
max. T-Wert	264	69.88	70.00	7.52	65.00	70.00	76.00
min. T-Wert	263	41.55	42.00	7.89	37.00	42.00	47.00
PR min. T	263	73.80	81.60	24.63	57.90	81.60	94.50
2.niedr. T-Wert	264	45.88	47.00	6.99	42.00	47.00	51.00
PR 2.niedr. T	264	76.02	88.50	24.89	61.80	88.50	95.20
Range	263	28.33	28.00	7.55	24.00	28.00	33.00
PR Range	263	46.27	46.00	26.32	27.60	46.00	67.40

Anmerkung:

5v...Untertest 5 vorwärts; 5r...Untertest 5 rückwärts; 7A...Untertest 7 Assoziationen; 7K...Untertest 7 Kodieren; max. T-Wert...maximaler T-Wert; min. T-Wert...minimaler T-Wert; PR min. T...Prozentrang untere Grenze der Intelligenzquantität; 2.niedr.T...zweitniedrigster T-Wert; PR 2.niedr. T...Prozentrang zweitniedrigste Untertestleistung; Range...Range Intelligenzquantität; PR Range...Prozentrang Range

An den mittleren T-Werten der Untertests und besonders deutlich am Mittelwert des Prozentrangs der unteren Grenzen der Intelligenzquantität (67.01 und 73.80) ist zu erkennen, dass die untersuchte Stichprobe höhere T-Werte erzielte als der Durchschnitt der Normstichprobe.

Es können deskriptive Unterschiede, die nicht auf Signifikanz geprüft wurden, festgestellt werden. Bezüglich der Fähigkeitsparameter ist in den Tabellen (Tabelle 28 bis Tabelle 33) sichtbar, dass die mittleren Fähigkeitsparameter in der

Zweittestung ansteigen, und zwar umso mehr je länger das Zeitintervall zwischen den Testungen ist.

Wenn man die mittleren T-Werte der Erst- und der Zweittestung miteinander vergleicht, sieht man, dass in der Gesamtstichprobe in den meisten Untertests (in Untertest **2, 3, 4, 8, 9, 10, 11** und **10a**) bei der Zweittestung im Mittel höhere Werte erreicht wurden.

Nach Intervallen getrennt (siehe Tabelle 22 und Tabelle 23 im Anhang), ist ersichtlich, dass bei zweimaliger Vorgabe der vier Untertests **1, 6** und **9** am selben Tag die mittleren T-Werte der Ersttestung ein wenig niedriger sind als die der Zweittestung. Die mittleren T-Werte der Ersttestung des Untertests 5 entsprechen in etwa denen der Zweittestung.

Nach einem Intervall von ca. einem Monat (siehe Tabelle 24 und Tabelle 25 im Anhang) können an den Mittelwerten in acht Untertests bei der Zweittestung Anstiege von mindestens zwei T-Werten festgestellt werden. Dies ist bei den Untertests **1, 2, 3, 4, 6, 7, 8** und **10** der Fall. Besonders hervorzuheben ist der Anstieg in Untertest **8**. Hier liegen zwischen Erst- und Zweittestung acht T-Werte. In den Untertests **4, 7** und **10** liegt zwischen Erst- und Zweittestung bei den mittleren T-Werten mindestens eine halbe Standardabweichung, das sind mindestens 5 T-Werte. Bei den mittleren Testkennwerten „maximaler T-Wert“, „minimaler T-Wert“ und „zweitniedrigste Untertestleistung“ kam es ebenfalls zu Anstiegen um vier T-Werte.

Bei zehn Untertests kam es nach ein bis drei Monaten (siehe Tabelle 26 und Tabelle 27) zu Anstiegen der mittleren Testwerte um mindestens zwei T-Werte. Dies war bei den T-Werten von Untertest **1, 2, 3, 4, 6, 7, 8, 10, 11** und **10a** der Fall. Der Anstieg liegt bei den Untertests **7, 8** und **10** über einer halben Standardabweichung von 5 T-Werten. Die mittleren Testkennwerte „maximaler T-Wert“, „minimaler T-Wert“ und „zweitniedrigste Untertestleistung“ stiegen um drei bis vier T-Werte an.

Nach einem Intervall von einem halben Jahr bis zu einem Jahr und elf Monaten (siehe Tabelle 28 und Tabelle 29) stieg der Mittelwert der Untertests **2, 3, 4, 6, 7, 8, 9, 10** und **11** um mindestens 2 T-Werte, doch lediglich im Untertest **9** betrug dieser

Anstieg eine halbe Standardabweichung von fünf T-Werten. Die mittleren Testkennwerte „maximaler T-Wert“, „minimaler T-Wert“ und „zweitniedrigste Untertestleistung“ stiegen um zwei T-Werte an.

Nach dem Intervall von zwei Jahren bis zu zwei Jahren und acht Monaten kam es zu einem Anstieg der mittleren T-Werte um mindestens zwei T-Werte in den Untertests **2, 4, 6, 8** und **11** (siehe Tabelle 30 und Tabelle 31). Im Untertest **8** war der Anstieg gerundet eine halbe Standardabweichung. Die mittleren Testkennwerten „maximaler T-Wert“, „minimaler T-Wert“ und „zweitniedrigste Untertestleistung“ waren bei der Zweittestung um zwei T-Werte höher als bei der Ersttestung.

Nach dem längsten Zeitintervall von zwei Jahren und neun Monaten bis zu sechs Jahren und fünf Monaten kam es bei den Untertests **6, 8, 10** und **11** zu Anstiegen der Mittelwerte um mehr als 2 T-Werte (siehe Tabelle 32 und Tabelle 33). Bei den Untertests **8** und **11** lagen die Mittelwerte der Zweittestung 5 T-Werte über denen der Ersttestung. Die mittleren Testkennwerten „maximaler T-Wert“, „minimaler T-Wert“ und „zweitniedrigste Untertestleistung“ waren auch hier bei der Zweittestung um zwei T-Werte höher, als bei der Ersttestung.

2.4.2 Retest-Reliabilität bei Vorgabe am selben Tag

Tabelle 16: Retest-Reliabilität bei Vorgabe am selben Tag

Subtests	Korrelationskoeffizienten									
	T-Werte					Fähigkeitsparameter				
	m.	w.	6-10	11-17	ges.	m.	w.	6-10	11-17	ges.
1	.856	.764	.832	.838	.839	.886	.856	.868	.864	.884
5 vorwärts	.819	.700	.875	.713	.792					
5 rückwärts	.579	.766	.729	.580	.632					
6	.822	.739	.810	.806	.807	.884	.784	.846	.836	.866
9	.821	.806	.724	.871	.817	.819	.814	.746	.859	.819

Anmerkung:

Subtest 1, 6 und 9 wurden in Parallelforn vorgegeben, 5 nachmals in Standardform

gesamt (ges.) n=151; männlich (m.): n=103; weiblich (w.): n=48;

Alter 6-10Jahre: n=65; Alter 10-17Jahre: n=86, Alter Mw=8.98, Sd=1.81;

für Fähigkeitsparameter: ges.: n=147; m.: n=99; w.: n=48; Alter 6-10Jahre: n=65; Alter 11-17Jahre: n=82

n steht für Stichprobenumfang, Mw für Mittelwert, Sd für Standardabweichung

In Tabelle 16 sind die Korrelationskoeffizienten der vier Subtests dargestellt, die am selben Tag zweimal bearbeitete wurden. Die Testpersonen bearbeiteten zuerst alle elf Subtests des AID 2, danach wurden Subtest **1 Alltagswissen** in der Parallelforn, **5 Unmittelbares Reproduzieren-numerisch vorwärts** und **rückwärts** in der Standardform, **6 Synonyme Finden** in der Parallelforn und Subtest **9 Funktionen Abstrahieren** in der Parallelforn in genannter Reihenfolge ein zweites Mal vorgegeben.

Die Korrelationskoeffizienten liegen, außer bei Untertest **5 Unmittelbares Reproduzieren-numerisch rückwärts** alle über 0.7. Auffallend ist, dass bei den älteren Testpersonen im Alter von 11 bis 17 Jahren die Korrelationskoeffizienten des Untertests **5 Unmittelbares Reproduzieren-numerisch rückwärts** am niedrigsten ausgefallen sind. Die niedrigeren Korrelationskoeffizienten der weiblichen Stichprobe im Vergleich zu der männlichen sind vermutlich auf die geringere Stichprobengröße zurückzuführen.

Festzuhalten ist, dass diese auffallenden Unterschiede rein deskriptiv sind und nicht auf Signifikanz überprüft wurden.

2.4.3 Stabilität nach einem Intervall von ca. einem Monat

Tabelle 17: Stabilität nach ca. einem Monat

	Korrelationskoeffizienten			Fähigkeitsparameter
	T-Werte			
Subtests	m.	w.	ges.	gesamt
1 Alltagswissen	.608	.966	.785	.955
2 Realitätssicherheit	.706	.519	.575	
3 Angewandtes Rechnen	.836	.615	.748	.920
4 Soziale und Sachliche Folgerichtigkeit	.647	.800	.695	.907
5 Unmittelbares Reproduzieren _{-numerisch} vorwärts	.587	.740	.633	
5 Unmittelbares Reproduzieren _{-numerisch} rückwärts	.414	.676	.511	
6 Synonyme Finden	.702	.859	.784	.921
7 Kodiermenge	.637	.539	.601	
7 Assoziationen	.652	.306	.540	
8 Antizipieren und Kombinieren	.771	.792	.812	.780
9 Funktionen Abstrahieren	.702	.589	.684	.869
10 Analysieren und Synthetisieren _{-abstrakt}	.663	.669	.695	.865
11 Soziales Erfassen und Sachliches Reflektieren	.787	.714	.746	.926
5a Unmittelbares Reproduzieren – figural abstrakt	.451	.315	.403	
5b Merken und Einprägen	.586	-.189	.292	
10a Strukturieren _{-visumotorisch}	.483	.234	.400	.484
Testkennwerte	m.	w.	ges.	
maximaler T-Wert	.486	.760	.680	
minimaler T-Wert	.583	.580	.617	
Prozentrang untere Grenze Intelligenzquantität	.607	.631	.648	
zweitniedrigster T-Wert	.591	.822	.723	
Prozentrang zweitniedrigste Untertestleistung	.720	.811	.782	
Range Intelligenzquantität	.430	-.266	.201	
Prozentrang Range	.444	-.302	.210	

Anmerkung:

gesamt (ges.): n=47; männlich (m.): n=27; weiblich (w.): n=20; Alter: Mw=10.94, Sd=3.1; Intervall in Tagen: Mw=31, Sd=4;

für Subtest 10a: ges.: n=34; m.: n=19; w.: n=15

für Testkennwerte: ges.: n=39; m.: n=22; w.: n=17

für Subtest 8, 9, 10, 11 Fähigkeitsparameter: ges.: n=46; m.: n=26; w.: n=20;

für 10a Fähigkeitsparameter ges.: n=33; m.: n=17; w.: n=16;

n steht für Stichprobenumfang, Mw für Mittelwert, Sd für Standardabweichung

In Tabelle 17 sind die Korrelationskoeffizienten der Testungen dargestellt, deren Zeitintervall ca. ein Monat betrug.

Die Gesamtstichprobe betrachtend kann festgestellt werden, dass fast alle T- Werte der Untertests einen Korrelationskoeffizient aufweisen, der höher als 0.6 ist. Niedrigere Korrelationskoeffizienten traten in den Untertests **2, 5 Unmittelbares Reproduzieren^{-numerisch} rückwärts, 7 Assoziationen** und in den drei Zusatztests **5a, 5b und 10a** auf. Besonders hervorzuheben sind die hohen Korrelationskoeffizienten in den Tests **1, 3, 6, 8 und 11** die höher als 0.74 sind.

Die Fähigkeitsparameter der Untertests weisen alle Korrelationen über 0.86 auf, außer der Zusatztest **10a**.

Die Korrelationskoeffizienten der Testkennwerte wurden zwecks Vollständigkeit erhoben. Überraschend hoch sind hier die Korrelationskoeffizienten Gesamtstichprobe der „zweitniedrigsten Untertestleistung“. Auch die Korrelationskoeffizienten der Gesamtstichprobe der Testkennwerte „maximaler T-Wert“ und „minimaler T-Wert“ liegen über 0.6. Auffallend sind die Unterschiede der Korrelationskoeffizienten zwischen den männlichen und den weiblichen Testpersonen in den Testkennwerten „maximaler T-Wert“ und „zweitniedrigster T-Wert“. Ebenfalls herausstechend sind die negativen Korrelationskoeffizienten der weiblichen Testpersonen bei den Testkennwerten „Range Intelligenzquantität“ und „Prozentrang Range“. Da der Stichprobenumfang nach Geschlechtern getrennt jedoch gering ist können diese Korrelationskoeffizienten nicht inhaltlich interpretiert werden.

2.4.4 Stabilität nach einem Intervall von ein bis drei Monaten

Tabelle 18: Stabilität nach ein bis drei Monaten

Subtests	Korrelationskoeffizienten			
	T-Werte			Fähigkeitsparameter
	m.	w.	ges.	gesamt
1 Alltagswissen	.450	.878	.714	.924
2 Realitätssicherheit	.636	.515	.554	
3 Angewandtes Rechnen	.839	.738	.798	.913
4 Soziale und Sachliche Folgerichtigkeit	.528	.641	.579	.853
5 Unmittelbares Reproduzieren-numerisch vorwärts	.517	.676	.572	
5 Unmittelbares Reproduzieren-numerisch rückwärts	.248	.642	.396	
6 Synonyme Finden	.507	.792	.661	.850
7 Kodiermenge	.558	.154	.414	
7 Assoziationen	.535	.240	.437	
8 Antizipieren und Kombinieren	.752	.727	.768	.759
9 Funktionen Abstrahieren	.651	.746	.711	.837
10 Analysieren und Synthetisieren-abstrakt	.605	.579	.642	.842
11 Soziales Erfassen und Sachliches Reflektieren	.806	.689	.756	.906
5a Unmittelbares Reproduzieren – figural abstrakt	.417	.277	.385	
5b Merken und Einprägen	.173	.113	.132	
10a Strukturieren-visumotorisch	.175	.512	.403	.585
Testkennwerte	m.	w.	ges.	
maximaler T-Wert	.463	.774	.674	
minimaler T-Wert	.659	.398	.597	
Prozentrang untere Grenze Intelligenzquantität	.351	.531	.494	
zweitniedrigster. T-Wert	.498	.850	.745	
Prozentrang zweitniedrigste Untertestleistung	.595	.691	.661	
Range Intelligenzquantität	.442	.070	.216	
Prozentrang Range	.343	.080	.232	

Anmerkung:

gesamt (ges.) n=74; männlich (m.): n=39; weiblich (w.): n=35; Alter: Mw=10.14 Sd=2.7; Intervall in Tagen: Mw=54, Sd=30;

für Subtest 10a: ges.: n=57; m.: n=30; w.: n=27;

für Testkennwerte, max. T-Wert, zweitniedrigster T-Wert, PR zweitniedrigste Untertestleistung, Range Intelligenzquantität: ges.: n=56; m.: n=28; w.: n=28;

für Testkennwerte min. T-Wert : ges. n=54; m.: n=28; w.: n=26;

für Testkennwerte PR untere Grenze Intelligenzquantität: ges.: n=55; m.: n=28; w.: n=27

für Fähigkeitsparameter: ges.: n=74; m.: n=39; w.: n=35

für Fähigkeitsparameter Subtest 9, 10 und 11: ges. n= 73; m.: n=38; w.: n=35

für Fähigkeitsparameter 10a: ges.: n=56; m.: n=28; w.: n=28

n steht für Stichprobenumfang, Mw für Mittelwert, Sd für Standardabweichung

In Tabelle 18 sind Korrelationskoeffizienten nach einem Zeitintervall von einem bis zu drei Monaten zu sehen.

Im Vergleich zu den Korrelationskoeffizienten nach einem Zeitintervall von einem Monat sind diese niedriger. Lediglich in den Untertests **3**, **11** und **10a** gibt es eine marginale Verbesserung. Die meisten Korrelationskoeffizienten der Fähigkeitsparameter für dieses Intervall sind geringer als die für das kürzere Intervall

von einem Monat. Die Korrelationskoeffizienten der Testkennwerte „maximaler T-Wert“ und „zweitniedrigste Untertestleistung“ sind auffallend hoch und liegen über 0.6.

2.4.5 Stabilität nach einem Intervall von 0.5 bis 1.92 Jahren

Tabelle 19: Stabilität nach 0.5 bis 1.92 Jahren

	Korrelationskoeffizienten			
	T-Werte			Fähigkeitsparameter
Subtests	m.	w.	ges.	gesamt
1 Alltagswissen	.576	.719	.604	.781
2 Realitätssicherheit	.416	.141	.350	
3 Angewandtes Rechnen	.692	.763	.706	.823
4 Soziale und Sachliche Folgerichtigkeit	.472	.048	.395	.707
5 Unmittelbares Reproduzieren-numerisch vorwärts	.716	.515	.676	
5 Unmittelbares Reproduzieren-numerisch rückwärts	.383	.418	.397	
6 Synonyme Finden	.524	.690	.535	.671
7 Kodiermenge	.656	.649	.674	
7 Assoziationen	.476	.513	.496	
8 Antizipieren und Kombinieren	.610	.664	.617	.755
9 Funktionen Abstrahieren	.471	.566	.498	.620
10 Analysieren und Synthetisieren-abstrakt	.689	.699	.686	.859
11 Soziales Erfassen und Sachliches Reflektieren	.660	.619	.625	.748
Testkennwerte	m.	w.	ges.	
maximaler T-Wert	.564	.376	.516	
minimaler T-Wert	.651	.622	.633	
Prozentrang untere Grenze Intelligenzquantität	.673	.524	.641	
zweitniedrigster T-Wert	.770	.686	.755	
Prozentrang zweitniedrigste Untertestleistung	.763	.593	.730	
Range Intelligenzquantität.	.326	.153	.274	
Prozentrang Range	.145	.190	.147	

Anmerkung:

gesamt (ges.): n=70; männlich (m.): n=49; weiblich (w.) n=21; Alter: Mw=8.47 Sd=1.99; Intervall in Jahren: Mw=1.44, Sd=0.29;

für Subtest 7: ges.: n=69; m.: n=49; w.: n=20

für Subtests 4, 5, 6, 8, 10, 11: ges.: n=68; m.: n=49; w.: n=19

für Subtest 9: ges.: n=67; m.: n=49; w.: n=18

für Fähigkeitsparameter: ges.: n=58; m.: n=41; w.: n=17

für Testkennwerte: ges.: n=60; m.: n=42; w.: n=18

N steht für Stichprobenumfang, Mw für Mittelwert, Sd für Standardabweichung

Tabelle 19 zeigt die Korrelationskoeffizienten nach einem Zeitintervall von einem halben Jahr bis zu einem Jahr und 11 Monaten.

Die Korrelationskoeffizienten der Fähigkeitsparameter weisen alle eine Höhe über 0.7 auf, außer die Untertests 6 und 9, die jedoch auch über 0.6 liegen.

Bei den Korrelationskoeffizienten der T-Werte ist jedoch ein Abfall der Stabilität im Vergleich zu dem Intervall von drei Monaten zu beobachten. Immerhin liegt der Korrelationskoeffizient des Untertests 3 über 0.7 und die Korrelationskoeffizienten der sechs Untertests: 1, 5 (Unmittelbares Reproduzieren-numerisch vorwärts), 7 (Kodiermenge), 8, 10 und 11 liegen über 0.6.

Die Korrelationskoeffizienten der Testkennwerte „minimaler T-Wert“ und „zweitniedrigste Untertestleistung“ liegen über 0.6.

2.4.6 Stabilität nach einem Intervall von 2 bis 2.67 Jahren

Tabelle 20: Stabilität nach 2 bis 2.67 Jahren

Subtests	Korrelationskoeffizienten			
	T-Werte			Fähigkeitsparameter
	m.	w.	ges.	gesamt
1 Alltagswissen	.770	.823	.793	.859
2 Realitätssicherheit	.348	.325	.340	
3 Angewandtes Rechnen	.509	.680	.589	.780
4 Soziale und Sachliche Folgerichtigkeit	-.010	.391	.149	.657
5 Unmittelbares Reproduzieren-numerisch vorwärts	.653	.465	.593	
5 Unmittelbares Reproduzieren-numerisch rückwärts	.407	.617	.460	
6 Synonyme Finden	.651	.494	.621	.818
7 Kodiermenge	.403	.638	.499	
7 Assoziationen	.214	.622	.429	
8 Antizipieren und Kombinieren	.442	.518	.478	.576
9 Funktionen Abstrahieren	.345	.535	.405	.633
10 Analysieren und Synthetisieren-abstrakt	.610	.601	.617	.753
11 Soziales Erfassen und Sachliches Reflektieren	.754	.496	.676	.828
Testkennwerte	m.	w.	ges.	
maximaler T-Wert	.401	.494	.470	
minimaler T-Wert	.589	.807	.676	
Prozentrang untere Grenze Intelligenzquantität	.569	.705	.618	
zweitniedrigster T-Wert	.698	.799	.742	
Prozentrang zweitniedrigste Untertestleistung	.635	.645	.643	
Range Intelligenzquantität	.303	.040	.284	
Prozentrang Range	.238	.042	.229	

Anmerkung:

gesamt (ges.): n=72; männlich (m.): n=49; weiblich (w.) n=23; Alter: Mw=8.71 Sd=2.32; Intervall in Jahren: Mw=2.64, Sd=0.22

für Subtests 3, 5, 11: ges.: n=71; m.: n=49; w.: n=22

für Testkennwerte: ges.: n=64; m.: n=43; w.: n=21

für Fähigkeitsparameter: ges.: n=65; m.: n=43; w.: n=22

n steht für Stichprobenumfang, Mw für Mittelwert, Sd für Standardabweichung

In Tabelle 20 sind die Korrelationskoeffizienten nach einem Zeitintervall von zwei Jahren bis zu zwei Jahren und acht Monaten dargestellt.

Bei den Korrelationskoeffizienten der T-Werte ist wieder ein Abfall der Stabilität im Vergleich zu den kürzeren Intervallen zu beobachten. Überraschend hoch ist in dieser Stichprobe der Korrelationskoeffizient des Untertests 1 mit 0.793. Hier ist die Langzeitstabilität höher auffallend hoch.

Außer diesem Untertest liegen noch drei weitere Korrelationskoeffizienten über 0.6. Das sind die der Untertests 6, 10 und 11. Besonders niedrig sind die Korrelationskoeffizienten der Untertests 2 und 4.

Die Fähigkeitsparameter weisen eine höhere Stabilität auf. Außer dem Korrelationskoeffizienten des Untertests **8** ($r_{tt}=0.576$) liegen alle über 0.63, drei sogar über 0.8.

Die Korrelationskoeffizienten den Testkennwerte „minimaler T-Wert“ und „zweitniedrigste Untertestleistung“ liegen über 0.6.

2.4.7 Stabilität nach einem Intervall von 2.75 bis 6.42 Jahren

Tabelle 21: Stabilität nach 2.75 bis 6.42 Jahren

	Korrelationskoeffizienten			
	T-Werte			Fähigkeitsparameter
Subtests	m.	w.	ges.	gesamt
1 Alltagswissen	.666	.652	.687	.667
2 Realitätssicherheit	.519	.322	.456	
3 Angewandtes Rechnen	.556	.791	.639	.621
4 Soziale und Sachliche Folgerichtigkeit	.235	.035	.179	.216
5 Unmittelbares Reproduzieren-numerisch vorwärts	.530	.643	.561	
5 Unmittelbares Reproduzieren-numerisch rückwärts	.251	.667	.370	
6 Synonyme Finden	.504	.277	.511	.661
7 Kodiermenge	.670	.288	.557	
7 Assoziationen	.252	-.054	.171	
8 Antizipieren und Kombinieren	.757	.454	.689	.733
9 Funktionen Abstrahieren	.599	.641	.599	.589
10 Analysieren und Synthetisieren-abstrakt	.577	.576	.586	.666
11 Soziales Erfassen und Sachliches Reflektieren	.369	.400	.380	.498
Testkennwerte	m.	w.	ges.	
maximaler T-Wert	.509	.054	.401	
minimaler T-Wert	.622	.676	.626	
Prozentrang untere Grenze Intelligenzquantität	.529	.545	.522	
zweitniedrigster T-Wert	.685	.750	.677	
Prozentrang zweitniedrigste Untertestleistung	.514	.711	.547	
Range Intelligenzquantität	.262	-.108	.171	
Prozentrang Range	.290	-.024	.213	

Anmerkung:

gesamt (ges.): n=69; männlich (m.): n=50; weiblich (w.) n=19; Alter: Mw=8.43 Sd=1.75; Intervall in Jahren: Mw=3.44, Sd=.66;

für Subtests 4, 5, 6, 7, 11: ges.: n=68; m.: n=50; w.: n=18;

für Testkennwerte: ges.: n=57; m.: n=41; w.: n=16

für Fähigkeitsparameter: ges.: n=54; m.: n=38; w.: n=16

n steht für Stichprobenumfang, Mw für Mittelwert, Sd für Standardabweichung

Tabelle 21 zeigt die Korrelationskoeffizienten nach dem längsten Zeitintervall dieser Studie. Hier betrug der Zeitabstand zwischen den beiden Testzeitpunkten von zwei Jahren und neun Monaten bis zu sechs Jahren und fünf Monaten.

Die Untertests, deren Korrelationskoeffizienten der T-Werte über 0.6 liegen sind **1, 3** und **8**. Auffallend niedrig sind die Korrelationskoeffizienten der Untertests **4** und **7 Assoziationen**.

Die Korrelationskoeffizienten der Fähigkeitsparameter der Untertests **1, 3, 6, 8** und **10** liegen über 0.6. Am niedrigsten ist der Korrelationskoeffizient des Fähigkeitsparameters des Untertest **4**.

Von den Testkennwerten liegen die Korrelationskoeffizienten des minimalen T-Werts und der zweitniedrigsten Untertestleistung über 0.6.

3 Diskussion

Hier werden kurz die aufgetretenen Probleme bei der Testpersonenakquirierung erörtert, danach werden die Ergebnisse anhand der Literatur interpretiert und zum Schluss werden die Ergebnisse dieser Arbeit in Bezug zur Praxis interpretiert und ein Ausblick in die Zukunft vorgestellt.

Das erste Problem, das bei dieser Untersuchung auftrat, war im Zuge der Testpersonenakquirierung. Wie Bortz und Döring (2005) berichteten, ist mit großen Ausfallzahlen zu rechnen und deshalb muss bei der Untersuchungsplanung eine große „Start-Stichprobe“ gezogen werden. Viele Psychologen wurden kontaktiert und von diversen Institutionen wurden zahlreiche Testpersonen angeschrieben. Dennoch blieb auch bei dieser Untersuchung das Problem, wie von Bortz und Döring (2005) beschrieben, dass überwiegend Testpersonen mit schlechten Testergebnissen ausfielen. Es konnten hauptsächlich Testpersonen zu einer zweiten Testung motiviert werden, die bei der ersten Testung sehr gute Testergebnisse erzielten und die wegen Hochbegabung in Beratungszentren vorgestellt worden waren. An den deskriptiven Ergebnissen der Stichproben ist zu erkennen, dass der mittlere Prozentrang der unteren Grenze der Intelligenzquantität in allen Stichproben über dem der Normstichprobe liegt. Auch konnten nicht gleich viele männliche und weibliche Personen getestet werden. Dadurch ist die Stichprobe verzerrt. Das ist vor allem ein Problem, da das Maß der Retest-Reliabilität stichprobenabhängig ist. Wenn mehrere Personen mit niedrigeren T-Werten in den Untertests teilgenommen hätten und somit die Varianzen der T-Werte in den Stichproben höher wären, wären eventuell auch die Korrelationskoeffizienten höher ausgefallen. Positiv anzumerken ist jedoch, dass durch die große Start-Stichprobe im Endeffekt Retestungen von 286 Testpersonen zur Analyse zur Verfügung standen. Ein weiteres Problem bei der Akquirierung der Testpersonen war, dass in den Beratungsstellen mehr AID 2 Testungen von männlichen Testpersonen vorlagen, als von weiblichen. Deshalb ist auch in dieser Studie das Geschlechterverhältnis nicht ausgewogen.

Intelligenz gilt als „trait“, somit wurden aufgrund der Literatur möglichst stabile Ergebnisse erwartet. Doch anhand der berichteten Studien wird auch deutlich, dass gerade in der Kindheit Schwankungen auftreten können, vor allem auf Grund von Entwicklungsveränderungen. Die Erkenntnisse in Punkt 1.1.3 verdeutlichen die rasanten Entwicklungsschritte von Kindern und Jugendlichen. Wenn alle Faktoren betrachtet werden, die einen Einfluss auf die Entwicklung eines Kindes haben, ist es verständlich, dass Forschungsergebnisse darauf hinweisen, dass Intelligenz im Kindes- und Jugendalter noch nicht so ein stabiles Merkmal ist, wie später im Erwachsenenalter. Um zusätzliche Verzerrungen der Stabilität zu vermeiden, wurden in dieser Stichprobe Testpersonen aus Kliniken (etwa aus der Kinderneuroonkologie) ausgeschlossen, da in der Literatur berichtet wurde, dass in dieser Population häufiger Verzerrungen der Testwerte auftreten (vgl. Knab, 1996; Weinstein & Teuber, 1957; Rist & Dirksmeier, 2001).

Nach den kürzesten Zeitintervallen, also nach der Retestung am selben Tag und der Stabilität nach einem Monat wurden die höchsten Korrelationskoeffizienten dieser Studie berechnet. Auch in vergleichbaren Studien sind die Korrelationskoeffizienten nach kürzeren Zeitintervallen relativ hoch und werden mit länger werdenden Zeitintervallen geringer (z.B. Anderson, 1939).

Die in Punkt 1.2.4 geforderte Reliabilität von mindestens 0.75 oder noch besser 0.8 konnte zum Großteil nach dem kürzesten Zeitintervall, also bei der Retestung am selben Tag errechnet werden. Beim Untertest **5 Unmittelbares Reproduzieren-numerisch rückwärts** war der Korrelationskoeffizient der Gesamtstichprobe der Testpersonen, die beide an demselben Tag den Untertest bearbeiteten, jedoch geringer, nämlich nur $r_{tt}=0.632$. Interessant ist hingegen, dass unter den weiblichen Testpersonen dieser Stichprobe der Korrelationskoeffizient mit $r_{tt}=0.766$ höher war als der der männlichen mit $r_{tt}=0.579$. Hier sei nochmals darauf hingewiesen, dass alle Unterschiede, die zwischen den Korrelationskoeffizienten, die in dieser Studie berechnet wurden, rein deskriptiv sind und nicht auf deren Signifikanz überprüft wurden. Durch die Vorgabe der Subtests **1, 6 und 9** wurde sowohl die Retest-Reliabilität als auch die Paralleltest-Reliabilität überprüft. Diese können nach den Forderungen von Weise (1975) und wenn von der Definition ausgegangen wird, dass

68% der Verteilung der Testwerte der Referenzpopulation den Durchschnittsbereich ausmachen (vgl. Kubinger, 2006), als gegeben angenommen werden. Ein T-Wert von 50 im Untertest **1** kann z.B. mit 95% Sicherheit als durchschnittlich klassifiziert werden, wenn man die Stabilität für denselben Tag berücksichtigen möchte. Die Kapazität der seriellen Informationsverarbeitung (im verbal-akustischen Bereich) des Arbeitsgedächtnis, die mit Untertest **5 Unmittelbares Reproduzieren-numerisch rückwärts** erfasst wird, scheint jedoch nicht so stabil zu sein, wie: die Fähigkeit sich Sachkenntnisse über heute alltägliche Dinge anzueignen (Untertest **1**), das elementare Sprachverständnis (Untertest **6**) und die Fähigkeit, durch Abstraktion zu einer Begriffsbildung zu gelangen (Untertest **9**). Dies kann möglicherweise dadurch erklärt werden, dass das Arbeitsgedächtnis situativen Einflüssen stärker unterlegen ist, als die anderen Fähigkeiten (vgl. z.B. Dixon, 2003). Auffallend ist, dass die Korrelationskoeffizienten der Fähigkeitsparameter der Subtests **6** und **9** bei der Vorgabe am selben Tag etwas geringer ausfielen als nach einem Intervall von einem Monat. Anhand Tabelle 22 und Tabelle 23 kann man sehen, dass die Fähigkeitsparameter in diesen Subtests bei der Vorgabe am selben Tag bei der Zweittestung im Mittel etwas niedriger waren als bei der Ersttestung. Der Grund dafür könnte also die Erschöpfung der Testpersonen sein.

Nach einem Intervall von einem Monat kann die Reliabilität nach dem Kriterium, dass diese mindestens 0.75 betragen soll (Wagner-Menghin, 2003) für die T-Werte der Untertests **1, 3, 6, 8** und **11** als ausreichend angesehen werden. Wenn man die Fähigkeitsparameter heranzieht, sind die Untertests **1, 3, 4, 6, 8, 9, 10** und **11** als zuverlässig zu bezeichnen. Auch der Prozentrang der zweitniedrigsten Untertestleistung ist ein zuverlässiger Messwert. Gründe, warum die anderen Untertests niedrigere Korrelationskoeffizienten aufweisen, können in der Literatur viele gefunden werden. Die Korrelationskoeffizienten können etwa durch Übungs- und Erinnerungseffekte verzerrt worden sein (Kubinger, 2006). Dies wurde auch von Amthauer (1957) und Catron (1978) berichtet. Bei zweimaliger Vorgabe eines Intelligenztests im Abstand von einem Monat kommt es nach der Literatur zu systematischen Mittelwertszunahmen von ca. der Hälfte einer Standardabweichung. Auch in dieser Diplomarbeitsstudie konnten Mittelwertszunahmen der T-Werte festgestellt werden. Anhand der deskriptiven Statistik sieht man, wenn man die

Mittelwerte der T-Werte der Untertests der ersten und zweiten Testung vergleicht, dass es zu Steigerungen von bis zu 8 T-Werten (in Untertest **8**) kam. Aber auch in den Untertests **2, 3, 4, 7, 8** und **10** kam es zur Steigerung der Mittelwerte der T-Werte in der zweiten Testung. In den Untertests **4, 7, 8** und **10** umfassten diese Steigerungen mindestens eine halbe Standardabweichung. Die Inhalte dieser Untertests (Bildgeschichten, Symbole kodieren, Puzzleteile und mit Würfeln Muster legen) waren für die Kinder anscheinend sehr ansprechend und somit leicht zu merken. Wie auch von Testleitern berichtet wurde, gab es Kinder, die sich nach der ersten Testung für die zweite Testung „schlau machten“ und gewisse Dinge übten. Da diese Übungsgewinne jedoch nach Leistungsstand, Lernfähigkeit und Gedächtnis interindividuell verschieden ausfallen, wie Amelang und Schmidt-Atzert (2006) feststellten, werden die Stabilitäts-Koeffizienten verzerrt.

Nach den Reliabilitätsanforderungen (Wagner-Menghin, 2003) dass eine Reliabilität höher als 0.75 sein müsse, wären, wenn man die Korrelationskoeffizienten der T-Werte betrachtet für die Stichprobe nach dem Intervall von maximal 3 Monaten die Untertests **3, 8** und **11** und der Prozentrang des zweitniedrigsten T-Wertes als zuverlässig zu bezeichnen. Die niedrigen Korrelationskoeffizienten des Untertest **7** ($r_{tt}=0.41$ und $r_{tt}=0.44$) können eventuell durch Lerneffekte begründet werden, da anhand der Deskriptivstatistik ein hoher Leistungszuwachs in diesem Untertest festgestellt werden kann, der womöglich nicht bei allen Testpersonen gleich verlaufen ist (vgl. z.B. Amelang & Schmidt-Atzert, 2006). Es ist gut vorstellbar, dass die Symbole, die zu kodieren sind, von einigen Testpersonen über ein bis drei Monate leichter im Gedächtnis behalten werden konnten als von anderen. Nach Betrachtung der Korrelationskoeffizienten der Fähigkeitsparameter erfüllen mehrere Untertests die Bedingung, dass die Reliabilität höher als 0.75 sein muss. Für das Intervall von drei Monaten trifft dies für Untertests **1, 3, 4, 6, 8, 9, 10** und **11** zu.

Nach dem Intervall von 0.5 bis 1.92 Jahren ist in der betreffenden Stichprobe kein Korrelationskoeffizient der T-Werte höher als 0.75, außer dem des zweitniedrigsten T-Werts. Für das Intervall von 0.5 bis 1.92 Monaten sind die Korrelationskoeffizienten der Fähigkeitsparameter der Untertests **1, 3, 6, 8, 10** und **11** größer als 0.75. Hier wird durch die Korrelationskoeffizienten jedoch nicht die

Stabilität der Testleistungen der Personen im Vergleich zu anderen Testpersonen berechnet. Diese hohen Korrelationskoeffizienten der Fähigkeitsparameter resultieren eher daraus, dass die Leistungszuwächse der Testpersonen dieser Stichprobe über das Zeitintervall untereinander ähnlich sind. Man kann daraus schließen, dass in den genannten Untertests die Entwicklungsverläufe der Testpersonen, die ein Alter von sechs bis zwölf Jahren hatten, homogen verliefen. Man kann von großen Entwicklungsfortschritten nach einem halben Jahr bis zu einem Jahr und elf Monaten ausgehen, da die deskriptiven Statistiken in Tabelle 28 und Tabelle 29 zeigen, dass die Mittelwerte der Fähigkeitsparameter in diesen Untertests in der Zweittestung um einiges höher sind als in der Ersttestung. Im Untertest 1 liegen z.B. zwischen den Mittelwerten der Fähigkeitsparameter in der ersten Testung und in der zweiten Testung gerundet 1.65 Einheiten. Dies kann wiederum durch das Überlappungsmodell (Anderson, 1939) und durch dessen nachfolgende Studien begründet werden. Die Testpersonen waren alle Klienten von Beratungsstellen, die zum Großteil zur Abklärung von Hochbegabung vorgestellt wurden. In Studien wurden bei besseren Probanden größere längsschnittliche Zugewinne festgestellt, bei Probanden mit niedrigeren Testwerten jedoch eher geringere Fortschritte (vgl. Amelang, 2006). Die meisten Probanden erbrachten bereits bei der Ersttestung eher höhere Leistungen und hatten auch, konform mit anderen Studien gleich gute Leistungszuwächse.

Nach dem Intervall von 2 bis 2.67 Jahren liegt der Korrelationskoeffizient des Untertests 1 über 0.75. In dieser Stichprobe war das Alltagswissen ein stabiles Merkmal. Leider kann diese Aussage nicht generalisiert werden, da im kürzeren Intervall von 0.5 bis 1.92 Jahren und im längeren Intervall von 2.75 bis 6.42 Jahren im Untertest 1 kein Korrelationskoeffizient über 0.75 berechnet wurde. In demselben Intervall von 2 bis 2.67 Jahren können die anhand der Korrelationskoeffizienten der Fähigkeitsparameter die Untertests 1, 3, 6 und 10 als reliabel bezeichnet werden, da sie über 0.75 liegen. Diese hohen Korrelationen der Fähigkeitsparameter können wie oben begründet werden, da es sich um eine Stichprobe mit hohen Leistungen handelt. Dass diese Koeffizienten jedoch geringer sind, als die der kürzeren Intervalle ist konform mit den berichteten Studien, dass kürzere Intervalle zu höheren Koeffizienten führen als längere (vgl. Amelang et al. 2006).

Im längsten Intervall von 2.75 bis zu 6.42 Jahren liegen alle Korrelationskoeffizienten der T-Werte der Untertests und der Testkennwerte unter 0.75. Auch die Korrelationskoeffizienten der Fähigkeitsparameter der Untertests liegen nach diesem längsten Zeitintervall unter 0.75. Dies ist ebenfalls konform mit den oben berichteten Studien, dass die Korrelationskoeffizienten abnehmen, umso länger die Zeitintervalle zwischen den Testungen sind (vgl. Amelang et al. 2006).

Dass der Untertest **5 Unmittelbares Reproduzieren-numerisch vorwärts** und **rückwärts** zwar am selben Tag recht hohe Korrelationskoeffizienten aufweist, bei der Langzeitstabilitätsuntersuchung jedoch stets niedrigere Werte aufzeigt, könnte daran liegen, dass die Informationsverarbeitungskapazität, die hier erfasst wird, eventuell stärker von der Tagesverfassung abhängt und auch eher Entwicklungsveränderungen durchläuft als die grundlegenden Fähigkeiten anderer Untertests (vgl. z.B. Dixon, 2003). Der Untertest **7** liefert sowohl für das **Kodieren** als auch für das **Assoziieren** mit länger werdendem Intervall niedrige Korrelationskoeffizienten. Auch dieser Test beruht zum Großteil auf Messung der Informationsverarbeitungskapazität, vor allem das Assoziieren.

Ebenfalls niedrige Korrelationskoeffizienten liefern die Untertests **2, 4** und **11**. Hier kann auch vermutet werden, dass das Verstehen sozialen Geschehens, das in allen drei Untertests eine wesentliche Rolle spielt, schwankende Entwicklungsveränderungen durchläuft.

Eine mögliche Erklärung für die niedrigen Korrelationskoeffizienten der Zusatztests könnte sein, dass die Testleiter Probleme bei der Testvorgabe und bei der Testauswertung hatten. Die Zusatztests wurden nur von Psychologiestudierenden vorgegeben, und vor allem die Zusatztests **5a** und **10a** sind in der Vorgabe etwas komplexer als die anderen. Andere Erklärungen könnten sein, dass diese Tests am Ende der Testung vorgegeben wurden, als die Testpersonen schon ermüdet waren und eventuell auch nicht mehr so motiviert waren, wie am Anfang der Testung. Außerdem können hier auch Erinnerungs- und Lerneffekte zum Tragen gekommen sein, da das Zeitintervall zwischen den beiden Testungen kurz war.

Die hier berechneten Korrelationskoeffizienten des AID 2 sind durchaus vergleichbar mit denen von Madest (1989, zitiert nach Melchers & Preuß, 1994) für das K-ABC, die jedoch nur nach einem Zeitintervall von einem Monat berechnet wurden. Leider können keine aussagekräftigen Vergleiche mit der Stabilität des HAWIK-IV durchgeführt werden, da im Manual die Retest-Reliabilität nur für drei Untertests angegeben wird (Petermann & Petermann, 2007). Korrelationskoeffizienten von $r_{tt}=0.9$, über die in vielen Studien berichtet wird (vgl. z.B. Amelang, 2006), beziehen sich zumeist auf die Stabilität eines Gesamt IQs und sind deshalb nicht mit den Korrelationskoeffizienten der einzelnen Untertests des AID 2 vergleichbar. Wegen des förderdiagnostischen Zugangs wird jedoch für den AID 2 kein Gesamt IQ berechnet. Dieser würde Defizite in einem Leistungsbereich durch Stärken in einem anderen Bereich ausgleichen und hätte somit keine Aussage für die Diagnose von Teilleistungsstörungen (Kubinger, 2006).

Für die Praxis kann aus diesen Ergebnissen geschlossen werden, dass die Testwerte mit einer gewissen Sicherheit nur für kurze Intervalle bis zu drei Monaten vorhergesagt werden können. Im Zweifelsfall sollte eine nochmalige Testung zur Überprüfung von Veränderungen durchgeführt werden, auch um situationsbedingte Leistungsschwankungen besser ausschließen zu können. Eine neue Erkenntnis ist, dass das Ausmaß der zweitniedrigsten Untertestleistung einen recht stabilen Messwert darstellt.

Eine nochmalige Untersuchung speziell der Zusatztests wäre wünschenswert. Eventuell könnten dann die Zusatztests am Beginn der Testung vorgegeben werden, oder es könnten überhaupt nur die Zusatztests vorgegeben werden, ohne die anderen Subtests vorzugeben, um Müdigkeitseffekte zu verhindern. Außerdem steht auch eine Untersuchung der Langzeitstabilität der Zusatztests aus, da diese hier nur nach maximal drei Monaten nochmals vorgegeben wurden.

Es wäre weiters noch interessant eine unmittelbare Wiederholung der Subtests **3**, **4**, **10** und **11** in derselben Sitzung durchzuführen, um die Kurzzeitstabilität zu untersuchen, da diese hier nicht untersucht wurde. Bei den Untertests **2**, **7** und **8** ist diese wegen Erinnerungseffekten und da es keine Parallelförmigkeit gibt wahrscheinlich nicht sinnvoll. Im Untertest **7** würde es wahrscheinlich zu großen Zunahmen in der

Merkleistung beim Assoziieren kommen. Die Entwicklung einer Parallelförm für diese Untertests wäre wünschenswert.

Auch eine genaue Analyse der Stabilität über die verschiedenen Altersstufen könnte weitere interessante Ergebnisse liefern, da altersspezifische Entwicklungsunterschiede in den verschiedenen Fähigkeiten und Fertigkeiten, die mittels der Untertests gemessen werden, festgestellt werden könnten.

4 Zusammenfassung

Ziel dieser Arbeit ist es, die Stabilität des Adaptiven Intelligenz Diagnostikums 2 (AID 2) von Kubinger und Wurst (2000) zu überprüfen.

Intelligenz ist ein weiter Begriff. In der vorliegenden Arbeit wird ein Überblick über Intelligenztests gegeben. Im Zusammenhang mit dem AID 2 werden unter Intelligenz kognitive Voraussetzungen verstanden, die zum Wissenserwerb dienen (Kubinger & Wurst, 2000). Forschungen zur Entwicklung von Intelligenz gibt es zahlreich. Neben der Anlage-Umwelt Debatte gibt es sechs Perspektiven, um die Entwicklung zu betrachten: die psychometrische, die Sichtweise nach Piaget, die Neo-Piaget Perspektive, die Informationsverarbeitungstheorie, die „klassische“ Lerntheorie und die kontextuelle Perspektive (Berg, 1992). In Studien zur Intelligenz konnten Stabilitätskoeffizienten von $r_{tt}=0.9$ errechnet werden. Mit dem Alter der Testpersonen steigen auch die Stabilitätskoeffizienten (Amelang, 2006).

In dieser Studie wurde die Stabilität über die Retest-Reliabilität und die Paralleltest-Reliabilität berechnet, die neben der Inneren Konsistenz Methoden der klassischen Testtheorie darstellen. Die Reliabilität ist der statistische Kennwert, der den Grad der Messgenauigkeit des Testwerts bestimmt (Bortz und Döring, 2005). Die Geschichte der Intelligenztests ist bereits hundert Jahre alt. In der neuen, probabilistischen Testtheorie gibt es bereits ganz andere Methoden, durch die die Messgenauigkeit eines Verfahrens berechnet werden kann (Kubinger, 2006). Diese wurden auch in vergleichbaren Tests verwendet. Vergleichsmöglichkeiten mit Studien über Retestreliabilitäten sind in Manualen von aktuellen, gleichwertigen Testbatterien spärlich (Kubinger & Wurst, 2000).

Der AID 2 ist eine Testbatterie für Kinder und Jugendliche im Alter von 6 Jahren bis zu 15 Jahren und elf Monaten. Mithilfe dieser Testbatterie wird eine Profilinterpretation ermöglicht, es kann ein Screening auf ausgewählte Teilleistungsstörungen vollzogen werden und es gibt weitere vielfältige Einsatzmöglichkeiten (Kubinger & Wurst, 2000).

Da für die meisten Untertests des AID 2 das dichotome logistische Modell von Rasch gilt, wurden über die probabilistische Testtheorie die Standardschätzfehler berechnet. Auch Studien zur split-half-Reliabilität und Retestreliabilität wurden bereits durchgeführt (Kubinger & Wurst, 2000). Die Untertests sind abwechslungsreich und heißen **1 Alltagswissen**, **2 Realitätssicherheit**, **3 Angewandtes Rechnen**, **4 Soziale und Sachliche Folgerichtigkeit**, **5 Unmittelbares Reproduzieren**_{-numerisch}, **6 Synonyme Finden**, **7 Kodieren und Assoziieren**, **8 Antizipieren und Kombinieren**, **9 Funktionen Abstrahieren**, **10 Analysieren und Synthetisieren**_{-abstrakt}, **11 Soziales Erfassen und Sachliches Reflektieren**, **5a Unmittelbares Reproduzieren**, **5b Merken und Einprägen** und **10a Strukturieren**_{-visumotorisch}.

Für diese Untersuchung wurden Daten von 286 Retestungen gesammelt. Diese wurden in Intervalle unterteilt, um Veränderungen in der Stabilität mit zunehmender Länge der Intervalle zu untersuchen. Dabei umfasste das erste Intervall ca. ein Monat, das nächste ein bis drei Monate, dann ein halbes Jahr bis zu einem Jahr und neun Monate, danach zwei Jahre bis zwei Jahre und 8 Monate und das längste Intervall umfasste zwei Jahre und neun Monate bis sechs Jahre und fünf Monate. Bei einigen Retestungen wurden die vier Subtests **5 Unmittelbares Reproduzieren**_{-numerisch} **vorwärts** und **rückwärts**, **1 Alltagswissen**, **6 Synonyme finden** und **9 Funktionen Abstrahieren** ein zweites Mal vorgegeben, um die unmittelbare Stabilität zu berechnen.

Die kürzesten Intervalle weisen im Einklang mit den Forschungsergebnissen die höchsten Korrelationskoeffizienten auf. Diese sind in der Höhe von ca. $r_{tt}=0.80$. Mit der Länge der Intervalle nimmt auch die Höhe der berechneten Korrelationskoeffizienten ab. Nach einem Intervall von sechs Monaten können die meisten Untertestergebnisse nur mehr mit ungenügender Genauigkeit vorhergesagt werden. Eine nochmalige Testung ist daher empfehlenswert.

In Zukunft sollten die Zusatztests **5a**, **5b** und **10a** nochmals untersucht werden. Auch eine Wiederholung weiterer Subtests in derselben Sitzung steht noch aus, um die unmittelbare Stabilität zu berechnen. Eine genaue Analyse der Stabilität über verschiedene Altersstufen wäre interessant.

Literaturverzeichnis

- Aebli, H. (1963). *Über die geistige Entwicklung des Kindes*. Stuttgart: Ernst Klett Verlag.
- Alsaker, F. D. (1992). Modeling Quantitative Developmental Change. In Asendorpf, J. B. & Valsiner (Eds.), *Stability and Change in Development. A study of Methodological Reasoning* (pp.88-109). California: SAGE.
- Amelang, M., Bartussek, D., Stemmler, G. & Hagemann, D. (Hrsg.) (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6. Auflage). Stuttgart: Kohlhammer.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Interventionen* (4., vollständig überarbeitete und erweiterte Aufl.). Heidelberg: Springer.
- Amthauer, R. (1957). Über die Prüfung der Zuverlässigkeit von Tests – erörtert am IST. *Psychologische Rundschau*, 8, 165-171.
- Anderson, J. E. (1939). The limitation of infant and preschool tests in the measurement of intelligence. *Journal of psychology*, 8, 351-379.
- Anderson, M. (1992). *Intelligence and Development: A cognitive Theory*. Oxford: Blackwell.
- Antretter, S. (2001). *Einsetzbarkeit des AID 2 bei Jugendlichen außerhalb des Altersgeltungsbereiches: Differenziertheit und Akzeptanz*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Berg, C. A. (1992). Perspectives for viewing intellectual development throughout the life course. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual Development* (pp. 1-15). New York: Cambridge University Press.
- Binet, A. & Simon, T. (1904). Methodes nouvelles pour le diagnostique du niveau intellectuel des anormaux. *Année Psychologique*, 11, 191-244.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Bortz, J. & Döring, N. (2005). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Brody, E. B. & Brody, N. (1976). *Intelligence. Nature, Determinants and Consequences*. London: Academic Press.

- Canfield, R. L. & Ceci, S. J. (1992). Integrating learning into a theory of intellectual development. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual Development* (pp. 278-300). New York: Cambridge University Press.
- Case, R. (1992). Neo-Piagetian theories of child development. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual Development* (pp. 161-196). New York: Cambridge University Press.
- Cattell, R. B. (1963). Theory of fluid and cristallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1-22.
- Catron, D. W. (1978). Immediate test-retest changes in WAIS scores among college males. *Psychological Reports*, 43 (1), 279-290
- Colom, R., Lluís-Font, J. M. & Andrés-Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence*, 33, 83-91.
- Daseking, M., Janke, N. & Petermann, F. (2006). Intelligenzdiagnostik. *Monatsschrift Kinderheilkunde*, 4, 314-319.
- Daseking, M., Petermann, U. & Petermann, F. (2007). Intelligenzdiagnostik mit dem HAWIK-IV. *Kindheit und Entwicklung*, 16 (4), 250-259.
- Dennis, W. & Najarian, P. (1957). Infant development under environmental handicap. *Psychological Monographs*, 71, 436
- Dietrich, K. (1972). *Intelligenz lässt sich lernen. Was Eltern über die geistige Entwicklung ihrer Kinder wissen müssen*. Stuttgart: Ernst Klett.
- Dixon, R. A. (2003). Themes in the Aging of Intelligence: Robust Decline With Intriguing Possibilities. In R. J. Sternberg, J. Lautrey & T. I. Lubart (Eds.), *Models of Intelligence: International Perspectives* (pp. 151-167). Washington: American Psychological Association.
- Ehrle, C. M. (1963). *Das geistige Erwachen des Kindes. Aufgaben der Erziehung in kindlichen Entwicklungsphasen*. Freiburg: Herder.
- Eysenck, H. J. (2004). *Die IQ-Bibel. Intelligenz verstehen und messen*. (K. Neff, Übers.) Stuttgart: Klett-Cotta. (Original erschienen 1998: *Intelligence. A New Look*)
- Flynn, J. R. (1984). The Mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95 (1), 29-51.
- Fröhlich, W. D. (2002). *Wörterbuch Psychologie*. München: dtv.

- Funke, J. & Vaterrodt-Plünnecke, B. (1998). *Was ist Intelligenz?* München: Beck.
- Gardner, M. K. & Clark, E. (1992). The psychometric perspective on intellectual development in childhood and adolescence. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual Development* (pp. 16-43). New York: Cambridge University Press.
- Garfinkel, R. & Thorndike, R. L. (1976). Binet item difficulty then and now. *Child Development*, 47 (3), 959-965.
- Gebetsberger, N. (2004). *Validität der Teilleistungsstörung Merkfähigkeit im AID 2*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Ginsburg, H. P. & Oppen, S. (1998). *Piagets Theorie der geistigen Entwicklung* (H. Kober & G. Hanenberg, Übers.). Stuttgart: Klett-Cotta. (Original erschienen 1988: Piaget's theory of intellectual development. An introduction)
- Grigorenko, E. L. (2003). Individual Fluctuations in Intellectual Functioning: Selectes Links Between Nutrition and the Mind. In R. J. Sternberg, J. Lautrey & T. I. Lubart (Eds.), *Models of Intelligence: International Perspectives* (pp. 91-115). Washington: American Psychological Association.
- Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53 (4), 267-293.
- Guthke, J. (1990). Grundzüge der Testkonstruktion und Testauswertung. In J. Guthke, H. R. Bötticher & L. Sprung (Hrsg.), *Psychodiagnostik* (Bd. 1) (S. 201-324). Berlin: Deutscher Verlag der Wissenschaften.
- Guthke, J. (1991). Leistungsmethoden. In J. Guthke, H. R. Bötticher & L. Sprung (Hrsg.), *Psychodiagnostik* (Bd. 2) (S. 172-244). Berlin: Deutscher Verlag der Wissenschaften.
- Guthke, J. (1996). *Intelligenz im Test: Wege der psychologischen Intelligenzdiagnostik*. Göttingen: Vandenhoeck & Ruprecht.
- Hay, D. (1999). The developmental genetics of intelligence. In M. Anderson (Ed.), *The Development of Intelligence* (pp. 75-104). East Sussex: Psychology Press Ltd.
- Hagmann-von Arx, P., Meyer, C. S. & Grob, A. (2008). Intelligenz- und Entwicklungsdiagnostik im deutschen Sprachraum. *Kindheit und Entwicklung*, 17 (4), 232-242.
- Heller, K. A. (Hrsg.) (2000). *Begabungsdagnostik in der Schul- und Erziehungsberatung* (2., vollst. überarb. Aufl.). Bern: Hans Huber Verlag.

- Holocher-Ertl, S., Kubinger, K. D. & Hohensinn, C. (2008). Hochbegabungsdagnostik: HAWIK-IV oder AID 2. *Kindheit und Entwicklung* 17 (2), 99-106.
- Howard, R. W. (1999). Preliminary real-world evidence that average human Intelligence really is rising. *Intelligence*, 27 (3), 235-250
- Jacobs, C., Heubrock, D. & Petermann, F. (2003). Testinformationen, *Diagnostica*, 49 (4), 184-188.
- Jäger, O. A. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35 (1), 21-35.
- Kail, R. & Bisanz, J. (1992). The information-processing perspective on cognitive development in childhood and adolescence. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual Development* (pp. 229-260). New York: Cambridge University Press.
- Knab, Christian (1996). *Langzeitstabilität des AID. Analyse der Retest-Reliabilität des AID über längere Zeiträume*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Kubinger, K. D. (1989). Die Optimierung der Messgenauigkeit beim „branched“-adaptiven Testen. In K. D. Kubinger (Hrsg.), *Moderne Testtheorie* (S. 187-218). Weinheim und Basel: Beltz.
- Kubinger, K. D. (1996a). *Einführung in die Psychologische Diagnostik*. Weinheim: PVU.
- Kubinger, K. D. (1996b). Methoden der Psychologischen Diagnostik. In E. Erdfelder, R., Mausfeld, T., Meiser, & G., Rudinger (Hrsg.), *Handbuch Quantitative Methoden* (S. 567-576). Weinheim: PVU.
- Kubinger, K. D. (1997). *Psychologische Diagnostik und Intervention in Fallbeispielen*. Weinheim. PVU
- Kubinger, K. D. (2003a). Adaptives Testen. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 1-9). Berlin: Beltz.
- Kubinger, K. D. (2003b). Gütekriterien. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 195-204). Berlin: Beltz.
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kubinger, K. D. & Wurst, E. (1991). *Adaptives Intelligenz Diagnostikum (AID)*. Weinheim: Beltz Test.

- Kubinger, K. D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum 2 (AID 2)* (2. überarb. Aufl.). Goettingen: Beltz Test.
- Lee, E. S. (1951). Negro intelligence and selective migration: a Philadelphia test of the Kingberg hypothesis. *American Social Review*, 16, 227-233.
- Leiss, U. (2003). *Erstellung und Erprobung einer optimalen Strategie zur Diagnostik von Teilleistungsschwächen*. Unveröff. Diss., Universität, Wien.
- Lienert, G. A. & Raatz, U. (1998). Testaufbau und Testanalyse (6. Auflage). Weinheim: PVU.
- Maryschka, C. (2000). AIDScore. Auswertungsprogramm zum AID 2. (Version 1.02) [Computer software]. Wien: Beltz Test.
- McCall, R. B. (1981). Nature-nurture and the two realms of development: A purposed integration with respect to mental development. *Child Development*, 52, 1-12.
- McCall, R. B., Hogarty, P. S. & Hurlburt, N. (1972). Transitions in infant sensorimotor development and the prediction of childhood IQ. *American Psychologist*, 27, 728-748.
- Melchers, P. & Preuß, U. (1994). *Kaufmann-Assessment Battery for Children (K-ABC), Deutsche Fassung* (2. korrigierte und ergänzte Aufl.). Frankfurt am Main: Swets & Zeitlinger.
- Milanovic, M. (1998). *Erfassung der Teilkomponente Serialität im AID 2*. Unveröff. Dipl.Arbeit. Universität, Wien.
- Mingroni, M. A. (2004). The secular rise in IQ: Giving heterosis a closer look. *Intelligence*, 32, 65-83.
- Nettlebeck, T. & Wilson, C. D. (2004). The Flynn effect: Smarter not faster. *Intelligence*, 32, 85-93.
- Owens, W. A. (1963). Age and mental abilities: A longitudinal study. *Genetic Psychology Monographs*, 48, 3-54.
- Pamperl, E. (2004). *Validierungsstudie zum AID 2 hinsichtlich der Teilkomponente Serialität*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Petermann, F. (2006) Intelligenzdiagnostik. *Kindheit und Entwicklung*, 15 (2), 71-75.
- Petermann, F. & Petermann, U. (Hrsg.). (2007). *Hamburg Wechsler Intelligenztest für Kinder – IV. Übersetzung und Adaptation der WISC-IV von David Wechsler*. Bern: Huber.

- Petermann, F. & Petermann, U. (2008). HAWIK-IV. *Kindheit und Entwicklung*, 17 (2), 71-75.
- Petrill, S. A. (2003). The Development of Intelligence: Behavioral Genetic Approaches. In R. J. Sternberg, J. Lautrey & T. I. Lubart (Eds.), *Models of Intelligence: International Perspectives* (pp. 81-89). Washington: American Psychological Association.
- Reznick, J. S. & Corley, R. (1999). What twins can tell us about the development of intelligence-a case study. In M. Anderson (Ed.), *The Development of Intelligence* (pp. 104-136). East Sussex: Psychology Press Ltd
- Rist, F. & Dirksmeier, C. (2001). Leistungsdiagnostik bei psychischen Störungen. In R.-D. Stieglitz, U. Baumann & H. J. Freyberger (Hrsg.), *Psychodiagnostik in Klinischer Psychologie, Psychiatrie, Psychotherapie* (S. 145-158). Stuttgart: Georg Thieme Verlag.
- Rodgers, J. L. (1998): A critique of the Flynn Effect: Massive IQ gains, methodological artefacts, or both? *Intelligence*, 26, 337-356.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. überarbeitete und erweiterte Aufl.). Bern: Huber.
- Schlagheck, W. & Petermann, F. (2006). Hochbegabtendiagnostik mit dem AID 2. *Kindheit und Entwicklung*, 15 (2), 93-99.
- Schröder, E. (2002). Selektive Differenzierung in der kognitiven Entwicklung: Eine längsschnittliche Analyse zur Entstehung individueller Unterschiede beim aussagenlogischen Denken. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 34 (3), 136-148.
- Schubhart, S. (2008). *Katamnestic Validierung des „Wiener Diagnosemodells zum Hochleistungspotenzial“*. Unveröff. Dipl. Arbeit, Universität, Wien.
- Simon, T., Binet, A. (1904). Sur la nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence. *L'Année psychologique*, 11, 163-190.
- Sommer, R. & Sommer, B. A. (1983). Early intervention, IQ, and psychology textbooks. *American Psychologist*, 38, 982-985.
- Sontag, L. W. , Barker, C. T. & Nelson, V. L. (1958). Mental growth and personality development: A longitudinal study. *Monographs of the Society for Research in Child Development*, 23, 1-143.
- Sternberg, R. J. (2002a). The Search for Criteria: Why Study the Evolution of Intelligence? In R. J. Sternberg & J. C. Kaufman (Eds.), *The Evolution of Intelligence* (pp. 1-7). Mahwah, NJ: Lawrence Erlbaum Associates.

- Sterberg, R. J. (2002b). The Theory of Successful Intelligence. In R. J. Sternberg, & E. L. Grigorenko (Eds.), *The General Factor of Intelligence: How General Is It?* Mahwah, NJ: Lawrence Erlbaum Associates.
- Sternberg, R. J. & Gardner, M. K. (1982). A Componential Interpretation of the General Factor in Human Intelligence. In H. J. Eysenck (Ed.), *A Model for Intelligence* (pp. 231-254). Berlin (u.a.): Springer.
- Steyer, R. & Eid, M. (1993). *Messen und Testen*. Berlin: Springer-Verlag.
- Sundet, J. M., Barlaug, D. G. & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32, 349-362.
- Teasdale, T. W. & Owen, D. R. (2000). Forty-year secular trends in cognitive abilities. *Intelligence*, 28, 115-120.
- Tewes, U., Rossmann, P. & Schallberger, U. (Hrsg.). (1999). *Hamburg-Wechsler Intelligenztest für Kinder – Dritte Auflage (HAWIK-III)*. Bern: Huber.
- Tuddenham, R. D. (1978). Soldier intelligence in World War I and World War II. *American Psychologist*, 3 (2), 54-56
- Wagner-Menghin, M. M. (2003). Messgenauigkeit. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 282-286). Berlin: Beltz.
- Weinert, F. E. & Hany, E. A. (2003). The Stability of Individual Differences in Intellectual Development: Empirical Evidence, Theoretical Problems and New Research. In R. J. Sternberg, J. Lautrey & T. I. Lubart (Eds.), *Models of Intelligence: International Perspectives* (pp. 169-181). Washington: American Psychological Association.
- Wertsch, J. V. & Kanner, B. G. (1992). A sociocultural approach to intellectual development. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual Development* (pp. 328-349). New York: Cambridge University Press.
- Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe
- Weinstein, S. & Teuber, H. L. (1975). Effects on penetrating brain injury on intelligence test scores. *Science*, 125, 1036-1037
- Wheeler, L. R. A. (1942). A comparative study of the intelligence of East Tennessee mountain children. *Journal of Educational Psychology*, 33 (5), 321-334.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I. & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32, 509-537.

5 Anhang

5.1 Abstract

Ziel dieser Arbeit war es, die Stabilität des Adaptiven Intelligenz Diagnostikums 2 (AID 2) von Kubinger und Wurst (2000) zu überprüfen. Anhand einer Stichprobe von 286 Retestungen wurden sechs unterschiedlich lange Intervalle gebildet, um die Veränderungen in der Stabilität mit zunehmender Länge der Intervalle zu untersuchen. Dabei umfasste das erste Intervall ca. einen Monat, das nächste einen bis drei Monate, dann ein halbes Jahr bis zu einem Jahr und elf Monate, danach zwei Jahre bis zwei Jahre und acht Monate und das längste Intervall umfasste zwei Jahre und neun Monate bis sechs Jahre und fünf Monate. Bei einigen Retestungen wurden vier Subtests im Anschluss an die Testung noch ein zweites Mal vorgegeben, um die unmittelbare Stabilität zu berechnen. Die kürzesten Intervalle weisen im Einklang mit den Forschungsergebnissen die höchsten Korrelationskoeffizienten um ca. $r_{tt}=0.80$ auf. Mit Zunahme der Länge der Intervalle nimmt die Höhe der berechneten Korrelationskoeffizienten ab. Nach einem Intervall von sechs Monaten sind die meisten Korrelationskoeffizienten der T-Werte der Subtests unter $r_{tt}=0.75$. Die Testwerte, die ein Psychologe oder eine Psychologin zur Beantwortung einer Fragestellung heranzieht, sollten demnach innerhalb des letzten halben Jahres erhoben worden sein. Danach sollte erneut eine AID 2 Testung durchgeführt werden.

It was the aim of this paper to figure out the stability of the Adaptive Intelligence Diagnosticum 2 (AID 2) by Kubinger and Wurst (2000). Samples were built out of 286 retests with different time periods, to be able to calculate the stability of the test after increasing time periods. The shortest period was about one month, the next one up to three months, then half a year to one year and eleven months, then two years to two years and eight months and the longest period was from two years and nine months to six years and five months. Four subtests were once again repeated after some retests, to find out the stability of the test at the same day. According to the results of the latest research the shortest periods show the highest coefficient of correlation, about $r_{tt}=0.80$. The longer the periods are, the smaller are the coefficients of correlation. After a time period of six months the reliability of the most subtest scores is not higher than $r_{tt}=0.75$. Therefore AID 2 test scores are only reliable for half a year.

5.2 E-Mail an PsychologInnen des BÖP

Sehr geehrte/r Frau/Herr Mag./Dr. XY!

Ich studiere Psychologie an der Universität Wien. Im Rahmen meiner Diplomarbeit bei Univ.-Prof. Dr. Mag. Klaus Kubinger untersuche ich die Stabilität des AID 2, um diesen im Sinne der Wissenschaft verbessern zu können. Dafür benötige ich Wiederholungstestungen von Kindern mit dem AID 2.

Im Anhang schicke ich Ihnen genauere Angaben: mein Ansuchen, meine Verpflichtungserklärung, den Untersuchungsplan und einen Elternbrief, den Sie bei Bedarf verwenden können.

Sind Sie bereit mein Vorhaben zu unterstützen?

Ich danke Ihnen schon im Voraus für Ihre Bemühungen und freue mich auf Ihre baldige Antwort!

Mit freundlichen Grüßen,
Claudia Parfuss

Handy.: 06763901208

Anschrift: Buchengasse 5, 3001 Mauerbach

5.3 Attachment des e-Mails an PsychologInnen

Ansuchen um Unterstützung meiner Diplomarbeit

Sehr geehrte PsychologInnen,

Ich studiere Psychologie an der Universität Wien und das Thema meiner Diplomarbeit bei Herrn Univ.-Prof. Dr. Mag. Klaus Kubinger ist die Stabilitätsmessung des **Adaptiven Intelligenz Diagnostikums 2** (AID 2) von Kubinger und Wurst (2000). Dafür benötige ich Retestungen von Kindern im Alter von 6 Jahren bis 15 Jahren und 11 Monaten.

Bei bereits vorhandenen Zweittestungen wäre es eine große Hilfe für meine Diplomarbeit, wenn Sie mir die anonymisierten Testergebnisse der Kinder zur Verfügung stellen würden.

Wenn Sie erst eine AID 2 Testung mit einem Kind oder Jugendlichen gemacht haben, könnten Sie den Eltern der Klienten meinen Elternbrief, den ich beigelegt habe, senden um sie zu einer Wiederholungstestung einzuladen. Ich versichere die Anonymität Ihrer Klienten zu wahren und die Daten zu schützen.

Diese Arbeit ist wichtig, um die Qualität des AID 2 überprüfen und gegebenenfalls verbessern zu können.

Um die Untersuchung möglichst schnell abschließen zu können ist eine baldige Zusage Ihrer Mitarbeit sehr wichtig.

Ich danke Ihnen schon im Voraus für Ihre Hilfe und Unterstützung und erwarte Ihre Antwort!

Beste Grüße,
Claudia Parfuss

Buchengasse 5
3001 Mauerbach
Handy: 06763901208
e-Mail: claudiaparfuss@gmx.at

Verpflichtungserklärung

Ich verpflichte mich hiermit, die erhobenen Daten ausschließlich im Rahmen meiner Diplomarbeit für wissenschaftliche Zwecke zu verwenden und diese an niemanden weiterzugeben.

Weiters verpflichte ich mich dazu, falls Sie es wünschen, Ihnen nach Beendigung meiner Untersuchung eine Kurzfassung der Ergebnisse zu übermitteln.

Untersuchungsplan

Ziel der Arbeit

Diese Arbeit soll folgende Frage beantworten:

- > Wie hoch sind die Stabilitäts-/Retest-Reliabilitäts-Koeffizienten des AID 2 (Kubinger & Wurst, 2000), die mittels Retest-Methode nach unterschiedlichen Untersuchungsintervallen berechnet werden?

Das Untersuchungsintervall soll dabei unterschiedliche Zeitspannen umfassen.

Der Untertest **5 Unmittelbares Reproduzieren** soll in einigen Retestungen zweimal vorgegeben werden. Die Untertests **1 Alltagswissen**, **6 Synonyme finden** und **9 Funktionen Abstrahieren** sollen ebenfalls in einigen Retestungen ein zweites Mal, jedoch in der Parallelform vorgegeben werden. Diese Subtests wurden ausgewählt, da die Vorgabe nicht viel Zeit beansprucht und deshalb die zweifache Vorgabe an ein Kind in derselben Sitzung zumutbar ist.

Außerdem soll in weiteren Fällen der gesamte Test im Intervall von einem Monat, einem Jahr, zwei Jahren, drei Jahren und mehr als vier Jahren wiederholt werden.

Stichprobe

Bei der Stichprobe soll darauf geachtet werden, dass die Kinder für die Population möglichst repräsentativ sind. Dabei sollen das Geschlecht und das Alter zwischen 6 und 16 Jahren gleichverteilt sein. Mögliche Testpersonen sind z.B.:

- > Kontrollgruppen aus Diplomarbeiten
- > Klienten von Beratungsstellen
- > Testpersonen des AID 2 Zertifizierungskurses

Vorgesehener Untersuchungszeitraum

Alle Retestungen sollten spätestens im November 2008 abgeschlossen sein.

Durchführung und Arbeitsaufwand der Untersuchung

Über Institutionen und praktizierende PsychologInnen werden potentielle Teilnehmer für die Untersuchung akquiriert.

Die Kinder, die erst einmal getestet wurden, werden von mir und gegebenenfalls von KollegInnen (DiplomandInnen aus dem Bereich der Psychologischen Diagnostik und TeilnehmerInnen des AID 2 Zertifizierungskurses) ein zweites Mal getestet. Die Testungen sollen je nach Bedürfnis der Klienten entweder bei den Kindern zu Hause, in der Schule, an den Institutionen oder samstags an der Testberatungsstelle der Universität Wien stattfinden.

Es soll eine kurze schriftliche Rückmeldung über die Testergebnisse erfolgen.

Im Allgemeinen möchte ich zum Schluss noch festhalten, dass ich bereit bin flexibel auf Anregungen, Wünsche und Beschwerden von Teilnehmern einzugehen.

5.4 Elternbrief der Autorin

Sehr geehrte Eltern, sehr geehrte Erziehungsberechtigte!

Im Rahmen meiner Diplomarbeit an der Universität Wien bei Univ.-Prof. Dr. Mag. Klaus Kubinger (dem Entwickler des **Adaptiven Intelligenz Diagnostikums 2**) untersuche ich die Stabilität dieses Intelligenztests für Kinder. Dabei wird gemessen, wie zuverlässig die Testergebnisse Ihres Kindes über einen gewissen Zeitraum sind.

Da Ihr Kind bereits an einer **AID 2** Untersuchung teilgenommen hat, lade ich Ihr Kind herzlich zu einer zweiten Testung ein. So erfahren Sie den aktuellen Leistungsstand Ihres Kindes und gegebenenfalls einige Fördermöglichkeiten.

Die Sitzung ist selbstverständlich gratis. Den Zeitpunkt und den Ort der Testung können wir flexibel nach Ihren Bedürfnissen wählen.

Aus den beiden Testergebnissen kann ich die Zuverlässigkeit des Tests berechnen, damit dieser im Sinne der Wissenschaft verbessert werden kann. Sie und Ihr Kind haben dadurch die Vorteile einer kostenlosen Leistungsrückmeldung. Ihre Daten werden geschützt, Sie bleiben anonym.

Damit wir das weitere Vorgehen besprechen können, bitte ich Sie, sich kurz bei mir zu melden:

telefonisch unter: 0676 390 1208

oder per e-mail an claudiaparfuss@gmx.at

oder über den Postweg: Claudia Parfuss, Buchengasse 5, 3001 Mauerbach.

Ich freue mich von Ihnen zu hören und sende freundliche Grüße,

Claudia Parfuss

5.5 Elternbrief des Kompetenzzentrums Wolkersdorf



BERATUNGSZENTRUM
für Schulfragen
Schlosspark 1
2120 Wolkersdorf
Tel 0 22 45 / 68 70
Fax 0 22 45 / 68 702

Wien, 18.4.08

Sehr geehrte Eltern!

Sie haben vor einiger Zeit eine psychologische Abklärung Ihres Kindes am Beratungszentrum für Schulfragen in Wolkersdorf durchführen lassen.

Im Rahmen der Untersuchung wurde der Intelligenztest AID 2 (Adaptives Intelligenzdiagnostikum 2) eingesetzt, den ein Wiener Forscherteam erstellt hatte.

Dieser Test soll nun auf wissenschaftlicher Grundlage an der Universität Wien / Psychologische Fakultät weiter entwickelt werden. Dazu ist eine nochmalige Vorgabe des Tests bei genau denselben Schülern und Schülerinnen, die den Test schon einmal gemacht haben, notwendig.

Diese nochmalige Untersuchung dient ausschließlich wissenschaftlichen Zwecken, hat keine schulische Bedeutung bzw. Verbindung und die Teilnahme ist selbstverständlich freiwillig. Die Untersuchung selbst wird von einer Diplomandin der Universität Wien an der Psychologischen Fakultät, Frau Claudia Parfuß, durchgeführt.

Wegen der Wahrung des strengen Datenschutzes gibt das Beratungszentrum die Testergebnisse der ersten Untersuchung im genannten Intelligenztest nur dann an die Universität weiter, wenn Sie sich bei der Diplomandin, Frau Parfuß persönlich melden und mit einer nochmaligen Testung einverstanden sind.

Alle anderen, von Ihnen bei der psychologischen Beratung gemachten Angaben oder weitere Testergebnisse werden nicht weitergegeben.

Wir ersuchen Sie um Unterstützung dieses wissenschaftlichen Vorhabens.

Mit freundlichen Grüßen

Dr. Renate Strasser
Leiterin des Beratungszentrums

5.6 Elternbrief der Schulpsychologie Wien



Ihr Zeichen, Unser Zeichen/GZ
Ihre Nachricht

BearbeiterIn
HR Dr. Mathilde Zeman
mathile.zeman@ssr-wien.gv.at

TEL 525 25
DW 77505
FAX 52525/99/77501

Datum
14.03.2008

Sehr geehrte Eltern!

Sie haben mit Ihrer Tochterschulpsychologische
Beratung des Stadtschulrates für Wien in Anspruch genommen.

Im Rahmen der schulpsychologischen Untersuchung wurde auch ein Intelligenztest (A-I-D 2)
eingesetzt.

Dieser Test soll nun auf wissenschaftlicher Grundlage an der Universität Wien /
Psychologische Fakultät weiter entwickelt werden. Dazu ist eine nochmalige Vorgabe des
Tests bei denselben Schülerinnen und Schülern, die den Test schon einmal gemacht haben,
notwendig.

Diese nochmalige Untersuchung dient ausschließlich wissenschaftlichen Zwecken, hat keine
schulische Bedeutung bzw. Verbindung und die Teilnahme ist selbstverständlich freiwillig.

Die Untersuchung selbst wird von einer Diplomandin der Universität Wien an der
Psychologischen Fakultät, Frau Claudia Parfuss durchgeführt.

Wegen der Wahrung des strengen Datenschutzes gibt die Abteilung Schulpsychologie Bildungsberatung die Testergebnisse der ersten Untersuchung im genannten Intelligenztest nur dann an die Universität weiter, wenn Sie sich bei der Diplomandin Frau Parfuss persönlich melden und mit einer nochmaligen Testung Ihrer Tochter einverstanden sind. Alle anderen von Ihnen bei der schulpsychologischen Beratung gemachten Angaben bzw. andere erhobene Testergebnisse werden nicht weitergegeben.

Wir ersuchen Sie um Unterstützung dieses wissenschaftlichen Vorhabens.

Mit freundlichen Grüßen

HR Dr. Mathilde Zeman
Leiterin der Abteilung
Schulpsychologie Bildungsberatung
Stadtschulrat für Wien

Beilage

5.7 Beispielrückmeldung⁴

Rückmeldung für Anatol R, geb. am

Anatol R wurde am 08. Mai 2006 im Rahmen einer Diplomarbeitsstudie der Fakultät für Psychologie mit dem in der Schulleistungsdiagnostik gängigen AID2 getestet.

Im Folgenden finden Sie eine kurze Beschreibung von Anatols Testleistungen. Es wird darauf hingewiesen, dass die durchgeführte Untersuchung wissenschaftlichen Zwecken dient und keine vollständige psychologisch-diagnostische Untersuchung darstellt und eine solche somit auch nicht ersetzen kann. Sollten Fragen zur Studie oder zur durchgeführten Testung auftreten, stehe ich Ihnen jederzeit gerne unter folgender Kontaktadresse (claudiaparfuss@gmx.at) zur Verfügung. Sofern eine ausführliche Einzelfalldiagnostik für Anatol gewünscht wird, wenden Sie sich zur weiteren Vermittlung bitte je nach Ihrem Wohnsitz an die Schulpsychologie Wien (+43 (0)1 52525-77505; <http://www.magwien.gv.at/bildung/stadtschulrat/schulpsychologie/index.html>) oder Niederösterreich (+43 (0)2742 280-4700; schulpsychologie.lsr-noe.gv.at).

Anatol hat während der gesamten Testung fleißig und gut konzentriert mitgearbeitet und ist vor allem durch seine Leistungsfreude positiv aufgefallen. Merkwürdig waren deutliche Sprachschwierigkeiten, welche wahrscheinlich durch die russische Muttersprache bedingt sind.

Im Vergleich zu Gleichaltrigen zeigt er insgesamt über alle Aufgaben hinweg deutlich unterdurchschnittliche Leistungen, welche sehr stark streuen.

Probleme hatte Anatol bei Aufgaben, bei denen man Synonyme für verschiedene Worte finden muss. Das weist auf einen kleinen passiven Wortschatz hin. Er tut sich auch schwer Sachzusammenhänge der gesellschaftlichen Umwelt zu begreifen und weiß über sozial angepasste Verhaltensweisen und gesellschaftliche Bedingungen weniger gut bescheid. Fragen über das Alltagswissen konnte er kaum beantworten.

Gute Ergebnisse erzielte er bei den Puzzleaufgaben, bei denen mehrere Teile so zusammengelegt werden müssen, dass eine Figur entsteht. Ihm fiel es auch leicht Bildgeschichten in die richtige Reihenfolge zu bringen. Er weiß daher über die Abfolge sozialen Geschehens gut bescheid und schätzt soziales Geschehen richtig ein. Bei Aufgaben, bei welchen er sich vorgesagte sinnfreie Silbe merken musste, schnitt er ebenfalls gut ab.

Anatol Rs schwächere Bereiche könnte man spielerisch fördern, indem man mit ihm vermehrt gemeinsam Wissensbücher liest und Kindermuseen besucht. Man sollte ihn auch zum Lesen ermuntern oder mit ihm Hörkassetten anhören. Wichtig ist, dass man ihm im Alltag viel erklärt, damit er besser über die Gründe von Regelungen und Verhaltensweisen bescheid weiß.

Insgesamt kann man sagen, dass er fleißig und ausdauernd mitgearbeitet hat. Anhand der Ergebnisse und der Verhaltensbeobachtung lässt sich eine Schwäche im sprachlichen Bereich vermuten.

Vielen Dank für die Teilnahme an der Studie.

Ich wünsche Ihnen alles Gute für die Zukunft!

Mit freundlichen Grüßen,
Claudia Parfuss

⁴ Namen und Inhalte wurden hier aus Datenschutzgründen verändert.

5.8 Deskriptive Statistik der Testwerte nach den unterschiedlichen Intervallen

Tabelle 22: Vorgabe am selben Tag; Ersttestung; Testwerte deskriptiv

Untertest		1		5v	5r
		FP	T-Wert	T-Wert	FP
Stichprobenumfang		147	151	151	151
Mittelwert		1.38	52.73	54.87	51.73
Standardabweichung		2.04	9.34	9.99	8.01
Minimum		-3.39	26.00	23.00	31.00
Maximum		7.00	75.00	81.00	73.00
Perzentile	25	-0.10	46.00	47.00	47.00
	50	1.40	54.00	54.00	52.00
	75	2.70	60.00	62.00	57.00
Untertest		6		9	
		FP	T-Wert	FP	T-Wert
Stichprobenumfang		147	151	147	151
Mittelwert		0.11	54.09	2.78	55.51
Standardabweichung		1.90	10.82	1.51	10.56
Minimum		-5.50	25.00	-1.60	22.00
Maximum		6.20	81.00	6.40	80.00
Perzentile	25	-0.80	47.00	2.00	49.00
	50	0.20	55.00	2.70	55.00
	75	1.30	61.00	3.70	63.00

Anmerkung: FP...Fähigkeitsparameter; 5v...Untertest 5vorwärts;
5r... Untertest 5 rückwärts

Tabelle 23: Vorgabe am selben Tag; Zweittestung; Testwerte deskriptiv

Untertest		1		5v	5r
		FP	T-Wert	T-Wert	FP
Stichprobenumfang		151	151	151	151
Mittelwert		1.61	54.17	55.77	51.57
Standardabweichung		2.11	9.96	10.81	8.77
Minimum		-4.09	32.00	19.00	19.00
Maximum		6.80	81.00	81.00	81.00
Perzentile	25	0.59	46.00	49.00	46.00
	50	1.60	54.00	54.00	52.00
	75	3.20	62.00	62.00	58.00
Untertest		6		9	
		FP	T-Wert	FP	T-Wert
Stichprobenumfang		147	151	147	151
Mittelwert		-0.14	52.47	2.72	54.89
Standardabweichung		2.08	11.59	1.50	10.50
Minimum		-6.79	19.00	-2.00	27.00
Maximum		5.09	79.00	6.20	81.00
Perzentile	25	-1.40	45.00	1.80	48.00
	50	-0.10	53.00	2.60	55.00
	75	1.19	60.00	3.70	64.00

Anmerkung: FP...Fähigkeitsparameter; 5v...Untertest 5vorwärts;
5r... Untertest 5 rückwärts

Tabelle 24: Intervall ca. 1 Monat;Ersttestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		47	47	47	47	47	47	47
Mittelwert		0.60	52.02	47.21	1.05	50.68	2.10	52.04
Standardabweichung		2.81	8.34	7.63	2.89	11.65	3.34	10.97
Minimum		-5.00	32.00	27.00	-5.09	20.00	-3.89	31.00
Maximum		5.50	68.00	61.00	5.70	77.00	7.79	72.00
Perzentile	25	-1.50	47.00	42.00	-.70	44.00	-1.39	44.00
	50	1.00	52.00	49.00	1.69	51.00	3.00	52.00
	75	2.50	58.00	52.00	3.10	56.00	4.29	62.00
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		47	47	47	47	47	47	47
Mittelwert		52.62	53.91	-1.00	49.49	48.68	49.32	-0.20
Standardabweichung		8.58	9.76	2.43	9.68	7.88	8.83	2.12
Minimum		23.00	31.00	-6.79	30.00	29.00	19.00	-4.70
Maximum		70.00	81.00	5.00	78.00	65.00	64.00	5.50
Perzentile	25	49.00	47.00	-2.89	44.00	45.00	46.00	-1.40
	50	54.00	53.00	-.80	51.00	49.00	51.00	-0.10
	75	58.00	59.00	.50	56.00	54.00	55.00	1.00
Untertests		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		47	47	47	47	47	47	47
Mittelwert		49.40	1.83	52.02	1.05	52.55	1.24	52.91
Standardabweichung		10.01	1.58	7.31	3.33	10.73	2.13	9.82
Minimum		24.00	-1.39	37.00	-4.60	25.00	-3.60	29.00
Maximum		77.00	5.40	72.00	5.90	81.00	6.59	78.00
Perzentile	25	43.00	0.80	48.00	-1.9	45.00	-0.50	49.00
	50	49.00	2.00	52.00	0.69	53.00	1.89	53.00
	75	54.00	3.10	57.00	4.09	59.00	2.89	61.00
Zusatztests		5a	5b	10a				
		T-Wert	T-Wert	FP	FP			
Stichprobenumfang		47	47	34	34			
Mittelwert		54.94	54.94	1.30	56.53			
Standardabweichung		9.65	9.65	1.52	7.72			
Minimum		40.00	40.00	-2.50	33.00			
Maximum		81.00	81.00	4.40	73.00			
Perzentile	25	47.00	45.00	0.05	52.25			
	50	53.00	51.00	1.20	57.00			
	75	64.00	57.00	2.70	61.50			
Testkennwerte		Max.	Min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		41	41	41	41	41	41	41
Mittelwert		65.78	38.32	63.41	27.46	44.09	42.10	61.78
Standardabweichung		7.39	6.94	25.08	6.85	26.26	6.13	27.46
Minimum		53.00	20.00	11.50	15.00	3.30	24.00	3.30
Maximum		81.00	51.00	98.90	44.00	94.50	52.00	96.00
Perzentile	25	60.00	34.00	42.10	21.00	15.00	39.00	46.00
	50	65.00	38.00	61.80	28.00	46.00	42.00	61.80
	75	72.00	43.00	85.05	31.00	61.80	47.00	88.50

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min. ... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 25: Intervall ca. 1 Monat; Zweittestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		47	47	47	47	47	47	47
Mittelwert		0.92	53.77	51.28	1.72	54.55	3.08	57.87
Standardabweichung		2.69	8.41	7.44	3.02	12.49	2.72	10.94
Minimum		-4.50	33.00	34.00	-4.50	21.00	-2.70	36.00
Maximum		5.90	72.00	66.00	6.79	77.00	7.80	78.00
Perzentile	25	-1.10	47.00	47.00	-1.00	49.00	0.69	52.00
	50	1.19	54.00	51.00	2.70	54.00	3.50	55.00
	75	3.10	61.00	55.00	4.00	64.00	5.00	67.00
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		47	47	47	47	47	47	46
Mittelwert		53.47	55.13	-0.45	52.60	56.30	54.43	1.43
Standardabweichung		11.73	9.56	2.32	9.18	10.16	8.39	3.07
Minimum		19.00	34.00	-6.00	35.00	37.00	34.00	-3.00
Maximum		81.00	77.00	4.59	75.00	77.00	70.00	14.00
Perzentile	25	50.00	48.00	-1.69	46.00	51.00	47.00	-0.84
	50	54.00	54.00	-0.20	52.00	55.00	55.00	1.00
	75	61.00	62.00	1.10	58.00	63.00	60.00	2.89
Untertests		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		47	46	47	46	47	46	47
Mittelwert		56.96	1.96	52.98	2.48	59.19	1.47	54.43
Standardabweichung		10.51	1.64	7.24	2.88	10.36	2.05	8.69
Minimum		36.00	-1.39	41.00	-3.20	40.00	-3.00	36.00
Maximum		81.00	5.40	72.00	5.90	81.00	4.79	70.00
Perzentile	25	50.00	0.80	47.00	0.20	50.00	-0.20	47.00
	50	54.00	2.39	52.00	3.00	59.00	1.65	54.00
	75	64.00	3.10	57.00	4.54	66.00	3.39	61.00
Zusatztests		5a	5b	10a				
		T-Wert	T-Wert	FP	FP			
Stichprobenumfang		47	47	36	36			
Mittelwert		56.43	54.02	1.59	58.33			
Standardabweichung		8.26	6.25	1.65	8.95			
Minimum		41.00	34.00	-2.50	38.00			
Maximum		81.00	66.00	6.20	81.00			
Perzentile	25	51.00	50.00	.73	53.00			
	50	57.00	56.00	1.39	58.00			
	75	62.00	57.00	2.72	64.00			
Testkennwerte		Max.	Min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		40	40	40	40	40	40	40
Mittelwert		69.75	42.65	78.20	27.10	41.29	46.80	79.65
Standardabweichung		7.77	6.55	22.20	6.72	23.25	5.57	21.03
Minimum		55.00	21.00	12.90	14.00	4.40	34.00	24.20
Maximum		81.00	53.00	99.30	47.00	96.40	57.00	99.10
Perzentile	25	64.00	38.00	61.80	23.00	24.20	42.25	63.65
	50	70.00	44.00	88.50	25.00	30.90	47.00	88.50
	75	75.75	47.75	95.33	29.75	55.93	51.00	95.20

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min. ... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 26: Intervall 1-3 Monate; Ersttestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		74	74	74	74	74	74	74
Mittelwert		-0.23	50.66	49.03	0.33	50.27	1.49	53.50
Standardabweichung		2.62	8.58	7.34	2.79	12.43	2.92	10.99
Minimum		-7.00	19.00	27.00	-6.90	19.00	-3.89	31.00
Maximum		5.50	68.00	64.00	5.70	81.00	7.79	81.00
Perzentile	25	-2.00	44.00	45.50	-1.67	43.00	-0.50	44.00
	50	-0.85	52.00	49.00	.80	49.50	1.00	52.00
	75	1.67	56.00	54.00	2.60	56.00	3.92	62.00
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		74	74	74	74	74	74	74
Mittelwert		53.84	54.22	-1.55	49.03	50.85	50.27	-0.38
Standardabweichung		8.01	9.26	2.25	9.92	8.71	8.09	1.90
Minimum		23.00	23.00	-6.79	19.00	29.00	19.00	-4.70
Maximum		77.00	81.00	5.00	78.00	70.00	64.00	5.50
Perzentile	25	50.00	48.75	-3.13	43.50	45.75	46.00	-1.40
	50	54.00	55.00	-1.50	51.00	50.00	53.00	-0.10
	75	57.00	58.25	-0.22	56.00	57.00	55.00	0.45
Untertest		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		74	74	74	74	74	74	74
Mittelwert		49.55	1.43	51.93	0.19	52.49	0.60	51.38
Standardabweichung		9.32	1.66	8.58	3.10	10.10	2.05	10.65
Minimum		24.00	-5.40	19.00	-4.60	25.00	-3.60	28.00
Maximum		77.00	5.40	72.00	5.90	81.00	6.59	78.00
Perzentile	25	44.75	0.67	48.00	-2.29	46.00	-.80	44.75
	50	50.00	1.30	52.00	-.30	53.00	0.50	53.00
	75	54.00	2.40	56.25	2.50	59.00	2.22	59.00
Zusatztest		5a	5b	10a				
		T-Wert	T-Wert	FP	T-Wert			
Stichprobenumfang		74	74	57	57			
Mittelwert		54.76	51.58	0.47	55.30			
Standardabweichung		9.67	8.39	1.81	9.29			
Minimum		36.00	36.00	-3.30	33.00			
Maximum		81.00	81.00	4.40	81.00			
Perzentile	25	47.00	45.00	-.85	50.00			
	50	53.50	51.00	0.29	55.00			
	75	61.00	56.25	1.40	61.00			
Testkennwert		Max.	Min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		66	65	66	66	65	66	66
Mittelwert		65.68	38.31	63.57	27.95	48.67	42.20	64.07
Standardabweichung		7.53	7.72	27.83	9.22	28.31	6.97	28.20
Minimum		52.00	19.00	1.00	13.00	3.30	19.00	3.30
Maximum		81.00	51.00	98.90	64.00	99.00	54.00	97.40
Perzentile	25	59.75	33.50	38.20	21.00	23.10	38.75	44.05
	50	65.00	40.00	70.85	27.50	46.00	44.00	75.80
	75	71.25	44.00	88.50	33.25	74.15	48.00	90.50

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min. ... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 27: Intervall 1-3 Monate; Zweittestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		74	74	74	74	74	74	74
Mittelwert		0.23	52.73	52.47	1.05	53.93	2.22	57.22
Standardabweichung		2.61	9.51	7.36	2.85	13.18	2.61	11.24
Minimum		-7.09	19.00	34.00	-5.29	19.00	-2.79	28.00
Maximum		5.90	73.00	52.47	6.79	77.00	7.80	78.00
Perzentile	25	-1.50	47.00	48.00	-1.44	45.50	0.20	49.00
	50	0.00	53.50	53.00	1.50	55.00	1.60	56.00
	75	2.29	59.25	58.00	3.39	64.25	4.40	66.25
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		74	74	74	74	74	74	73
Mittelwert		54.61	54.64	-1.12	51.31	53.57	55.69	0.92
Standardabweichung		11.05	9.31	2.22	9.83	8.17	9.88	2.64
Minimum		19.00	34.00	-6.00	27.00	34.00	34.00	-3.10
Maximum		81.00	81.00	4.59	75.00	70.00	79.00	14.00
Perzentile	25	50.00	49.00	-2.89	44.00	47.75	50.75	-0.70
	50	54.00	53.00	-1.19	51.00	55.00	55.00	0.40
	75	61.00	59.00	.50	57.00	60.00	61.00	1.90
Untertest		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		74	73	74	73	74	73	74
Mittelwert		55.64	1.69	53.31	1.50	58.39	0.8	53.11
Standardabweichung		9.55	1.54	7.96	2.86	10.02	1.97	9.95
Minimum		35.00	-2.39	31.00	-4.00	37.00	-4.00	19.00
Maximum		81.00	5.40	72.00	5.90	81.00	4.79	72.00
Perzentile	25	50.00	085	48.00	-0.10	51.50	-.40	46.75
	50	54.00	1.80	52.50	1.39	59.00	.60	53.00
	75	63.00	2.70	57.00	4.09	65.25	2.29	60.25
Zusatztest		5a	5b	10a				
		T-Wert	T-Wert	FP	T-Wert			
Stichprobenumfang		74	74	61	61			
Mittelwert		55.15	52.93	0.80	57.00			
Standardabweichung		8.14	7.79	1.91	9.98			
Minimum		40.00	34.00	-3.39	35.00			
Maximum		81.00	77.00	6.20	81.00			
Perzentile	25	48.00	45.00	0.05	52.00			
	50	55.00	55.00	.90	59.00			
	75	61.00	57.00	1.60	64.00			
Testkennwert		Max.	min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		57	56	56	56	56	57	57
Mittelwert		68.81	41.73	74.57	27.05	42.78	46.05	77.52
Standardabweichung		7.42	7.67	25.37	6.88	24.13	6.50	23.12
Minimum		54.00	19.00	11.00	12.00	4.40	19.00	5.00
Maximum		81.00	53.00	99.30	47.00	96.40	57.00	99.10
Perzentile	25	64.00	37.25	58.88	23.00	24.20	42.00	61.80
	50	69.00	43.50	86.80	25.00	34.55	47.00	88.50
	75	74.50	48.00	95.60	30.75	61.80	51.00	95.20

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min. ... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 28: Intervall 0.5 bis 1.92 Jahre; Ersttestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		59	70	70	59	70	58	68
Mittelwert		-0.74	54.44	50.47	0.17	56.84	0.74	53.96
Standardabweichung		2.50	11.15	8.25	2.96	11.59	3.17	11.67
Minimum		-7.60	26.00	27.00	-6.40	24.00	-6.70	19.00
Maximum		4.59	96.00	73.00	8.40	81.00	7.80	81.00
Perzentile	25	-2.50	49.50	45.00	-1.90	48.75	-1.70	46.50
	50	-0.90	54.00	50.50	0.10	57.00	-0.40	55.00
	75	0.80	61.00	54.00	2.40	64.00	3.00	60.00
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		68	68	58	68	69	69	59
Mittelwert		57.13	56.79	-1.80	51.44	56.04	54.23	-0.55
Standardabweichung		10.25	9.16	2.20	9.17	9.11	8.78	2.32
Minimum		23.00	38.00	-6.00	26.00	40.00	34.00	-6.50
Maximum		81.00	81.00	3.60	72.00	81.00	70.00	4.40
Perzentile	25	51.25	50.00	-3.60	45.50	49.00	48.00	-2.30
	50	54.00	57.00	-1.90	51.00	57.00	55.00	-0.70
	75	61.75	61.75	-0.30	57.75	62.50	60.00	1.00
Untertest		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		69	59	69	59	69	58	68
Mittelwert		51.36	0.78	52.35	-0.19	54.87	0.21	53.68
Standardabweichung		10.39	1.69	10.25	3.32	11.17	1.75	9.85
Minimum		19.00	-3.00	31.00	-6.00	19.00	-3.30	31.00
Maximum		74.00	6.40	75.00	5.90	81.00	5.70	81.00
Perzentile	25	45.00	-0.20	46.00	-2.70	49.00	-1.00	47.25
	50	54.00	0.70	51.00	-1.30	53.00	0.25	53.50
	75	60.00	2.00	61.00	2.60	62.00	1.24	59.00
Testkennwert		Max.	min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		60	60	60	60	60	60	60
Mittelwert		68.68	39.25	66.95	29.43	48.93	43.82	68.73
Standardabweichung		7.88	7.87	27.12	8.06	28.39	6.62	26.32
Minimum		52.00	19.00	1.10	16.00	4.40	23.00	2.20
Maximum		81.00	52.00	99.10	48.00	97.00	59.00	99.50
Perzentile	25	63.00	34.25	43.58	22.25	19.85	40.25	51.00
	50	67.00	41.00	75.80	29.00	50.00	44.00	75.80
	75	75.50	44.00	88.50	35.00	72.50	48.00	90.50

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min. ... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 29: Intervall 0.5 bis 1.92 Jahre; Zweittestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		68	70	70	68	70	68	69
Mittelwert		0.91	54.81	53.09	2.19	58.71	2.88	58.04
Standardabweichung		2.04	9.60	7.94	2.42	12.93	2.34	12.43
Minimum		-3.39	37.00	36.00	-5.09	19.00	-3.89	25.00
Maximum		6.90	78.00	77.00	7.50	81.00	7.80	81.00
Perzentile	25	-0.35	47.75	48.00	1.35	52.75	0.70	51.00
	50	0.80	55.00	52.50	2.60	59.50	3.30	60.00
	75	2.18	61.25	58.00	3.50	65.25	4.40	66.00
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		69	69	68	69	70	70	68
Mittelwert		57.12	54.45	-0.43	54.93	54.76	55.69	0.81
Standardabweichung		8.91	9.02	1.75	10.67	8.88	9.96	2.34
Minimum		40.00	38.00	-5.50	25.00	27.00	36.00	-4.70
Maximum		81.00	81.00	5.40	81.00	77.00	81.00	5.50
Perzentile	25	51.50	48.50	-1.18	49.00	48.00	48.00	-0.70
	50	56.00	53.00	-.50	55.00	55.00	55.00	1.00
	75	62.00	58.00	.58	62.50	60.00	63.00	2.20
Untertest		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		69	68	69	68	69	68	69
Mittelwert		54.43	2.43	56.70	1.67	57.22	1.41	55.62
Standardabweichung		11.00	1.57	10.61	3.18	11.33	1.60	11.10
Minimum		28.00	-1.60	31.00	-4.59	28.00	-2.60	27.00
Maximum		77.00	5.70	79.00	5.90	81.00	5.00	80.00
Perzentile	25	46.00	1.35	50.00	-1.98	48.00	0.40	48.50
	50	54.00	2.40	56.00	2.20	57.00	1.40	55.00
	75	61.00	3.10	64.50	4.09	65.50	2.00	62.00
Testkennwert		Max.	min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		67	67	67	67	67	67	67
Mittelwert		71.33	41.39	73.69	29.97	51.34	45.94	76.38
Standardabweichung		7.25	7.61	24.93	7.38	25.27	6.63	25.11
Minimum		50.00	19.00	1.10	18.00	7.50	27.00	7.50
Maximum		81.00	58.00	99.90	51.00	98.90	59.00	99.50
Perzentile	25	66.00	37.00	57.90	25.00	30.90	42.00	61.80
	50	72.00	42.00	81.60	29.00	50.00	47.00	88.50
	75	77.00	47.00	94.50	34.00	69.20	50.00	94.50

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min.... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 30: Intervall 2 bis 2.67 Jahre; Ersttestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		66	72	72	66	72	66	72
Mittelwert		-0.16	54.82	51.14	0.65	57.03	1.16	55.31
Standardabweichung		2.78	10.29	8.45	2.99	13.24	2.77	10.97
Minimum		-5.70	24.00	23.00	-6.40	23.00	-3.90	37.00
Maximum		5.30	81.00	73.00	6.80	81.00	6.10	81.00
Perzentile	25	-2.00	47.00	47.00	-1.05	49.00	-0.70	45.00
	50	-0.15	54.00	50.00	0.80	59.00	0.70	54.50
	75	1.90	61.75	56.50	2.70	65.00	3.40	65.00
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		72	72	66	72	72	72	66
Mittelwert		57.14	56.51	-1.53	52.40	55.61	53.93	-0.56
Standardabweichung		9.11	11.15	2.63	11.61	9.79	8.95	2.00
Minimum		43.00	19.00	-7.80	19.00	38.00	27.00	-5.20
Maximum		81.00	81.00	3.80	75.00	81.00	73.00	4.40
Perzentile	25	51.00	49.25	-3.13	44.00	47.00	48.00	-2.30
	50	54.00	55.50	-1.30	53.50	55.50	55.00	-0.10
	75	61.75	62.00	0.10	62.00	62.00	60.00	0.70
Untertest		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		72	66	72	66	72	66	72
Mittelwert		50.33	0.95	51.96	0.28	55.75	0.43	53.79
Standardabweichung		8.73	2.02	11.04	3.02	10.54	2.12	12.02
Minimum		30.00	-5.60	19.00	-4.60	36.00	-4.00	26.00
Maximum		72.00	6.40	75.00	5.90	82.00	6.60	80.00
Perzentile	25	46.00	-0.20	46.00	-2.30	47.25	-0.78	46.00
	50	50.00	1.25	52.00	0.20	55.00	0.40	54.50
	75	57.00	1.92	59.00	2.63	62.00	1.53	63.75
Testkennwert		Max.	min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		65	65	65	65	65	65	65
Mittelwert		69.40	39.94	69.02	29.46	50.10	68.57	43.69
Standardabweichung		7.47	7.47	25.61	7.42	27.35	27.77	7.31
Minimum		54.00	19.00	1.10	15.00	3.30	0.50	19.00
Maximum		81.00	55.00	99.70	48.00	97.00	99.30	58.00
Perzentile	25	65.00	37.00	57.90	23.00	24.20	54.00	41.00
	50	68.00	41.00	75.80	29.00	50.00	75.80	44.00
	75	76.00	46.00	92.50	35.50	74.15	94.50	50.00

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min.... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 31: Intervall 2 bis 2.67 Jahre; Zweittestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		71	72	72	70	71	69	72
Mittelwert		1.71	54.42	54.31	2.96	57.41	3.50	56.86
Standardabweichung		2.33	10.59	8.22	1.89	11.26	2.28	10.71
Minimum		-3.00	26.00	34.00	-1.20	32.00	-2.70	37.00
Maximum		7.00	75.00	66.00	7.50	81.00	7.79	78.00
Perzentile	25	-0.40	47.00	49.00	1.69	50.00	2.20	49.25
	50	1.90	54.00	54.00	3.30	58.00	3.50	56.00
	75	3.39	63.00	61.00	4.05	64.00	5.05	65.00
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		71	71	71	72	72	72	69
Mittelwert		55.79	54.17	0.48	56.43	54.96	54.17	1.12
Standardabweichung		10.83	9.69	2.02	10.75	10.40	9.24	1.92
Minimum		23.00	31.00	-4.60	32.00	30.00	27.00	-3.90
Maximum		81.00	81.00	6.20	81.00	81.00	70.00	5.50
Perzentile	25	49.00	49.00	-0.80	48.25	47.25	48.00	-0.69
	50	54.00	53.00	0.50	56.50	55.00	56.00	1.00
	75	62.00	58.00	1.60	62.75	61.00	60.00	2.60
Untertest		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		72	71	72	69	72	70	71
Mittelwert		54.56	2.73	55.40	2.47	57.26	2.18	56.87
Standardabweichung		9.37	1.72	10.59	2.72	10.46	1.87	11.079
Minimum		31.00	-1.90	30.00	-2.70	35.00	-2.60	27.00
Maximum		75.00	6.40	80.00	5.90	81.00	6.60	81.00
Perzentile	25	48.25	1.89	48.00	0.59	50.00	0.98	50.00
	50	54.00	2.70	55.00	2.20	58.50	2.00	56.00
	75	60.00	3.90	63.00	5.90	64.50	3.32	64.00
Testkennwert		Max.	min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		71	71	71	71	71	71	71
Mittelwert		69.56	41.86	73.79	27.66	43.92	46.08	75.56
Standardabweichung		8.05	7.32	22.35	7.64	28.00	7.29	25.61
Minimum		53.00	23.00	15.60	14.00	2.20	27.00	7.50
Maximum		81.00	59.00	99.90	45.00	95.10	64.00	99.90
Perzentile	25	62.00	38.00	61.80	22.00	18.40	42.00	61.80
	50	70.00	41.00	75.80	27.00	42.10	47.00	88.50
	75	77.00	48.00	95.60	32.00	65.50	51.00	95.20

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min.... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 32: Intervall 2.75 bis 6.42 Jahre; Ersttestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		58	69	69	58	69	57	68
Mittelwert		-0.70	54.74	52.72	0.08	57.91	1.09	57.66
Standardabweichung		2.00	10.30	9.62	2.67	14.15	2.61	12.81
Minimum		-4.20	35.00	22.00	-6.90	19.00	-3.90	28.00
Maximum		4.20	74.00	73.00	5.50	81.00	7.80	81.00
Perzentile	25	-2.05	46.50	47.00	-1.45	48.50	-0.50	48.75
	50	-0.80	55.00	53.00	2.20	59.00	0.70	58.00
	75	0.32	62.50	60.00	1.63	68.00	2.75	66.75
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		68	68	57	68	68	68	58
Mittelwert		56.62	55.12	-1.78	53.32	54.72	54.32	-0.78
Standardabweichung		7.68	8.70	2.10	9.92	8.41	9.06	1.82
Minimum		41.00	38.00	-6.00	33.00	34.00	33.00	-5.60
Maximum		77.00	94.00	3.30	75.00	82.00	76.00	3.60
Perzentile	25	51.00	49.00	-3.25	47.25	50.00	49.00	-1.40
	50	54.00	55.00	-1.60	53.50	55.50	54.50	-0.70
	75	61.75	58.75	-0.30	60.00	60.00	61.00	0.03
Untertest		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		69	58	69	58	69	57	68
Mittelwert		49.88	1.22	55.17	-0.77	53.86	-0.02	52.38
Standardabweichung		8.73	1.72	11.01	2.57	9.73	1.63	9.89
Minimum		25.00	-5.40	19.00	-4.60	28.00	-3.30	30.00
Maximum		69.00	5.40	81.00	5.90	74.00	3.40	74.00
Perzentile	25	46.00	0.33	47.00	-2.30	47.50	-1.25	45.00
	50	50.00	1.35	55.00	-1.30	54.00	0.20	52.00
	75	56.00	2.20	62.50	0.83	61.00	0.85	60.75
Testkennwert		Max.	min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		57	57	57	57	57	57	57
Mittelwert		68.11	40.14	68.77	27.96	44.39	44.40	70.91
Standardabweichung		7.29	7.70	25.91	7.34	27.36	7.27	26.21
Minimum		50.00	19.00	1.10	14.00	2.20	19.00	0.50
Maximum		81.00	55.00	99.70	47.00	96.40	57.00	99.10
Perzentile	25	62.50	36.00	50.00	22.00	18.40	41.00	54.00
	50	68.00	40.00	72.50	28.00	46.00	44.00	75.80
	75	74.50	45.50	91.50	32.00	65.50	50.00	94.50

Anmerkung: FP... Fähigkeitsparameter; PR... Prozentrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min.... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

Tabelle 33: Intervall 2.75 bis 6.42 Jahre; Zweittestung; Testwerte deskriptiv

Untertest		1		2	3		4	
		FP	T-Wert	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		64	69	69	64	69	63	69
Mittelwert		2.50	55.39	51.84	3.41	56.03	4.58	56.88
Standardabweichung		1.69	9.72	9.29	1.52	10.70	1.62	11.08
Minimum		-1.69	33.00	19.00	-1.50	23.00	0.70	19.00
Maximum		5.79	78.00	73.00	6.79	81.00	7.80	78.00
Perzentile	25	1.19	50.00	47.00	2.60	49.50	3.50	51.00
	50	2.39	56.00	51.00	3.50	55.00	5.00	56.00
	75	3.70	62.50	57.00	4.43	64.00	6.09	62.50
Untertest		5v	5r	6		7A	7K	8
		T-Wert	T-Wert	FP	T-Wert	T-Wert	T-Wert	FP
Stichprobenumfang		69	69	64	69	69	69	63
Mittelwert		56.59	53.42	0.87	55.74	53.04	53.43	1.40
Standardabweichung		9.29	10.51	1.42	9.75	10.11	10.43	1.74
Minimum		40.00	34.00	-3.29	31.00	31.00	19.00	-3.89
Maximum		81.00	81.00	5.50	81.00	77.00	70.00	5.50
Perzentile	25	51.00	46.00	-0.05	52.00	46.00	47.00	0.40
	50	55.00	53.00	0.60	56.00	54.00	53.00	1.39
	75	64.00	60.00	1.84	61.50	59.00	60.00	2.29
Untertest		8	9		10		11	
		T-Wert	FP	T-Wert	FP	T-Wert	FP	T-Wert
Stichprobenumfang		69	64	69	63	69	64	69
Mittelwert		54.10	3.39	56.52	3.45	56.59	2.84	57.86
Standardabweichung		9.66	1.21	9.88	2.25	11.08	1.47	10.97
Minimum		27.00	0.29	22.00	-1.10	31.00	-0.20	30.00
Maximum		78.00	6.40	74.00	5.90	81.00	5.70	81.00
Perzentile	25	49.50	2.70	51.00	1.50	50.00	1.93	51.00
	50	54.00	3.10	56.00	4.09	57.00	2.65	58.00
	75	60.00	4.50	64.50	5.90	62.00	3.79	64.00
Testkennwert		Max.	min./unt. Grenze		Range		2.niedr. Leistung	
		T-Wert	T-Wert	PR	T-Wert	PR	T-Wert	PR
Stichprobenumfang		68	68	68	68	68	68	68
Mittelwert		69.87	41.85	74.37	28.01	45.81	45.75	75.88
Standardabweichung		7.15	7.48	25.06	7.16	26.40	7.24	24.63
Minimum		54.00	19.00	1.10	11.00	0.80	19.00	2.20
Maximum		81.00	58.00	99.90	46.00	95.80	59.00	99.50
Perzentile	25	65.00	36.25	51.98	24.00	27.60	42.00	61.80
	50	70.00	42.50	83.35	28.50	48.00	47.50	89.50
	75	74.75	46.75	94.00	33.00	67.40	51.00	95.20

Anmerkung: FP... Fähigkeitsparameter; PR... Prozenrang; 5v...Untertest 5 vorwärts; 5r... Untertest 5 rückwärts; 7K... 7 Kodieren; 7A... 7 Assoziationen; Max. ... Maximaler T-Wert; Min.... Minimaler T-Wert; unt. Grenze... untere Grenze der Intelligenzquantität; 2.niedr. Leistung... 2.niedrigste Untertestleistung

5.9 Lebenslauf

CURRICULUM VITAE

Claudia Parfuss
Buchengasse 5
3001 Mauerbach
Tel.: + 43 676 390 1208
E-mail: claudiaparfuss@gmx.at

Persönliche Daten

Geburtsdatum:	26. 02. 1985
Geburtsort:	Wien
Staatsbürgerschaft:	Österreich
Familienstand:	ledig

Ausbildung

- 09/1995 – 06/2003 **Gymnasium der Dominikanerinnen, Wien**
Matura mit gutem Erfolg, neusprachlicher Zweig
- 10/2003 – dato **Universität Wien**
Studium: Psychologie, Pflegewissenschaften, Pädagogik, Französisch, Ungarisch
- 16.2.2006 **1. Diplomprüfung Psychologie**
Schwerpunkte im 2. Abschnitt: Angewandte Kinder- und Jugendpsychologie; Psychologische Diagnostik
Diplomarbeit in Psychologischer Diagnostik zum Thema: „Die Stabilität des Adaptiven Intelligenz Diagnostikums 2“ (AID 2)
Betreuung durch Univ.-Prof. Dr. Mag. Klaus Kubinger

Berufserfahrung

- 09/1999 – 06/2002 **Tutorenprogramm „Schüler helfen Schülern“**
Tätigkeit: Nachhilfe für Schüler der Unterstufe in Deutsch, Englisch, Mathematik
- 08/2003 **Verein zur Förderung freiwilliger sozialer Dienste**
Sommereinsatz: Verein Ferienheim Maien, Vorarlberg
Tätigkeit: Betreuung und Erziehung von Kindern
- 02/2004 **Freie Dienstnehmerin bei "Sale for you", Studentenjob**
Tätigkeiten: Kundenakquirierung
- 09/2004 - dato **Wiener Sozialdienste Förderung und Begleitung GmbH, Behindertenbetreuung/Persönliche Assistenz**
Tätigkeit: Betreuung von psychisch kranken und mehrfach behinderten Menschen
- 08/2007 – 11/2007 **Psychologisches Praktikum bei Hill International, Wien**
Tätigkeiten: Berufs- und Karriereberatung; Normberechnung; Übersetzung von Tests (ungarisch); Manualgestaltung; Itemgenerierung; Vorbereitung und Mitarbeit an Trainings
- 03/2008 – 05/2008 **Freie Dienstnehmerin an der Universität Wien**
Tätigkeit: Durchführung von AID 2 Testungen in englischer Sprache

Weitere Qualifikationen

- Sprachen: **Deutsch**, verhandlungsfähig
Ungarisch, verhandlungsfähig
Englisch, verhandlungsfähig
Französisch, fließend in Wort und Schrift
Latein, Grundkenntnisse
- EDV: **MS Office, SPSS**

Interessen

- Sport: Schifahren, Schwimmen, Tauchen, Laufen, Aerobic
Sonstiges: Reisen, Kunst, Gesellschaftsspiele, Tanzen, Freunde, Familie