

Titel der Diplomarbeit

**Einzel- versus Gruppenvorgabe der Untertests *Antonyme*,  
*Synonyme Finden* und *Funktionen Abstrahieren* des AID 3**

Verfasserin

Jana Goldmann

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, April 2011

Studienkennzahl: A 298

Studienrichtung: Psychologie

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



## Danksagung

An dieser Stelle möchte ich mich bei meinem Betreuer, **Univ. Prof. Dr. Mag. Kubinger**, bedanken, der mich zu diesem interessanten Thema inspiriert hat und mir im Laufe der Diplomarbeit mit seinem Expertenwissen stets zur Verfügung stand. Des Weiteren gebührt Frau **Univ. Prof. Dr. Holocher- Ertl** für ihre fachliche Unterstützung großer Dank.

Auch meinen Eltern gebührt großer Dank für die Unterstützung in allen Bereichen, die sie mir im Laufe des gesamten Studiums geboten haben.

Meiner Arbeitgeberin Frau Univ. Doz. Dr. Margot Schmitz möchte ich an dieser Stelle ebenfalls Dank aussprechen, da sie mir durch flexible Dienstenteilungen und viel Verständnis mein Studium erleichtert hat.

Großer Dank gebührt außerdem meinem Partner Mag. Dr. Dennstedt für seine moralische Unterstützung.

Auch den Direktoren, Lehrern und Schüler der Europaschule Linz Praxisvolksschule der Pädagogischen Hochschule Oberösterreich Lederergasse 35 4020 Linz und des Wirtschaftskundlichen Realgymnasiums der Franziskanerinnen Vogelweiderstraße 2-4, 4600 Wels möchte ich danken, dass sie mir die Möglichkeit gegeben haben die Testungen an ihren Schulen durchzuführen.



## **Abstract**

Es ist bekannt, dass Gruppentests auf Grund der erhöhten Ökonomie einen großen Vorteil gegenüber Individualtestungen mit sich bringen. Allerdings gibt es in der Literatur nur wenige Studien, die sich damit befassen, ob ein Gruppentest auch als Individualtest eingesetzt werden kann und umgekehrt, ob der Einsatz eines Individualtests auch als Gruppentest möglich ist. In der vorliegenden Studie wurden 147 Schüler der dritten, fünften und achten Schulstufe randomisiert den Bedingungen Einzel- und Gruppentest zugewiesen. Es zeigte sich in zwei der drei untersuchten Subtests, dass die Vorgabe des Individualtests als Gruppentest durchführbar ist. Im dritten Untertest zeigten sich jedoch signifikante Leistungsunterschiede zu Gunsten der Individualbedingung. Die Frage worauf diese Unterschiede zurückzuführen sind, konnte im Rahmen der Studie nicht beantwortet werden. Es ist jedoch auszuschließen, dass sie aus der Möglichkeit des Nachfragens seitens des Testleiters resultieren.

## **Abstract**

The benefit of the economy in group tests compared to individual tests is widely known. Anyway there are only a few studies that outline, if it is possible to use an individual test as a group test and contra wise to use a group test as an individual test. In the following thesis 147 students of the third, fifth and eighth grade were randomly assigned to the individual- or the group testing situation. The results of the analyses indicate that from the AID 3, two out of three inspected subtests can also be used as a group test. In the third subtest there was a significant difference in the performance of the testes between the two conditions in favor of the individual test situation. The question wherefrom this benefit arises could not been explained in this study. Anyway it could be shown that it did not result from the opportunity that the test-administrator was permitted to inquire the testes, if they feel confident that their answer is definitely correct.

# Inhaltsverzeichnis

I.	Einleitung .....	1
II.	Theoretischer Teil .....	3
	1. Begriffliche Unterscheidung zwischen Einzel- und Gruppenverfahren.....	3
	2. Zeitbegrenzung und die Möglichkeit des Abschreibens .....	6
	3. Äquivalenzprüfung .....	7
	4. Störeinflüsse und Messungenauigkeiten .....	8
	5. Maßgeschneiderte Testsituation, Besondere Testmaterialien, Beobachtung und die Möglichkeit des Nachfragens .....	10
	6. Adaptives-, Branched- und Tailored- Testen .....	13
	7. Benachteiligung bestimmter Personengruppen .....	15
	8. Bedürfnis nach sozialen Vergleichsprozessen .....	16
	9. Anonymität gegenüber der Gruppe .....	17
	10. Anonymität gegenüber dem Testleiter .....	17
	11. Social facilitation – Leistungsaktivierung .....	18
III.	Empirischer Teil .....	22

1. Ziel der Untersuchung / Forschungshypothesen .....	22
2. Methode.....	23
2.1. Versuchsanordnung.....	23
2.2. Testmaterialien/ Erhebungsinstrumente.....	27
2.2.1. Kurzbeschreibung des AID 2 .....	27
2.2.1.1. Synonyme Finden .....	29
2.2.1.2. Funktionen Abstrahieren: .....	29
2.2.1.3. Antonyme: .....	30
2.2.2. Vorgabe und Konstruktion der Parallelttestformen.....	30
2.3. Testdurchführung .....	34
3. Ergebnisse .....	36
3.1. Stichprobenbeschreibung .....	36
4. Statistische Auswertung der Ergebnisse .....	38
5. Diskussion .....	45
6. Zusammenfassung.....	47
IV. Literaturverzeichnis.....	49
V. Anhang 1 .....	V-2



1.	Ansuchen um die Genehmigung einer Untersuchung in oberösterreichischen Schulen	
	V-2	
2.	Erteilte Genehmigung .....	V-4
3.	Elternbrief.....	V-6
4.	Lehrerbrief.....	V-8
VI.	Anhang 2 .....	VI-1
1.	t - Test zwischen t1 und t3 .....	VI-1
1.1.	Deskriptive Statistiken .....	VI-1
1.2.	Ergebnisse des t - Test inkl. Levene-Test .....	VI-1
2.	Multivariate Varianzanalyse.....	VI-2
2.1.	Deskriptive Statistiken .....	VI-2
2.2.	Levene-Test auf Gleichheit der Fehlervarianzen .....	VI-2
2.3.	Box-Test auf Gleichheit der Kovarianzenmatrizen .....	VI-3
2.4.	Multivariate Tests .....	VI-3
3.	Kreuztabellen .....	VI-4
3.1.	Schulstufe.....	VI-4
3.2.	Schulform.....	VI-4

3.3. Geschlecht.....	VI-5
3.4. Muttersprache .....	VI-6
4. Qualitative Untersuchung des Untertest Funktionen Abstrahieren .....	VI-6
VII. Abbildungsverzeichnis .....	VII-9
VIII. Tabellenverzeichnis.....	VIII-1

## I. Einleitung

Mit dieser Arbeit soll untersucht werden, ob es möglich ist, den für die Individualvorgabe konzipierten AID 3 als Gruppeninstrument vorzugeben. In der Literatur findet sich nur wenig über die Gültigkeit des Einsatzes von Individualverfahren als Gruppenverfahren.

Untersucht wurden die Leistungsunterschiede zwischen Volksschülern der dritten Schulstufe, Hauptschülern und Gymnasiasten der fünften, sowie der achten Schulstufe, in den Untertests *Funktionen Abstrahieren, Synonyme Finden* und *Antonyme* des adaptiven Intelligenzdiagnostikums 3 (AID 3, Kubinger & Holocher-Ertl, in Vorb.). Anhand dieser Untertests werden die sprachlichen Fertigkeiten der Kinder erfasst.

Wenn es möglich ist diesen Test als Gruppentest vorzugeben, wäre vor allem der immense Vorteil der Ökonomie zu nennen, der für den Einsatz der Gruppentestung spricht. In derselben Zeit könnten statt einem Schüler / einer Schülerin eine ganze Klasse getestet werden.

Falls die Leistungen der Schüler jedoch zwischen den Vorgabemodi (Einzel und Gruppentestung) stark streuen, muss eine neue Normierung für den Einsatz als Gruppentest vorgenommen werden, oder gar von einer Gruppenvorgabe Abstand genommen werden. In der folgenden Arbeit wird auf den bisherigen Stand der Forschung, sowie auf die Vor- und Nachteile der erwähnten Vorgabemodi eingegangen.

Die Arbeit gliedert sich im Wesentlichen in folgende Teile: im theoretischen Teil wird ein Überblick über den aktuellen Stand der Forschung, zu diesem Thema, gegeben. Im

empirischen Teil wird dargestellt, ob Leistungsunterschiede zwischen Einzel- und Gruppenbedingungen, bei der Anwendung des AID 3, existieren.

## II. Theoretischer Teil

### 1. Begriffliche Unterscheidung zwischen Einzel- und Gruppenverfahren

Um Verwirrungen zu vermeiden, sollen zuerst die Begriffe *Einzelfalldiagnostik* und *Gruppendiagnostik* skizzenhaft voneinander abgegrenzt werden:

Unter Einzelfalldiagnostik versteht Huber (1988, S. 162) „alle Ansätze innerhalb der Psychologischen Diagnostik, die auf die Prüfung singulärer und ideographischer Hypothesen ausgerichtet sind“. Dabei können vier Untersuchungseinheiten bei Einzelfallanalysen unterschieden werden: eine einzelne Person, eine Personengruppe (zum Beispiel eine Schulklasse oder Therapiegruppe), eine Organisation (Partei, Gewerkschaft, Jugendverband), oder eine Gesellschaft bzw. Kultur.

Auf die Unterschiede zwischen den möglichen diagnostischen Vorgehensweisen wird hier jedoch nicht eingegangen. Die Unterscheidung Einzelfalldiagnostik versus Gruppendiagnostik steht hier, nach der Definition von Huber, also noch in keinem Zusammenhang mit dem Vorgabemodus Einzel- versus Gruppentest.

Jäger (1986, S. 314) versteht unter *Einzelfalluntersuchung* die „Untersuchungen von einzelnen Personen, die mit dem Ziel verbunden sind, über Datengewinnung, Urteilsbildung und Entscheidung zu einer Bewertung über den weiteren Fortgang oder den Abbruch des therapeutischen Prozesses zu gelangen.“

Im Gegensatz dazu sieht Jäger das Ziel der Gruppendiagnostik darin gewonnene Daten so auszuwerten, dass „Statistiken von Personenstichproben resultieren, die als Schätzungen für Parameter von Populationen dienen“ (Jäger 1986, S. 85). Eine fast idente Definition gibt Kühn (1988) indem er postuliert, dass mit der Diagnostik von Gruppen auch das Bestreben verbunden ist allgemeine Aussagen zu treffen bzw. Gesetzmäßigkeiten zu finden und Häcker und Stapf (2004, S.437) definiert Gruppentests als die „Bez. für jenen Test, der so aufgebaut ist, dass er gleichzeitig einer großen Probandenanzahl vorgelegt werden kann“.

In der Literatur wird meist nicht genau auf die Unterscheidung zwischen Einzel- und Gruppentest eingegangen. Es existieren nur sehr wenige Untersuchungen, die sich mit diesem Thema befassen und nur wenige Autoren gehen explizit auf diesen Unterschied ein. Auch in den psychologischen Wörterbüchern werden diese Begriffe nur unzureichend voneinander abgegrenzt. So ist ein Individual-Tests laut Häcker und Stapf (2004, S.437) die Bezeichnung „für die individuelle Anwendung eines bestimmten Tests“. Hier werden weder Vor- noch Nachteile genannt, noch auf die speziellen Rahmenbedingungen oder die formalen Gestaltungsrichtlinien eingegangen, die mit dieser Unterscheidung einhergehen. Etwas genauer wird der Unterschied zwischen Individual- und Gruppentestung im Wörterbuch Psychologie (Fröhlich 2005, S. 227) herausgearbeitet. Hier werden Vorteile wie die Ökonomie der Testvorgabe als auch Nachteile, wie zum Beispiel jene von Störeinflüssen bei Gruppentests, wie folgt erwähnt:

Bezeichnung für solche Tests, die - meist aus Ökonomiegründen [vgl. hierzu (Kubinger 2009b, S. 98; Lienert & Ratz, 1998, S.12)] gleichzeitig bei Gruppen von Individuen angewendet werden können. Es handelt sich bei der Mehrheit der Fälle um sog. „Bleistiftpapier Tests“ d.h. solche, die schriftlich vollzogen werden; zum Unterschied von

individuellen Tests, die nur einzeln angewendet werden dürfen, da die Gegenwart anderer Individuen einen störenden Einfluß hätte.

Kubinger grenzt die Begriffe Gruppen- vs. Individualverfahren wie folgt voneinander ab: „Gruppenverfahren können von einem einzigen Testleiter (eventuell unter Hinzuziehen von Hilfspersonal) mehreren Testpersonen gleichzeitig vorgegeben werden, wohingegen „Individualverfahren“ auf Grund ihrer besonderen Gestaltung pro Testperson einen eigenen Testleiter erfordern - natürlich ist bei einem Gruppenverfahren auch eine Einzelvorgabe möglich“ (Kubinger 2009b, S.26).

Reindl (1998) ging der Frage nach, ob Niveauunterschiede in einem Leistungstest anhand des Vorgabemodus (Einzel- vs. Gruppentestung) feststellbar sind. Hierfür wurde AHS Schülern eine Kurzform des BIS (Berliner Intelligenz- Strukturtest, Jäger, Süß & Beauducel, 1997) vorgegeben. Mit diesem Test sollen die Fähigkeiten Verarbeitungskapazität, Einfallsreichtum, Bearbeitungsgeschwindigkeit, Merkfähigkeit, Sprachgebundenes Denken, Zahlengebundenes Denken, Anschauungsgebundenes Denken, figural-bildhaftes Denken und allgemeine Intelligenz erfasst werden. Es zeigten sich zwar einige signifikante Unterschiede zwischen den Versuchsbedingungen, allerdings sind diese Unterschiede geringer als der, von Reindl festgelegte, kritische Betrag von fünf T-Werten und daher nicht von inhaltlicher Relevanz.

In den nachfolgenden Abschnitten sollen nun vor allem die Vor- und Nachteile von Gruppen- als auch Individualtestsituationen bzw. -verfahren besonders beleuchtet und herausgearbeitet werden.

## 2. Zeitbegrenzung und die Möglichkeit des Abschreibens

Die Testvorgabe in der Gruppe setzt meist notwendiger Weise eine Zeitbegrenzung bei der Testbearbeitung voraus.

Bei speed-Tests wird durch eine begrenzte Zeitvorgabe neben der Qualität der Leistung auch die Geschwindigkeit erhoben, mit der eine Leistung erbracht wird. Bei power - Tests zählt dagegen nur, ob die Aufgaben richtig oder falsch gelöst wurden, und nicht wie viel Zeit die Person dafür benötigt. Reine power - Tests sind schon aus technischen Gründen kaum durchführbar, da jede Testvorgabe eine zeitliche Begrenzung haben muss (Rost, 2004, S. 43).

Die Testadministration unter Zeitbeschränkung birgt jedoch zwei Probleme mit sich, die sich komplementär zueinander verhalten: einerseits wird jenen Testpersonen, die langsamer arbeiten nicht ausreichend Zeit für die Bearbeitung der Items gegeben und andererseits kann es umgekehrt leicht passieren, dass jene Testpersonen, die früher mit der Bearbeitung fertig sind Unruhe erzeugen und so die Konzentration der langsamer Arbeitenden stören.

Eine zusätzliche Problemquelle bei der Gruppenvorgabe eines Tests wird explizit noch von Amelang und Schmidt-Atzert (2006) erwähnt, nämlich die Gefahr, dass die Testpersonen voneinander abschreiben und damit das Testergebnis beeinflusst wird. Bei Einzeltestungen ist das Abschreiben unter den zu testenden Personen hingegen nicht möglich und es können so keine individuellen Verfälschungen der Testergebnisse einer Person entstehen (deswegen ist auch, falls die Ergebnisse mehrerer Testpersonen aggregiert werden sollen, nicht mit einer Verfälschung oder Verzerrung der Ergebnisse der gesamten Leistung einer Personengruppe zu rechnen). Amelang und Schmidt-Atzert (2006) weisen in diesem Zusammenhang jedoch



auf die Möglichkeit hin, in Gruppentestungen Parallelformen zum Einsatz zu bringen, erwähnen aber auch, dass „viele Paralleltests nicht in dem Maße äquivalent [sind] wie dies wünschenswert wäre. Manche Autoren behelfen sich deshalb mit der Konstruktion sog. „Pseudoparallelformen“, bei denen zwar die selben Aufgaben verwendet, aber anders angeordnet werden“. (Amelang und Schmidt-Atzert, 2006, S. 368)

Da in der vorliegenden Arbeit die Äquivalenz der Testwerte zwischen der Vorgabe als Einzel- und als Gruppentestung untersucht wird, wurden zu diesem Zweck vier Paralleltestformen für die drei vorzuziehenden Untertests entwickelt, um die Möglichkeit des Abschreibens in der Gruppenbedingung zu mindern. Deshalb soll im nachfolgenden Abschnitt das Thema der Äquivalenz von Paralleltestformen genauer beleuchtet werden.

### 3. Äquivalenzprüfung

Die Frage der Äquivalenz kann sich auf die Konstruktion von Parallelformen eines Tests, auf die Äquivalenz verschiedener Testbedingungen oder auf die interkulturelle Äquivalenz von Verfahren beziehen (vgl. Wagner-Menghin, 2003). Im folgenden Abschnitt soll jedoch nur auf Äquivalenz im Sinne der Überprüfung der Gleichwertigkeit zweier Versionen eines psychologisch–diagnostischen Verfahrens genauer eingegangen werden.

Unter einer Parallelform versteht man „psychologische Testformen, die einander in allen relevanten testtheoretischen Kriterien äquivalent sind, so dass es im Einzelfall egal ist, welcher der Formen einer Testperson vorgegeben wird“ (Hahne, 2003, S. 316). Der Einsatz von Parallelformen eines Tests kann zum Beispiel dann notwendig sein, wenn eine Person wiederholt, oder eine Gruppe, getestet wird. Die Äquivalenz der verschiedenen

Testbedingungen bezieht sich hierbei vorrangig auf die „Gleichwertigkeit verschiedener Methoden der Administration eines Verfahrens“ (Wagner-Menghin, 2003, S.32). Dabei ist vor allem die Frage nach der Möglichkeit des Einsatzes von Computern für die Vorgabe und Auswertung psychologischer Tests, die in ihrer ursprünglichen Form als Papier- Bleistift Tests konzipiert wurden, zentral. Hier ist vor allem die *psychometrische Äquivalenz*, also „die Gleichwertigkeit der Parameter der Item- und Skalenwerte sowie der Reliabilitäts- und Validitätskennwerte bei verschiedenen Vorgabebedingungen“ (Wagner-Menghin, 2003, S.33) vorrangig. Dies gilt auch für die vorliegende Arbeit, da überprüft werden soll, ob eine Gleichwertigkeit der Parameter in der Einzel- und in der Gruppenvorgabe erreicht werden kann.

Ergänzend hierzu wird von Honaker (1988) noch die *perzeptierte Äquivalenz* gefordert, welche „die subjektive Bearbeitungsleichtigkeit der Pbn.“ erhebt (Häcker & Stapf, 2004, S. 61). Sie umfasst „Aspekte der Wahrnehmung, Einstellung und emotionalen Bewertung des Verfahrens durch die Testpersonen“ (Wagner-Menghin, 2003, S.33)

#### 4. Störeinflüsse und Messungenauigkeiten

Wie bereits in Abschnitt 2 erwähnt, kann die Testdurchführung durch eine Vielzahl von Störeinflüssen behindert werden. Zu den klassischen äußeren Störeinflüssen zählen zum Beispiel Lärm- oder Lichteinflüsse, welche in Testsituationen optimaler Weise auszuschalten, oder zumindest konstant zu halten, sind. Für den Testleiter ist es deshalb ein wesentlicher Punkt darauf zu achten, dass die Testung so weit wie möglich standardisiert von statten geht. In Gruppentestungen können jedoch noch zusätzliche Störeinflüsse durch andere Gruppenmitglieder, wie zum Beispiel unruhiges Sitzen oder lautes Umblättern,

hinzukommen. Diese störenden Einflüsse, die Auswirkungen auf die Testleistungen der Testpersonen haben können, können jedoch auch seitens des Testleiters nur schwer ausgeschaltet werden.

So schreibt auch Rost (2004):

Wenn die Testergebnisse nicht unabhängig vom Testleiter, von Situationsmerkmalen, von störenden Randbedingungen, vom Testauswerter oder sonstigen Personen ist, so wird der Test auch keine interne Validität und keine besonders hohe Meßgenauigkeit erlangen können. Insofern ist *Objektivität* der Testdurchführung eine *logische Voraussetzung für Reliabilität und Validität*. (S. 39)

Auch Scheuer (1988) geht in seiner Definition eines psychologischen Tests auf die Notwendigkeit standardisierter Bedingungen während der Testsituation ein. Er beschreibt einen psychologischen Test als ein spezifisches psychologisches Experiment<sup>1</sup>, das der Erkundung und Beschreibung individueller psychischer Merkmale dient. Laut Scheuer besteht es im Wesentlichen darin, „daß unter standardisierten Bedingungen eine Informationsstichprobe über einen Probanden (Pb) erhoben wird, die einen wissenschaftlich begründeten Rückschluß auf die Ausprägung eines oder mehrerer psychischer Merkmale des Pb gestattet.“ (Scheuer,1988, S. 208) Des Weiteren schreibt Scheuer, dass ein Test bezüglich seiner Durchführung, Auswertung und Interpretation standardisiert sein muss und deshalb Regeln zur Durchführung, Auswertung und zur Interpretation der Ergebnisse vorliegen

---

<sup>1</sup>Als Urvater des klassischen psychologischen Experiments gilt Wundt (1898). Es bezeichnet eine methodisch angelegte Untersuchungsanordnung und ermöglicht als einzige wissenschaftliche Versuchsanordnung zuverlässige Kausalaussagen (Ursache-Wirkungs-Beziehungen). Es muss messbare Ergebnisse liefern, nachvollziehbar, wiederholbar und objektiv sein.

müssen und fordert damit, dass ein Test ein objektives kontrolliertes Verfahren sein muss. (vgl. Scheur,1988)

In diesem Zusammenhang sei auch auf das Problem der Verrechnungssicherheit von Individual- bzw. Gruppenverfahren hingewiesen. Laut Kubinger (2009b, S.46) sind Gruppenverfahren meist notwendigerweise nach dem Multiple- Choice-Format gestaltet und somit verrechnungssicherer als Individualverfahren. In Individualverfahren wird hingegen das meist viel aufschlussreichere freie Antwortformat benützt und folglich weniger Verrechnungssicherheit erreicht.

Die Antworten beim freien Antwortformat in schriftlicher Weise können schwer leserlich sein, oder die Testpersonen können die Antworten orthographisch nicht korrekt wiedergeben, so dass eine eindeutige Verrechnung oft schwierig ist. Auch dieser Faktor beeinflusst die Messgenauigkeit des Tests und infolge auch die Objektivität der Testauswertung.

## 5. Maßgeschneiderte Testsituation, Besondere Testmaterialien, Beobachtung und die Möglichkeit des Nachfragens

Im Rahmen einer Einzeltestung hat der Testleiter die Möglichkeit auf die Testperson und ihre speziellen Probleme einzugehen, was, wie in diesem Kapitel gezeigt werden soll, von großem Vorteil sein kann. Auf der anderen Seite bringen Einzeltestungen den auf der Hand liegenden Nachteil der sinkenden Testleiterunabhängigkeit mit sich. (vgl. Kubinger 2009b, S. 39 ff) Amelang und Schmidt-Atzert (2006) führen dazu an, dass in Einzeltestsituationen die Möglichkeit besteht für die Testperson „möglichst günstige Bedingungen zu schaffen, um ...[ihr] Gelegenheit zu maximaler Entfaltung ...[ihrer] Fähigkeiten und Fertigkeiten zu

geben.“ (Amelang und Schmidt, 2006, S. 369) Dieser Weg soll laut den Autoren dann eingeschlagen werden, wenn eine Person in einer Bewährungssituation (hier: Testsituation) versagt hat und es gilt, experimentell die Annahme genereller Fähigkeits- und Fertigungsdefizite zu falsifizieren. Nach Holzkamp (1966, zitiert nach Amelang und Schmidt 2006, S. 370) ist der diagnostische Prozess dagegen so zu gestalten, dass er für die Testperson in der speziellen Bewährungssituation repräsentativ ist. „Nicht die speziellen Bedürfnisse des Probanden, sondern die kriterialen Bedingungen hätten danach die diagnostische Situation zu bestimmen“. So wäre zum Beispiel für einen einzelnen Schüler die Leistungsfeststellung in Form einer Klassenarbeit für eine Einzeluntersuchung nicht repräsentativ. Amelang und Schmidt (2006) folgern daraus, dass Testaufgaben wie externe Bedingungen, den im Kriterium herrschenden Anforderungen entsprechen sollen, um in Folge aus dem Testverhalten valide Vorhersagen auf das Kriteriumsverhalten treffen zu können.

Wie in vielen anderen Intelligenztests, die für die Individualvorgabe konzipiert sind ist es auch im AID 2 gestattet, dass der Testleiter, wenn die Testperson, hier Kinder im Alter von 6 bis 15.11 Jahren, keine eindeutig richtige Antwort gibt, mit Phrasen, wie „Erklär mir das genauer“, nachfragt. Außerdem ist es dem Testleiter gestattet die Instruktion während der Testphase zu wiederholen, wenn dieser den Eindruck hat, dass die Testperson die Aufgabenstellung aus den Augen verloren oder nicht richtig verstanden hat.<sup>2</sup> Dies ist im Rahmen einer Gruppentestung nicht möglich. Hier muss der Testleiter die Antworten als richtig oder falsch kodieren, ohne der Testperson durch Nachfragen die Möglichkeit zu geben ihre Antwort „selbständig“ zu korrigieren. Auch eine erneute Erläuterung der Instruktion

---

<sup>2</sup> An dieser Stelle sei auch auf das damit in Zusammenhang stehende Problem der Testleiterunabhängigkeit und der Objektivität verwiesen (s. Kubinger, 2009, S. 39ff), denn in Einzeltestungen stehen der Beobachter und der Beobachtete in einer sozialen Interaktion.

während der Testphase ist nicht möglich, da die anderen Testpersonen dadurch gestört würden. Wenn einem Kind in der Individualtestung beim Untertest *Funktionen Abstrahieren* beispielsweise die Frage gestellt wird: „Was ist das gemeinsame an Alterspension und Stipendium?“ Dann könnte das Kind antworten, dass man bei beiden Geld bekommt. Diese Antwort ist zwar nicht falsch, dennoch erfasst sie nicht den wesentlichen Aspekt, dass das Geld hier eine staatliche Unterstützung darstellt. Diese Antwort würde laut Manual als nicht gelöst verrechnet werden, da keine eindeutig korrekte Antwort erfolgt, die den zentralen Aspekt der staatlichen Unterstützung bei der Alterspension und beim Stipendium beschreibt. Im Rahmen einer Einzeltestung hätte die Testperson unter Umständen durch gezieltes Nachfragen ihre Antwort konkretisieren und somit unter Umständen die Aufgabe korrekt lösen können.

Ein weiterer Vorteil der Einzeltestung besteht darin, dass spezielle Testmaterialien, wie zum Beispiel Puzzles oder Würfel, wie sie in den Intelligenztests HAWIE-R, HAWIK-III und AID 2 Usus sind, eingesetzt werden können. Gerade jene Tests die auf dem Intelligenzkonzept von Wechsler basieren, welches von einem Verbal- und einem Handlungsteil ausgeht, sind in Gruppentestsituationen, auf Grund der eingeschränkten Möglichkeit der Materialvorgabe und der begrenzten individuellen Beobachtungsmöglichkeiten seitens des Testleiters, ungeeignet.

Daher liegt auch ein weiterer wesentlicher Vorteil der Einzeltestung auf der Hand nämlich, dass das Verhalten der Testperson während der Testung beobachtet werden kann und so während der Testung zusätzlich Hypothesen über das Arbeitsverhalten der Testperson generiert werden können. So kann zum Beispiel die Reaktion auf Frustration während der Testvorgabe eines Individualtests vom Testleiter beobachtet werden. Die Beobachtung stellt hier jedoch einen eigenständigen Prozess dar und ist nicht als Teil des psychologischen

Experiments (also der standardisierten Testsituation) anzusehen. Verhaltensbeobachtung ist „die auf das Verhalten eines oder mehrerer Menschen gerichtete, nicht dem Zufall überlassene, methodisch kontrollierte Wahrnehmung [...] mit der Absicht, dadurch etwas für die Persönlichkeit Charakteristisches zu erfahren“ (Arnold et al. 1997, S. 2454)

## 6. Adaptives-, Branched- und Tailored- Testen

Im Gegensatz zur konventionellen Vorgabe der Tests, bei der allen Testpersonen die gleichen Items in gleicher Reihenfolge zur Bearbeitung vorgelegt werden, wird beim adaptiven Testen die Itemschwierigkeit dem Leistungsniveau der Testperson sukzessive angepasst. Der Testperson wird anfänglich ein mittelschwieriges Item geboten und je nachdem, ob die Person in der Lage ist dieses zu lösen erfolgt als nächstes die Vorgabe eines schwierigeren oder leichteren Items. Nach der Bearbeitung jedes einzelnen Items erfolgt eine neue Schätzung des Fähigkeitsparameters der Person und es wird eine Aufgabe aus dem Itempool gewählt, deren Schwierigkeit der momentanen Schätzung der Fähigkeit der Person am ehesten entspricht, also ein Item, das für die Fähigkeitsschätzung der Person zu diesem Zeitpunkt „maßgeschneidert“ („tailored“) ist. Diese Anpassung der Itemschwierigkeiten an den „wahren“ Leistungsgrad der Testperson erfolgt im Laufe der Testvorgabe mit zunehmender Genauigkeit, indem der Fähigkeitsparameter der Testperson mit zunehmender Itemanzahl immer genauer bestimmt werden kann. Diese Art der Testvorgabe bringt den Vorteil mit sich, dass auch in extremen Leistungsbereichen große Messgenauigkeit erzielt wird. Im Vergleich dazu besteht bei der konventionellen Vorgabe der Aufgaben die Gefahr, dass die Testpersonen deutlich überfordert oder unterfordert werden, was sich wiederum auf die Motivation und somit auch auf die Leistungsgüte auswirken kann. Trotz der Tatsache, dass

unterschiedlichen Testpersonen, im Zuge des adaptiven Testens, auch unterschiedliche Items zur Lösung vorgelegt werden, können die Leistungen der Personen anhand ihrer resultierenden Fähigkeitsparameter miteinander verglichen werden.

Die Vorgabe der Items mittels adaptiven Testens bringt den Vorteil mit sich, dass die Testung ökonomischer ist als bei der konventionellen Vorgabe, bei der standardmäßig versucht wird einen sehr großen Fähigkeitsbereich abzudecken. Es müssen also bei der adaptiven Testung weniger Items von der Testperson bearbeitet werden um eine hohe Messgenauigkeit zu erreichen.

Das adaptive Testen ist jedoch an den Einsatz von Computern gebunden, mittels derer die fortlaufende Fähigkeitsschätzung berechnet wird. Eine spezielle Form des adaptiven Testens, die nicht an den Einsatz von Computern gebunden ist, stellt die Methode des so genannten „*Branched-Testings*“ dar.

Hier handelt es sich gegenüber dem *Tailored-testing* um eine suboptimale Strategie, bei der nicht nach der Bearbeitung jedes einzelnen Items das aktuell informativste Item ausgewählt und als nächstes vorgegeben wird, sondern die Items in Gruppen zusammengefasst werden und zwischen diesen je nach Leistungsgüte in fest verzweigter Weise verwiesen wird (Kubinger, 2009b, S. 108).

Beim *Branched-Testing* erfolgt die Anpassung an das Leistungsniveau der Testperson nicht nach jeder einzelnen Vorgabe eines Items, sondern erst nach einer Gruppe von wenigen Items. Je nachdem wie viele Items, einer solchen Gruppe, die Testperson im Stande ist zu lösen, wird nach einem fest verzweigten Schema die nächste Gruppe an Items ausgewählt, die dem Fähigkeitsbereich der Testperson am besten entspricht (vgl. Kubinger & Wild, 1988).



Als „Urvater“ des Branched-Testings kann der AID (aktuell AID 2) angesehen werden. Bei diesem wird der Testperson je nach Altersgruppe eine Aufgabengruppe, bestehend aus fünf Items, zur Lösung vorgelegt. Löst die Testperson keines oder nur eines der Items dieser Gruppe, ist davon auszugehen, dass ihr diese Aufgabengruppe zu schwierig ist und sie bekommt im nächsten Schritt eine leichtere Aufgabengruppe zur Lösung vorgelegt. Löst sie hingegen vier oder gar alle fünf Items aus dieser Gruppe, dann kann davon ausgegangen werden, dass diese Aufgabengruppe zu leicht für die Testperson ist und es wird ihr als nächstes eine Aufgabengruppe mit schwierigeren Items geboten. Bei der Lösung von zwei oder drei Items einer solchen Gruppe entspricht die Schwierigkeit der Items dem Leistungsniveau der Testperson und ihr wird folglich eine Aufgabengruppe mit ähnlich schwierigen Aufgaben zur Bearbeitung vorgelegt. In den nach dem Branched-Testing gestalteten Untertests des AID 2 werden der Testperson insgesamt drei solche Aufgabengruppen vorgegeben. So kann mit der Vorgabe von nur 15 Items eine gute Messgenauigkeit erzielt werden (vgl. Kubinger, 2009b, S. 108ff).

## 7. Benachteiligung bestimmter Personengruppen

Amelang und Schmid (2006) gehen explizit auf die Nachteile einer Gruppentestung ein und nennen als Beispiel zwei Untertests aus dem LPS bzw. dem PSB. Bei diesen Untertests wird die Wortschatzkenntnis einer Person dadurch erfasst, dass sie den in einem Wort eingebauten Rechtschreibfehler markiert und dadurch ihre Kenntnis des Wortes dokumentiert. Wenn aber eine Person das Wort kennt, seine orthographische Schreibweise der Testperson jedoch unbekannt ist, dann wird sie hier, durch die mangelnde Kenntnis der Schreibweise, benachteiligt. Des Weiteren gehen sie auf das Problem der mangelnden Lesefähigkeit und das

daraus resultierende verminderte Instruktionsverständnis ein: „Personen ohne ausreichende Lesefertigkeiten werden offenbar durch die in Gruppentests häufig schriftlich gegebenen Instruktionen in ihrer Leistungsfähigkeit beeinträchtigt und sollen besser mit Hilfe von Einzeltests untersucht werden, deren Aufgaben mündlich zu applizieren sind.“ (Amelang & Schmidt-Atzert, 2006, S 368)

Die Autoren erwähnen auch die Möglichkeit, dass sich Personen, in Gegenwart potentieller Konkurrenten, nicht eingestehen wollen die Instruktion nicht verstanden zu haben und damit ihre Erfolgchancen schmälern. Dieses Verhalten kann ebenfalls zu Störeinflüssen, welche durch die Gruppe erzeugt werden, gezählt werden und leitet auch direkt in den nächsten Abschnitt, welcher das Bedürfnis nach sozialen Vergleichsprozessen beleuchtet, über.

## 8. Bedürfnis nach sozialen Vergleichsprozessen

Laut Festingers Theorie der sozialen Vergleichsprozesse (1954), besteht bei Menschen das Bedürfnis ihre Meinungen durch den Vergleich mit Bezugspersonen oder Bezugsgruppen, die einem selbst hinsichtlich wichtiger Merkmale ähnlich sind, zu vergleichen. Steht eine solche Bezugsgruppe nicht zur Verfügung kann Verhaltensunsicherheit auftreten. Umgelegt auf die Situation der Einzeltestung ist es für Personen möglicherweise schwierig soziale Vergleichsprozesse herzustellen. Dies kann Meinungs- und Verhaltensunsicherheit im Bezug auf die Testsituation bewirken. Die Schulklasse als solche kann eine geeignete Bezugsgruppe für das Individuum darstellen und ein angenehmes und positives Gruppenklima schaffen, wodurch es zu einer Steigerung der Motivation und somit auch zu einer Steigerung der Leistung kommen kann. Es kann jedoch auch, wie im vorherigen Abschnitt beschrieben, zu gegenteiligen Effekten kommen.

## 9. Anonymität gegenüber der Gruppe

Vor allem bei schlechten Schülern kann die Angst, dass sie durch unzureichende Leistung vor ihren Klassenkameraden bloßgestellt werden, dazu beitragen, dass sie in Leistungstests schlechter abschneiden. Diese Angst vor dem sozialen Vergleich kann in einer Gruppentestung geschmälert werden, indem die Anonymität der Testergebnisse gegenüber der Gruppe versichert wird. Wenn den Schülern mitgeteilt wird, dass die Schulkameraden die Testergebnisse der anderen nicht erfahren, kann dies zu einer entspannten Leistungssituation und somit zu besseren individuellen Leistungen beitragen. Auch in der Individualtestung besteht nicht die Gefahr vor den Klassenkameraden bloßgestellt zu werden. Auf der anderen Seite müssen schlechte Leistungen hier gegenüber dem Testleiter offengelegt werden. Auf diese Problematik soll im nächsten Kapitel eingegangen werden.

## 10. Anonymität gegenüber dem Testleiter

Viele Testpersonen, vor allem Kinder, reagieren ängstlich und unsicher wenn sie psychologische Tests absolvieren sollen, da sie einer neuartigen Situation ausgesetzt werden. Testpersonen, die sich zum ersten Mal einer psychologischen Untersuchung unterziehen, fürchten vielleicht dem Psychologen ausgeliefert zu sein. Diese Ängste und Befürchtungen können sich im Leistungsniveau der Personen niederschlagen. Hier kann die Anonymität die man in der Gruppe erfährt, im Vergleich zu einer Individualtestung, angstreduzierend wirken. Durch die Gruppentestsituation kann also diese Angst entschärft werden. Die Testperson begreift sich als Teil der Gruppe und durch die Erfüllung des Bedürfnis‘ nach sozialem

Vergleich können auch Meinungs- und Verhaltensunsicherheiten, wie bereits im Abschnitt 8 erwähnt, reduziert werden.

## 11. Social facilitation – Leistungsaktivierung

Der Begriff *social facilitation* beschreibt das Phänomen, dass „bei bloßer Anwesenheit von anderen Personen gelegentlich beobachtete höhere Leistungen“ (Häcker & Stapf, 2004, S. 266) erbracht werden.

In einem von Allport (1920, zit. n. Herkner, 2001) durchgeführten Experiment konnte gezeigt werden, dass die Versuchspersonen, die in Anwesenheit anderer Personen Aufgaben lösten, in den meisten Fällen schneller arbeiteten, als die isoliert arbeitenden Versuchspersonen. Allport konnte so zeigen, dass die „Leistungsmenge“ in Gruppensituationen bei den meisten Aufgaben ansteigt. Herkner (2001) erklärt dieses Phänomen der Steigerung der Leistungsmenge dadurch, dass es sich bei jenen Aufgaben um „sehr gut gelernte, reibungslos ablaufende Tätigkeiten handelt“ (Herkner, 1991, S. 475). Bei komplexeren und unbekannteren Aufgaben, wie dem Erkennen falscher Syllogismen, waren jedoch die isoliert arbeitenden Personen im Vorteil.

In einer späteren Studie von Pessin (1933) mussten die Versuchspersonen Listen mit sinnfreien Silben auswendig lernen. Hier zeigte sich, dass jene Personen, die alleine lernten, die Silben rascher lernten und besser behielten, als jene die sich in einer Beobachtungssituation befanden. Allerdings zeigte sich auch, dass die Testpersonen in der Zuschauerbedingung zu einem späteren Zeitpunkt wesentlich mehr Silben behalten hatten. Zajonc (1965) erklärt dieses Phänomen anhand der *Theorie der sozialen Erleichterung*. Diese

geht davon aus, dass Leistung von Aktivierung abhängig ist und dass in Gruppensituationen in der Regel eine erhöhte Aktivierung vorherrscht und bei bloßer Anwesenheit anderer Personen, bei der Bearbeitung einfacher Aufgaben, bessere Resultate erzielt werden. Alternative Erklärungsansätze für dieses Phänomen liefern unter Anderem die *Bewertungs-Erwartungshypothese* (die beobachtete Person macht sich Gedanken über die Erwartungen der Zuschauer und ist erregt, da sie sich nicht blamieren möchte) und die *Ablenkungsthese* (die beobachtete Person beobachtet die anderen Personen in ihrem Umfeld wiederum selbst und ist deswegen von der Bearbeitung der Aufgabenstellung abgelenkt). (vgl. Zajonc, 1965)

Der Gegenbegriff zur *sozialen Erleichterung* ist das *soziale Faulenzen* (Latané, 1979). Sobald Individuen im Kollektiv mit anderen auf ein gemeinsames Ziel hinarbeiten und ihre Einzelleistungen nicht bekannt werden, reduziert sich ihre physiologische Anspannung, was zu einem Leistungsabfall bei der Bearbeitung einfacher und umgekehrt zu einer Leistungssteigerung bei komplexen Aufgaben führt. In einer Meta-Analyse zu Experimenten bezüglich des sozialen Faulenzens konnten Karau und Williams (1993) zeigen, dass Menschen in Gruppen weniger motiviert sind die volle Leistung zu erbringen. Später konnte Ohlert (2009) auch noch zeigen, dass soziales Faulenzen schon dann auftritt, wenn man sich, als Teil der Gruppe, alleine auf eine Aufgabe vorbereitet, dies zeigt, dass sich Individuen auf Gruppenaufgaben schlechter vorbereiten als auf Einzelaufgaben.

Sanders, Baron und Moore (1977) beleuchten in ihrer Studie das Phänomen des *Ablenkungs-Konflikts* (distraction-conflict), bei dem das Individuum einem Konflikt zwischen Aufgabenbearbeitung und Ablenkung ausgesetzt ist. Die Ablenkung (distraction) wird als jede Art von Reiz beschrieben, der irrelevant für die eigentliche Aufgabenbearbeitung ist, die Testperson aber dazu verleitet diesem irrelevanten Reiz Aufmerksamkeit zukommen zu

lassen. Diese Ablenkung kann sozialer oder nicht sozialer Natur, durch einen externen Stimulus oder internen Gedanken verursacht, sein. Der Konflikt wird als ein Verlangen beschrieben gleichzeitig, zwei verschiedenen Reizen, die ganze Beachtung schenken zu wollen. Es wird davon ausgegangen, dass diese Art von Konflikt zu einem erhöhten Antrieb führt.

Sanders, Baron und Moore (1977) führten diesbezüglich zwei Experimente durch. Im ersten Experiment lag der Fokus auf dem Bedürfnis nach sozialem Vergleich. Hier arbeitete ein Teil der Versuchspersonen isoliert in einem Raum, der andere Teil arbeitete in Kleingruppen zu zwei bis vier Personen. Des Weiteren wurde zwischen leichten und komplexen Aufgaben unterschieden. Signifikant bessere Ergebnisse waren bei den leichten Aufgaben bei jenen Personen zu verzeichnen, die zusammen mit anderen Personen in einem Raum die gleichen Aufgaben bearbeiteten. Die Hauptideen aus dieser Studie besagen, dass die reine Anwesenheit anderer Personen nicht ausreicht, um einen erhöhten Arbeitsantrieb zu erreichen. Erst wenn die Anwesenden idente Aufgaben bearbeiten, führt dies zu einer Leistungssteigerung.

In einem zweiten von ihnen durchgeführten Experiment wurde ebenfalls zwischen leichten und schwierigen Aufgaben unterschieden, wobei die Versuchspersonen den folgenden Bedingungen zugewiesen wurden: alleine in einem Raum arbeitend, in Anwesenheit anderer an identen Aufgaben (together-same) arbeitend oder in Anwesenheit anderer an verschiedenen Aufgaben (together-different) arbeitend.

Die Ergebnisse zeigen, dass die Versuchspersonen die in Anwesenheit anderer die Aufgaben bearbeiteten (together-different und together-same) sowohl bei den einfachen als auch bei den

komplexeren Aufgaben angaben, sich konkurrenzfähiger zu fühlen als jene Personen, die die Aufgaben isoliert bearbeiteten. Wie auch im ersten Experiment, konnte gezeigt werden, dass die Versuchspersonen, die die leichten Aufgaben bearbeiteten und in der together-same Versuchsbedingung waren, die besseren Leistungen erbrachten. Auch das Phänomen social facilitation zeigte sich nur bei jenen Personen in der together-same Bedingung.

Beide Experimente zeigen, dass die Aufmerksamkeit in Leistungssituationen nicht auf andere Personen gelenkt wird, wenn diese andere Aufgaben als man selbst bearbeitet, sondern dass die Aufmerksamkeit nur auf andere Personen gelenkt wird, wenn diese idente Aufgaben, wie man selbst bearbeiten. Diese Aufmerksamkeitsverschiebung, von der Aufgabe zu anderen Personen, führt bei leichten Aufgaben zu einer Leistungsverbesserung, bei schwierigen Aufgaben führt dies jedoch zu einer erhöhten Anhäufung von Fehlern und somit zu einer Leistungsverschlechterung im Vergleich zu den isoliert arbeitenden Personen.

### III. Empirischer Teil

#### 1. Ziel der Untersuchung / Forschungshypothesen

Wie im Theorieteil dieser Arbeit geschildert, gibt es keine eindeutigen Belege dafür ob psychologisch- diagnostische Verfahren, die für den Individualeinsatz konzipiert wurden, ohne weiteres auch in Gruppensituationen eingesetzt werden können. In der Literatur wurde dieses Thema noch nicht ausreichend behandelt, um eine klare Antwort geben zu können. Falls sich im Rahmen dieser Untersuchung zeigen sollte, dass sich das Leistungsniveau in Einzelsituationen signifikant von jenem in Gruppensituationen unterscheidet, müssten getrennte Normen für die zwei Vorgabemodi entwickelt werden oder es müsste gar Abstand von der Vorgabe dieser Untertests als Gruppentest genommen werden.

Nach Durchsicht der vorhandenen Literatur und unter Berücksichtigung des bisherigen Forschungsstandes resultierten folgende Forschungshypothesen:

- 1.) H1: Der Vorgabemodus (Einzel- vs. Gruppentestung) übt einen Einfluss auf die Leistung in den Untertests *Synonyme Finden*, *Funktionen Abstrahieren* und *Antonyme* des AID 3 aus.

Diese Hypothese bildet die Hauptfragestellung in der vorliegenden Arbeit.

Durch den speziellen Vorgabeplan, der in dieser Studie Anwendung findet, ergibt sich eine weitere interessante Hypothese:

- 2.) H1: Die Leistungen zwischen Testzeitpunkt 1 (kurz t1) und Testzeitpunkt 3 (kurz t3) der Einzeltestungen unterscheiden sich voneinander.



Sollte sich in dieser Studie herausstellen, dass bessere Ergebnisse in den Einzeltestungen, als in den Gruppentestungen vorliegen entsteht eine weitere zu prüfende Hypothese:

- 3.) H1: Die besseren Ergebnisse, in der Einzelbedingung, sind auf die Möglichkeit des Nachfragens, seitens des Testleiters, zurückzuführen.

## 2. Methode

Im folgenden Abschnitt soll nun die Versuchsanordnung genau dargestellt und erläutert werden. Ebenso werden die Erhebungsinstrumente und Testmaterialien detailliert dargestellt und auch eine Beschreibung des Stichprobenplans erfolgt in diesem Kapitel.

### 2.1. Versuchsanordnung

Geplant war es, 120 Schüler aus der dritten, fünften und achten Schulstufe, der Volksschule, der Hauptschule und des Gymnasiums, randomisiert den Bedingungen Einzel- vs. Gruppentest zuzuweisen. Die Einzeltestungen sollten zum Testzeitpunkt eins und zum Testzeitpunkt drei erhoben werden und die Gruppentestungen erfolgten zum Testzeitpunkt zwei (kurz t2). Auf die Darstellung des Randomisierungsplans wird an späterer Stelle eingegangen.

Dafür wurde Kontakt zu den Landesschulräten und Direktoren oberösterreichischer Schulen aufgenommen. Das Schreiben an den oberösterreichischen Landesschulrat befindet sich im

Anhang (s. Anhang 1, Abschnitt 1). Nach Erhalt der Genehmigung (s. Anhang 1, Abschnitt 2) und erneuter Rücksprache mit den Direktoren konnte die Testung an folgenden drei Schulen durchgeführt werden:

- Volksschule:  
Europaschule - Linz - Praxisvolksschule der Pädagogischen Hochschule Oberösterreich; Lederergasse 35, 4020 Linz;
- Hauptschule:  
Europaschule - Linz - Praxisvolksschule der Pädagogischen Hochschule Oberösterreich; Lederergasse 35, 4020 Linz;
- Gymnasium:  
Wirtschaftskundliches Realgymnasium der Franziskanerinnen Vogelweiderstraße 2-4, 4600 Wels

Vor der Testung wurde je ein Informationsschreiben für Eltern (s. Anhang 1, Abschnitt 3) und für Lehrer (s. Anhang 1, Abschnitt 4) verfasst, welches den Zweck der Untersuchung erörterte, die Anonymität der Testergebnisse zusicherte und eine Kontaktadresse für eventuelle Rückfragen enthielt.

Als Testleiter fungierten zwei weibliche Psychologiestudentinnen, die beide den AID 2 Zertifizierungskurs<sup>3</sup> besucht hatten und im Laufe ihres Studiums schon Erfahrung mit diesem psychologischen Testverfahren sammeln konnten.

---

<sup>3</sup> Dieser Kurs richtet sich in erster Linie an Studierende des Diplomstudiums in Psychologie an der Universität Wien und dient dazu, umfassende und detaillierte praktische Kenntnisse für die Anwendung und Interpretation des AID 2 und ein schriftliches Zertifikat darüber zu erwerben.

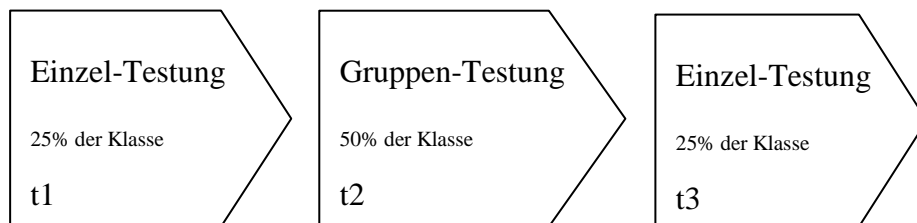
Vor Beginn der Testung werden noch die soziodemographischen Eckdaten - Alter, Geschlecht, Muttersprache und Staatsbürgerschaft - der Kinder mittels eines kurzen Fragebogen erfasst.

Der Untersuchungsplan gliederte sich wie folgt: Zu Beginn der ersten Schulstunde erfolgte eine kurze Vorstellung seitens der Testleiterinnen sowie eine kurze standardisierte Instruktion der Vorgehensweise. Im Anschluss erfolgte die randomisierte Zuteilung der Schüler zu den drei Gruppen. Mit Hilfe von Dosenspitzen in drei verschiedenen Farben wurde die Klasse in jeweils drei gleich große Gruppen aufgeteilt, indem alle Schüler der jeweiligen Klasse aus einer undurchsichtigen Tüte einen Dosenspitzen ziehen mussten. Die Farbe des Dosenspitzen entschied über die Zuordnung zur Gruppen- bzw. zu einer der Einzeltestbedingungen. Da die Kinder die Dosenspitzen behalten durften, könnte dies vielleicht zu einer Steigerung der Motivation zur Testteilnahme beigetragen haben, da aber alle drei Gruppen diese Verstärkung erhielten ist mit keiner systematischen Verzerrung dadurch zu rechnen.

Die Gruppentestungen fanden in einem Klassenzimmer, unter Aufsicht der zwei Testleiterinnen, statt. Für die Einzeltestungen wurden von der Schule geeignete Räumlichkeiten zur Verfügung gestellt, in denen ein ungestörtes Arbeiten möglich war. Der ursprüngliche Altersbereich des AID 2 umfasst Kinder im Alter von sechs bis 16 Jahren. Da bei Kindern der ersten und zweiten Schulstufe aber noch nicht von einer ausreichenden Lese-, Verständnis- und Schreibleistung ausgegangen werden kann, wurden in dieser Untersuchung nur Schüler ab der dritten Schulstufe für die Untersuchung herangezogen.

Um das Problem des Weitersagens der Aufgaben unter den Schülern und vor allem zwischen den Versuchsbedingungen zu reduzieren, wurden die Schüler zum einen gebeten, die Items

für sich zu behalten und nicht an ihre Mitschüler zu verraten, und zum anderen wurde ein spezielles Untersuchungsdesign gewählt: Zu Beginn des Schultages und nach erfolgreicher Randomisierung erfolgte die erste Hälfte der Einzeltestungen (t1). Im Anschluss wurde die Gruppentestung (t2) vorgenommen und danach erst die zweite Hälfte der Einzeltestungen (t3) durchgeführt, um so den Vorteil eines potentiellen Lerneffekts durch die Kenntnis der Aufgaben der letzten Gruppe durch die erste wett zu machen. Die nachfolgende Grafik Abbildung 3. soll den Testaufbau der Studie nochmals illustrieren:



**Abbildung 1.1** Darstellung der randomisierten Zuteilung der einzelnen Schulklassen zu den Testbedingungen

Die zwei Testleiter suchten mit den Kindern aus der Einzeltestbedingung jeweils einzeln einen ruhigen Raum, in dem die Testungen ungestört stattfinden konnten und der von der Schule zur Verfügung gestellt wurde (zum Beispiel: Biologiesaal, Physiksaal, ...), auf. In dieser Zeit fand der Unterricht mit den anderen Kindern wie gewohnt statt. Vor der Testung erhielten die Kinder jeweils standardisierte Informationen in denen ihnen die Anonymität der Testergebnisse zugesichert und eine Testinstruktion gegeben wurde. Die Instruktion war stark an die Originalinstruktion des AID 2 angelehnt.

Geplant waren Erhebungen in der dritten Klasse Volksschule sowie in der ersten und vierten Klasse Hauptschule und Gymnasium, wobei jeweils ca. 40 Kinder getestet werden sollten und dabei jeweils die Hälfte der Gruppenbedingung sowie der Einzelbedingung (hier je 10 Kinder zu t1 und t3) zugeordnet werden sollte.

Im nachfolgenden Abschnitt soll nun eine Beschreibung des AID 2, den daraus verwendeten Untertests, sowie der für diese Studie konstruierten Paralleltestformen, gegeben werden.

## 2.2. Testmaterialien/ Erhebungsinstrumente

Die hier zum Einsatz gelangten Untertests stammen aus dem AID 3, da dieser aber noch nicht veröffentlicht ist, wird im Folgenden eine genaue Darstellung des AID 2 gegeben, der dem AID 3 in seinem Aufbau und der Messintention im Wesentlichen gleicht.

### 2.2.1. Kurzbeschreibung des AID 2

Der AID 2 ist eine Testbatterie für Kinder und Jugendliche, im Alter von 6.0 bis 15.11 Jahren, zur Erfassung komplexer und basaler Kognitionen (*Intelligenz*). Er besteht aus elf Untertests. Für acht dieser elf Untertests existiert eine adaptive Aufgabenvorgabe, im Sinne des Branched-Testings (s. Abschnitt 6). Die anderen Untertests werden mittels konventioneller Vorgabe (vgl. dazu ebenfalls Abschnitt 6) durchgeführt. Der Großteil der Untertests lässt sich einem der folgenden vier Faktoren zuordnen: *Informationsverarbeitung in der gesellschaftlichen Umwelt*, *Informationsverarbeitung neuer Inhalte*, *Auffassungskapazität* und *(Re)Produktionsfähigkeit durch Strukturierung*. Außerdem lassen sich die Untertests in förderungsabhängige und förderungsunabhängige Tests unterteilen. Das Testkonzept ist an

jenen von Wechsler (1944) angelehnt (vgl. dazu Abschnitt 5). Im Gegensatz zu Spearman's Generalfaktortheorie (1927) steht Kubinger der Berechnung eines *Intelligenzquotienten*, als Durchschnittswert aller geprüften Fähigkeiten, kritisch gegenüber und präferiert die Profilinterpretation<sup>4</sup> der einzelnen Untertests, um so mögliche Stärken und Schwächen der Testperson besser identifizieren zu können (vgl. Kubinger, 2009a) und folgt damit einem förderungsorientierten Ansatz. Als globales Maß der Intelligenz bietet der AID 2 die sogenannte *Intelligenzquantität*<sup>5</sup>, die Aufschluss über die kognitive Mindestfähigkeit der Testperson gibt, sowie den *Range*<sup>6</sup> der Intelligenz, der die Differenziertheit aller erfassten Fähigkeiten widerspiegelt.

Eine weitere Besonderheit des AID 2 liegt in seinem Vorgabemodus welcher auf dem Prinzip des Branched-Testings (siehe Abschnitt 6) basiert und somit eine adaptive Vorgehensweise, ohne den Gebrauch eines Computers, erlaubt. Dieses Konzept ermöglicht es, den Testpersonen nur jene Aufgabenblöcke vorzulegen, die auf ihren individuellen Leistungsgrad abgestimmt sind. Dies bringt zum Einen den Vorteil erhöhter Ökonomie in der Testvorgabe mit sich, da die Testperson nur die Items bearbeitet, die in hohem Maße Aufschluss über ihre Fähigkeiten geben, zum Anderen werden Items, die wenig Information über die Fähigkeiten

---

<sup>4</sup> Profilinterpretation meint ursprünglich die graphische Darstellung von Testergebnissen. Heutzutage bezieht sich dieser Begriff auf eine detaillierte Gegenüberstellung der Ergebnisse in mehreren (Unter-) Tests bzw. Skalen und den daraus resultierenden Schlussfolgerungen. (vgl. Kubinger, 2009b, S. 82)

<sup>5</sup> Die Intelligenzquantität „gibt jeweils in Prozenträngen, den Anteil der Referenzpopulation an, mit dem andere Personen weniger niedrigere *schlechtere Untertestleistungen* erbringen als die betreffende Testperson.“ (Kubinger, 2009a, S. 31) Sie wird als die kognitive Mindestfähigkeit, mit der bei einer bestimmten Person gerechnet werden kann, interpretiert.

<sup>6</sup> „Der *Range* der „Intelligenz“ gibt jeweils in Prozenträngen den Anteil der Referenzpopulation an, mit dem andere Personen noch größere Differenzen zwischen *bester* und *schlechtester Untertestleistung* aufweisen als die betroffene Testperson (Kubinger, 2009a, S. 32)

der Testperson bieten, nicht vorgegeben und damit bleibt auch die Leistungsmotivation der Testpersonen aufrecht erhalten. (vgl. Kubinger, 2009a)

In dieser Studie kommen drei Untertests des neu entwickelten und noch nicht publizierten AID 3 zum Einsatz, für welche jedoch bereits aus Vorstudien Parameterschätzungen existieren und mit dem vorrangig sprachliche Fertigkeiten erfasst werden: *Synonyme Finden*, *Antonyme* und *Funktionen Abstrahieren*. Im Original des AID 2, sowie auch im AID 3, werden diese adaptiv vorgegeben. An dieser Stelle werden die Untertests kurz vorgestellt.

#### **2.2.1.1.      *Synonyme Finden***

Der Untertest 6, *Synonyme Finden*, prüft das „elementare Sprachverständnis“, das heißt inwieweit die Testperson die Bedeutung sprachgebundener Begriffe erfasst, bzw. über einen Wortschatz verfügt, der es ihr erlaubt solche Begriffe durch Alternativen auszudrücken.

Der Testperson werden Worte vorgegeben, für die sie jeweils ein anderes Wort mit (ziemlich) derselben Bedeutung finden soll. Die Verrechnung erfolgt dichotom (*richtig* oder *falsch*). Das heißt es werden lediglich die korrekten Antworten einer Testperson verrechnet (keine Teillösungen). Dieser Untertest misst nachweislich eindimensional. (vgl. Kubinger, 2009a, S. 17)

#### **2.2.1.2.      *Funktionen Abstrahieren:***

Mit dem Untertest 9, *Funktionen Abstrahieren*, soll die Fähigkeit erfasst werden, durch Abstraktion zu einer Begriffsbildung zu gelangen. Der Testperson werden jeweils zwei Objekte genannt, deren gemeinsame wesentliche Funktion sie erkennen und beschreiben soll. Es wurde darauf geachtet, dass die Lösungen geringe Ansprüche an den Wortschatz der

Testperson stellen. Auch hier erfolgt die Verrechnung dichotom (*richtig* oder *falsch*). Auch dieser Untertest misst ebenfalls nachweislich eindimensional. (vgl Kubinger, 2009a, S. 18)

### **2.2.1.3.     *Antonyme:***

Der Untertest, *Antonyme*, der im AID 2 noch nicht vorkommt, jedoch Bestandteil des AID 3 ist, prüft ebenfalls das elementare Sprachverständnis. Mit ihm wird die Fähigkeit erfasst „inwieweit die Testperson imstande ist, die Bedeutung eines Begriffes zu erfassen und die gegensätzliche Bedeutung dieses Begriffes wiedergeben zu können.“ (Weber, 2011, S. 36)

### **2.2.2. *Vorgabe und Konstruktion der Paralleltestformen***

Aus den Items der Untertests *Synonyme Finden* und *Funktionen Abstrahieren* wurden pro Altersgruppe vier Parallelformen entwickelt. Für den Untertest *Antonyme* existierten bereits zwei solcher Paralleltestformen weshalb hier keine neuen entwickelt wurden.

An dieser Stelle sei eine genaue Darstellung, der den Altersstufen entsprechenden Aufgabenblöcke der Untertests *Synonyme Finden* und *Funktionen Abstrahieren*, sowie deren Überschneidungsbereiche, gegeben. Jeder dieser Itemblöcke umfasst fünf Items (vgl. dazu auch die tabellarische Darstellung in Tabelle 2.1). Die grau markierten Blöcke in Tabelle 2.1 entsprechen zwar dem jeweiligen Altersbereich, sind aber nur der Vollständigkeit halber angeführt, da sie aufgrund ihrer sehr hohen oder geringen Lösungswahrscheinlichkeit in dieser Altersgruppe nicht vorgegeben wurden.



3. Schulstufe	2	10	3	11	12	6	5	4			
5. Schulstufe			3	11	12	6	5	4	13	7	
8. Schulstufe					12	6	5	4	13	7	8

**Tabelle 2.1** Darstellung der den Altersstufen entsprechenden Aufgabenblöcke und deren Überschneidungsbereiche der drei Untertests *Synonyme finden* und *Funktionen Abstrahieren*.

Aus diesem Pool wurden vier Parallelförmungen, pro Altersgruppe, für die vorliegende Untersuchung generiert (vgl. dazu Tabelle 2.2, Tabelle 2.3 sowie Tabelle 2.4). Wie hier ersichtlich werden sechs verschiedene Itemblöcke pro Schulstufe herangezogen, wobei ein Block immer in zwei verschiedenen Testformen, je Altersgruppe, vorkommt. Vergleicht man nun die dritte mit der fünften Schulstufe wird ersichtlich, dass sich hier vier Blöcke überschneiden, was einem Überschneidungsanteil von  $\frac{2}{3}$  entspricht (vgl. dazu Tabelle 2.2 und Tabelle 2.3). Das gleiche Konstruktionsprinzip wurde auch für die Gestaltung der Testformen der achten Schulstufe angewandt. Auch hier ergibt sich ein Überschneidungsbereich von  $\frac{2}{3}$  der Items, zwischen der fünften und achten Schulstufe (vgl. dazu Tabelle 2.3 und Tabelle 2.4). Zwischen der dritten und der achten Schulstufe besteht immerhin noch ein Überschneidungsbereich von 2 Blöcken, was einem Drittel entspricht. Dieser ist essentiell um eine Vergleichbarkeit zwischen den Testergebnissen und den Altersstufen gewährleisten zu können.

<b>Testform</b>	<b>Itemblock</b>
A	3, 4, 5
B	10, 4, 12
C	3, 11, 5
D	10, 11, 12

**Tabelle 2.2** Parallelttestformen für die dritte Schulstufe

<b>Testform</b>	<b>Itemblock</b>
A	4, 5, 6
B	11, 5, 13
C	4, 12, 6
D	11, 12, 13

**Tabelle 2.3** Parallelttestformen für die fünfte Schulstufe

<b>Testform</b>	<b>Itemblock</b>
A	5, 6, 7
B	12, 6, 8
C	5, 13, 7
D	12, 13, 8

**Tabelle 2.4** Paralleltestformen für die achte Schulstufe

Für den Untertest *Antonyme* bestehen bereits Paralleltestformen für Sechs- bis Siebenjährige, Acht- bis Elfjährige und Zwölf- bis 15-jährige Kinder. Den Kindern der dritten Schulstufe wurden die ersten drei Blöcke der Acht- bis Elfjährigen zur Bearbeitung vorgegeben und jene der fünften Schulstufe sollten die letzten drei Blöcke dieser Testform lösen. Den Schülern der achten Schulstufe wurden die letzten drei Blöcke der 12- bis 15-jährigen Kinder zur Bearbeitung vorgelegt. So wurden hier an Stelle von vier nur zwei Testformen vorgegeben, wobei alle Testformen 15 zu bearbeitende Items enthalten.

Pro Untertest gab es eine Zeitbegrenzung von jeweils 15 Minuten in der Gruppentestungsphase, um die Gefahr des Abschreibens und Redens zu minimieren. Für den Fall, dass alle Schüler unter der vorgegebenen Zeit mit der Bearbeitung fertig werden sollten, wurden die Testleiterinnen angewiesen abzurechnen und mit dem nächsten Untertest fortzufahren. Den Schülern wurde je eine der vier bzw. zwei Testversionen in der Reihenfolge *Synonyme Finden*, *Funktionen Abstrahieren* und *Antonyme* vorgegeben.

Alle drei Untertests bestehen aus jeweils 15 altersentsprechenden Aufgaben. Die Items sind je Untertest bezüglich ihrer Schwierigkeiten aufsteigend geordnet, um die Demotivation<sup>7</sup> während der Bearbeitung zu mindern.

In der Einzelversion wurden die Schüler für den Zeitraum der Testung einzeln aus der Klasse geholt und die Vorgabe erfolgte in einem separaten Raum, ohne Zeitbegrenzung, da hier die Gefahr des Schummelns nicht gegeben war.

### *2.3. Testdurchführung*

Nach erfolgreicher Vorstellung der Testleiter und Randomisierung der Versuchspersonen wurde mit der ersten Hälfte der Einzeltestungen (t1) begonnen. Die Kinder erhielten eine standardisierte Instruktion, die jener des AID 2 sehr ähnlich ist, wobei die Anonymität der Testergebnisse zugesichert wurde. Danach wurde mit der Testvorgabe begonnen, wobei die Items in den Einzeltestungen von den Testleiterinnen mündlich vorgegeben wurden und die Möglichkeit des Nachfragens seitens der Testleiterinnen bestand.

In der darauffolgenden Gruppenbedingung (t2) wurde mit der Hälfte der Klasse ein ruhiger Raum aufgesucht und die Testung erfolgte, nach Vorgabe der allgemeinen Instruktion und erneuter Zusicherung der Anonymität der Testergebnisse, im Beisein der zwei Testleiterinnen. Es war eine Bearbeitungszeit von 15 Minuten pro Untertest geplant. Allerdings zeigte sich, dass die Kinder für die Bearbeitung meist weniger Zeit brauchten, weshalb, ohne die vollen

---

<sup>7</sup> Demotivation könnte einerseits dann entstehen, wenn den Testpersonen zu Beginn zu schwierige Items vorgelegt werden, welche sie nicht lösen können und andererseits, wenn die Items am Anfang eines Tests eine zu hohe Lösungswahrscheinlichkeit haben, wodurch sich die Testperson nicht ernstgenommen fühlen könnten. In beiden Fällen kann sich dies demotivierend auf die Testpersonen auswirken.

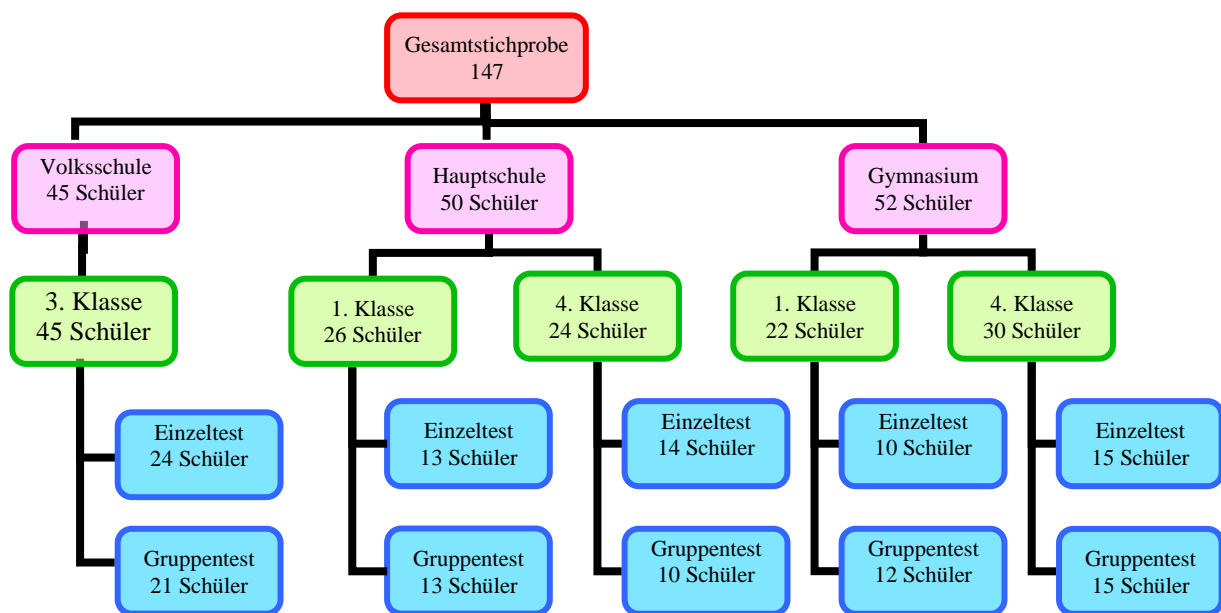
15 Minuten abzuwarten, zum nächsten Untertest übergegangen wurde. Nach der Durchführung der Gruppentestung erfolgte der zweite Teil der Einzeltestungen (t3). Das Vorgehen hier ist ident mit dem der Einzeltestungen zum Zeitpunkt (t1). Alle drei Testungen, wurden an ein und demselben Tag, in einer Schulklasse, durchgeführt, um den Kindern so wenig Zeit wie möglich zu geben, die Items ihren Mitschülern zu verraten.

## 3. Ergebnisse

### 3.1. Stichprobenbeschreibung

Insgesamt wurden 147 Schüler, aus drei Schulstufen (dritte Klasse Volksschule, erste Klasse Hauptschule bzw. Gymnasium und die vierte Klasse Hauptschule bzw. Gymnasium), welche randomisiert zu den Testbedingungen zugewiesen wurden, getestet. Aus der dritten Klasse Volksschule wurden insgesamt 45 Kinder getestet, wobei 24 Schüler der Einzelbedingung und 21 Schüler der Gruppenbedingung randomisiert zugewiesen wurden. Aus der ersten Klasse Hauptschule wurden jeweils 13 Kinder der Einzel- bzw. Gruppenbedingung zugewiesen und aus der vierten Klasse Hauptschule 14 der Einzel- und 10 der Gruppentestung. In der ersten Klasse Gymnasium erfolgten 10 Einzeltestungen und eine Gruppentestung mit 12 Schülern. In der vierten Klasse Gymnasium 15 Einzeltestungen und eine Gruppentestung mit 15 Teilnehmern statt. (vgl. dazu auch Abbildung 3.)

Die Auswertung erfolgte mittels SPSS Version 18 für Windows. Insgesamt wurden 147 Kinder aus der dritten, fünften und achten Schulstufe getestet, wobei 76 (51.7%) der Kinder der Individualbedingung und 71 (48.3%) der Kinder der Gruppenbedingung zugeordnet wurden. Diese Differenz in der Anzahl zwischen Einzel- und Gruppentestungen ergibt sich zum einen daraus, dass in einigen Klassen keine gerade Anzahl an Schülern vorhanden war und zum anderen daraus, dass nicht alle Schüler zum Zeitpunkt der Gruppentestung anwesend waren. Diese Kinder wurden deshalb der Gruppe der Individualtestungen zugewiesen.



**Abbildung 3.2** Darstellung der randomisierten Zuteilung der einzelnen Schulklassen zu den Testbedingungen

Die Verteilung in den Schulstufen setzt sich aus 30.6% der Schüler der dritten. Schulstufe, 32.7 % der fünften und 36.7 % der achten Schulstufe zusammen. Auch die Verteilung in den Schultypen kann, mit 30.6 % an Volksschülern, 34.0 % an Hauptschülern und 35.4 % an Gymnasiasten, als relativ ausgeglichen gewertet werden.

Ein deutliches Ungleichgewicht ist hingegen in der Geschlechterverteilung feststellbar. Diese Differenz, mit 11.5% an männlichen und 88.5% an weiblichen Schülern, resultiert daraus, dass das Gymnasium, in dem die Testungen stattfanden (Wirtschaftskundliches Realgymnasium der Franziskanerinnen), lange Zeit als reine Mädchenschule geführt wurde. Erst in den letzten Jahren hat man sich dazu entschlossen auch Buben aufzunehmen, weswegen die achte Schulstufe noch ein reiner Mädchenjahrgang ist und sich in der fünften Schulstufe nur vier Buben befinden. Auch in der Volksschule gibt es, mit 60.0% an Mädchen

und nur 40.0% an Buben, kein ausgewogenes Verhältnis zwischen den Geschlechtern. In der Hauptschule ist es, mit 50.0% männlicher und 50.0% weiblicher Schüler, hingegen ausgeglichen.

Da die für diese Untersuchung verwendeten Untertests vorrangig sprachliches Verständnis und sprachliche Ausdrucksfähigkeit prüfen, wurden nur Schüler mit österreichischer Staatsbürgerschaft in die Auswertung aufgenommen. 82.3% der getesteten Schüler sind deutscher Muttersprache. 8.2 % haben als Muttersprache Kroatisch, Serbisch oder Slowenisch und je 1.4% haben als Muttersprache Türkisch oder Slowakisch/Tschechisch. Die restlichen 6.8% haben eine andere Muttersprache als die oben erwähnten.

Insgesamt wurden 40 Schüler (27.2%) zum ersten Testzeitpunkt, 71 (48.3%) zum zweiten und 36 (24.5%) Schüler dem dritten Testzeitpunkt zugewiesen. Somit nahmen 51.7% der Schüler an den Einzeltestungen und 48.3% an den Gruppentestungen teil.

#### 4. Statistische Auswertung der Ergebnisse

Zur Beantwortung der Forschungshyopthese, ob sich die Ergebnisse der Einzeltestungen zum Testzeitpunkt eins und vom Testzeitpunkt drei auf Grund des Weitersagens der Items unter den Schülern signifikant voneinander unterscheiden, wurde ein t-Test für unabhängige Stichproben gerechnet. Hierfür wurden die bereits vorhandenen Fähigkeitsparameter zur Berechnung herangezogen. Die Ergebnisse werden in t-Werten, den zugehörigen Freiheitsgraden (df) und deren zweiseitigen p-Werten, mit einem gewählten Signifikanzniveau von  $\alpha = 0.05$ , berichtet.

Der t-Test (für gleiche Varianzen) ist in den Untertests *F Synonyme Finden* ( $t = -.94$ ;  $p = .35$ ;



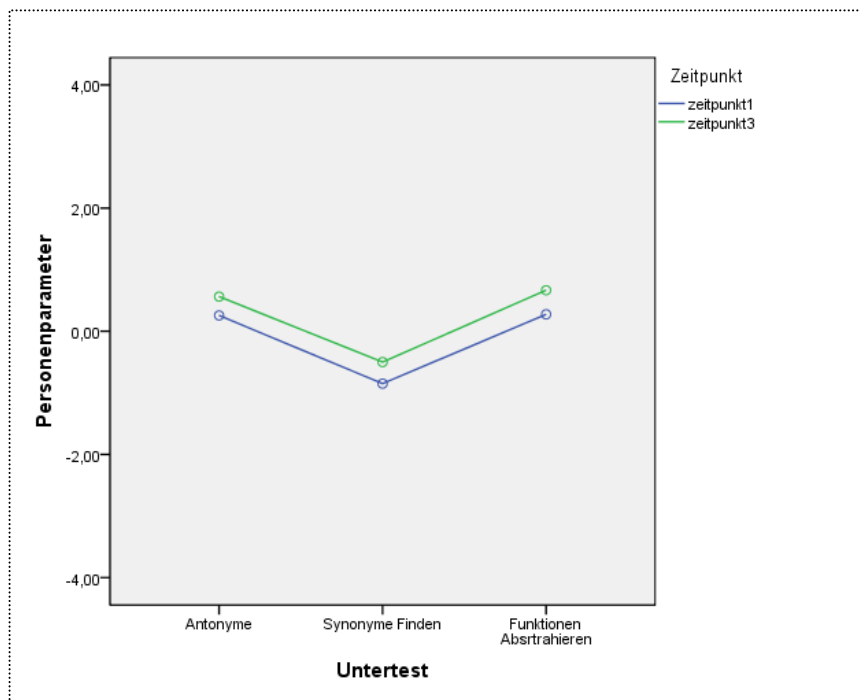
df = 74) und H *Funktionen Abstrahieren* ( $t = -1.18$ ,  $p = .24$ ,  $df = 74$ ) nicht signifikant. Im Untertests C *Antonyme finden* wurde der t-Test für ungleiche Varianzen zur Interpretation der Ergebnisse herangezogen (da der zugehörige Levene-Test signifikant ausfiel). Dieser war jedoch ebenfalls nicht signifikant ( $t = -.97$ ,  $df = 71.57$ ,  $p = .33$ ). Somit zeigten sich keine signifikanten Unterschiede in den einzelnen Untertests, zwischen den Testzeitpunkten (vgl. Anhang 2 VI.1). Auffällig hierbei sind jedoch die großen Standardabweichungen in den einzelnen Untertests.

Eine graphische Darstellung der zwei Zeitpunkte der Einzeltestungen befindet sich in

Abbildung 4.1. Daraus ist ersichtlich, dass die Testpersonen zum Zeitpunkt eins zwar in allen Untertests schlechter abschneiden als zum Testzeitpunkt drei, dieser Unterschied ist aber als statistisch nicht signifikant zu werten und könnte auf die anfängliche Vermutung zurückzuführen sein, dass die Kinder nach den ersten beiden Testdurchgängen (t1 und t2) die Items ihren Mitschülern verraten haben und die Schüler des Testzeitpunkts drei somit einen Vorteil gegenüber jenen der ersten Testung hatten. Der augenscheinlich naheliegende Schluss, dass im Untertest *Synonyme Finden* schlechteren Ergebnisse, im Vergleich zu den beiden anderen Untertests *Funktionen Abstrahieren* und *Antonyme*, resultieren, ist auf Grund der unterschiedlichen Kalibrierung pro Untertest nicht zulässig.

Des Weiteren ist nun die Frage der möglichen Differenz, zwischen der Individual- und der Gruppenvorgabe, von Interesse. Da keine signifikanten Unterschiede zwischen den zwei Zeitpunkten der Einzeltestung bestehen, können diese zu einem aggregiert werden, um den Unterschied zwischen den Werten der Einzeltestung und der der Gruppentestung genauer zu betrachten. Zu diesem Zweck wird eine Multivariate Varianzanalyse durchgeführt. Dabei

handelt es sich um ein Verfahren mit dem mehrere abhängige Variablen bei der Auswertung gleichzeitig berücksichtigt und eventuelle wechselseitige Beziehungen aufgezeigt werden können.



**Abbildung 4.1** Vergleich der Mittelwerte in den Testergebnissen zum ersten und dritten Testzeitpunkt

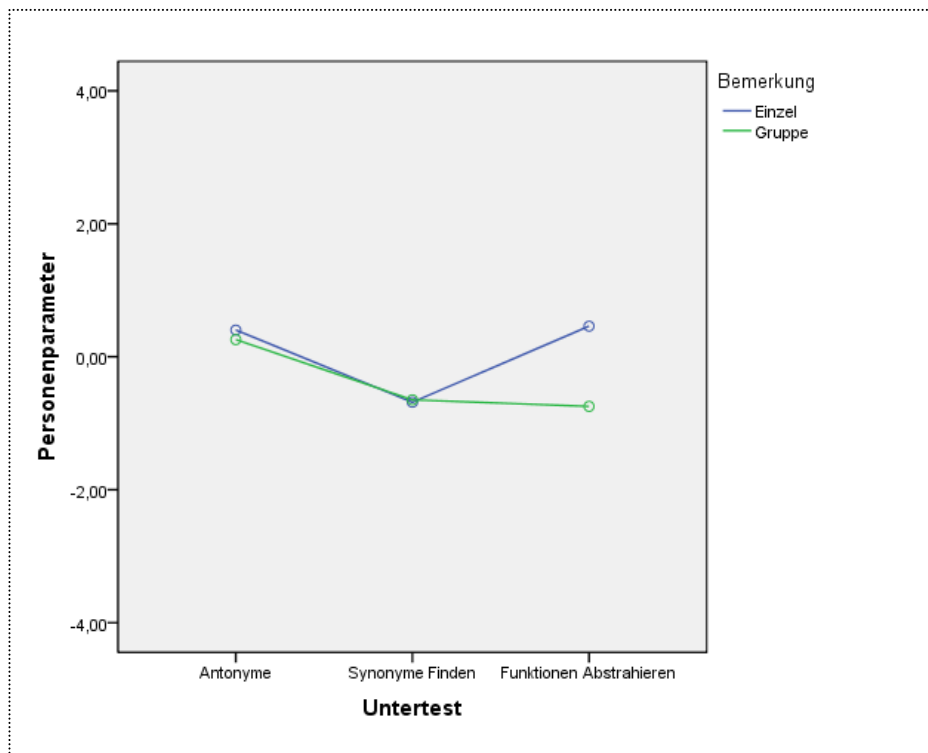
Der Box-M-Test (vgl. Anhang 2 VI.2.3) liefert ein signifikantes Ergebnis, das bedeutet, dass die Voraussetzung einer homogenen Varianz-Kovarianzmatrix nicht erfüllt ist. Es wurde daher eine multipel univariat gerechnet. Bei genauerer Betrachtung, der einzelnen Untertests in Tabelle 4.1 zeigt sich, dass in den Untertests, C *Antonyme* ( $F = .28$ ,  $df = 1$ ,  $df2 = 145$ ,  $p = .60$ ) und F *Synonyme Finden* ( $F = .01$ ,  $df1 = 1$ ,  $df2 = 145$ ,  $p = .91$ ), keine signifikanten Leistungsunterschiede, zwischen der Gruppen- und den Einzeltestbedingung, bestehen. Es

besteht allerdings ein signifikanter Unterschied, zwischen den Leistungen in der Individual- und der Gruppentestung, im Untertest H *Funktionen Abstrahieren* ( $F = 21.89$ ,  $df_1 = 1$ ,  $df_2 = 145$ ,  $p = .00$ ).

	Abhängige Variable	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Untertest	C	.76	1	.76	.28	.60
	F	.05	1	.05	.01	.91
	H	53.48	1	53.48	21.89	.00
Fehler	C	392.45	145	2.71		
	F	552.62	145	3.81		
	H	354.32	145	2.44		
Gesamt	C	409.52	147			
	F	618.00	147			
	H	410.00	147			

**Tabelle 4.1** Ergebnisse der multiplen univariaten Varianzanalyse

In Abbildung 4.2 werden die unterschiedlichen Leistungen in den Einzel- und Gruppenbedingungen in den drei vorgegebenen Untertests graphisch veranschaulicht. Hier wird deutlich, dass sich die Leistungen der Schüler, in den Subtests C *Antonyme* und F *Synonyme Finden*, nur unwesentlich voneinander unterscheiden. Im Untertest H *Funktionen Abstrahieren* ist der Unterschied zwischen den Vorgabemodi signifikant und auch in der Abbildung sehr deutlich zu erkennen.



**Abbildung 4.2** Vergleich der Mittelwerte der Testergebnisse zwischen der Einzel- und der Gruppentestbedingung

Diese Ergebnisse werfen die Frage auf, worauf diese signifikanten Leistungsunterschiede zwischen den Gruppen, im Untertest *Funktionen Abstrahieren*, zurückzuführen sind.

Um zu überprüfen ob sich die Kinder der Gruppentestung, im Hinblick auf die Zusammensetzung der soziodemographischen Variablen, von jenen der Einzeltestung unterscheiden, werden die Variablen Schulstufe, Schulform, Geschlecht und Muttersprache mittels Kreuztabellen verglichen (s. Anhang 2 VI.3.1 bis VI.3.4). Für die Variable Schulstufe ist der Chi-Quadratwert ( $\chi^2 = .82$ ,  $p = .81$ ) nicht signifikant, was bedeutet, dass sich die Schulstufen, in der Einzel- und in der Gruppenbedingung, nicht wesentlich voneinander unterscheiden. Auch bei den Variablen Schulform ( $\chi^2 = .43$ ,  $p = .81$ ) und Geschlecht ( $\chi^2 =$

2.30,  $p = .85$ ) gibt es keine signifikanten Unterschiede zwischen den Vorgabemodi. Die Variable Muttersprache wurde dichotomisiert (mit den Ausprägungen „Deutsche Muttersprache“, und „andere Muttersprache“). Auch hier zeigt sich kein signifikanter Unterschied ( $\chi^2 = .04$ ,  $p = .85$ ) zwischen den Einzeltestungen und der Gruppentestung.

Die Leistungsdifferenzen zwischen der Individual- und der Gruppentestung stehen damit also nicht mit den soziodemographischen Variablen in Zusammenhang.

Um die dritte Hypothese, dass die signifikanten Leistungsunterschiede zu Gunsten der Individualbedingung, zwischen der Einzel- und Gruppenvorgabe, im Untertest *Funktionen Abstrahieren*, auf das Nachfragen seitens der Testleiterinnen während der Testung, zurückzuführen sind, zu überprüfen, wurde wie folgt vorgegangen: Die Antworten der Schüler der Gruppentestung, wurden daraufhin bewertet, ob man im Fall einer Individualtestung bei dieser Antwort genauer nachgefragt hätte, was das Kind damit meint, um ihm so die Chance zu geben seine Antwort zu konkretisieren und möglicherweise zu korrigieren.

Alle Schüler der Gruppentestung bei denen die Testleiterinnen unter einer Individualbedingung mehr als fünf Mal (das entspricht einem Drittel der Items) nachgefragt hätten, wurden aus der Auswertung ausgeschlossen, mit dem Ziel nur jene Testwerte der Gruppenbedingung in der Auswertung zu behalten, die zumindest zu zwei Drittel eindeutig als richtig oder falsch zu klassifizieren sind. Interessant ist nun, ob sich die Mittelwerte in den beiden Bedingungen annähern oder ob sich die beiden Gruppen, nach Ausschluss jener Fälle bei denen man häufig nachgefragt hatte, noch immer signifikant voneinander unterscheiden.

Der Mittelwertsvergleich zeigt jedoch keine Annäherung zwischen dem Originalsample (MW =  $-.75$ , SD =  $1.68$ ,  $n = 71$ ) und dem Datenpool nach dem Ausschluss (MW =  $-.78$ , SD =  $1.74$ ,  $n = 49$ ). Auch der Welch- Test zur Überprüfung der Mittelwertsunterschiede zwischen den beiden Gruppen, ist signifikant ( $v = 17.46$ ,  $df1 = 1$ ,  $df2 = 87.63$ ,  $p = .00$ ). Die Leistungsunterschiede sind somit nicht auf das Nachfragen unter Individualbedingung zurückzuführen (vgl. AnhangVI.4).

## 5. Diskussion

Im Rahmen der vorliegenden Studie soll der Frage auf den Grund gegangen werden ob es möglich ist, die für die Individualvorgabe konzipierten Items der Untertests *Antonyme*, *Synonyme Finden* und *Funktionen Abstrahieren* des in Entwicklung befindlichen AID 3 in Form einer Gruppentestung vorzugeben. In der Literatur gibt es bislang nur wenige Forschung zu diesem Thema. Vor allem die Ökonomie spricht für eine Vorgabe eines Tests als Gruppenvorgabe. Die vorliegende Studie soll klären ob eine solche Umsetzung möglich ist, oder ob getrennte Normen für die Vorgabe als Gruppentest entwickelt werden müssen. Dabei stellt sich heraus, dass es keine signifikanten Unterschiede zwischen der Einzel- und der Gruppenvorgabe, der Untertests *Antonyme* und *Synonyme Finden*, gibt. Diese beiden Untertests könnten, auf Grund der Ergebnisse dieser Studie, auch zur Vorgabe in einer Gruppentestung genützt werden. Im Untertest *Funktionen Abstrahieren* zeigen sich hingegen signifikante Leistungsdifferenzen zwischen den beiden Vorgabemodi. Hier müssten getrennte Normen entwickelt werden oder es müsste gar Abstand von der Vorgabe als Gruppentest genommen werden. Worauf dieser Unterschied zurückzuführen ist, konnte im Rahmen dieser Studie nicht geklärt werden, es ist jedoch nicht anzunehmen, dass Ermüdung dabei eine wesentliche Rolle gespielt hat, da der besagte Untertest als zweiter der drei Untertests vorgegeben wurde. Auch der Einfluss von Störeinflüssen schien während der gesamten Testung konstant gering zu sein und die qualitative Überprüfung zeigte augenscheinlich kein erhöhtes Maß an Übereinstimmung der Antworten in den Gruppenbedingungen, was ein Hinweis auf Schummeln wäre. Die anfängliche Vermutung, dass der Leistungsvorteil in der Individualtestung dadurch zu Stande kommt, dass hier die Möglichkeit seitens der Testleiterinnen besteht, bei einer unklaren Antwort nachfragen zu können, konnte nicht

bestätigt werden. Auch die Zusammensetzung der Gruppen in den erhobenen Variablen bietet hierfür keine Erklärung. Weitere Untersuchungen zu diesem Thema sind für die Psychologische Diagnostik auch in Zukunft von hoher Relevanz, da sich in der vorliegenden Untersuchung zeigt, dass sich zwei der drei untersuchten Untertests als Gruppentest eignen, wodurch die Ökonomie der Testvorgabe immens gesteigert werden kann



## 6. Zusammenfassung

Die vorliegende Studie sollte die Frage klären ob es möglich ist den für die Individualvorgabe konzipierten Untertests *Antonyme*, *Synonyme Finden* und *Funktionen Abstrahieren* als Gruppentest durchzuführen. In der Literatur finden sich bislang nur wenige Studien zu diesem Thema. Reindl (1998) stellte in ihrer Diplomarbeit zwar signifikante Unterschiede zwischen den Vorgabemodi fest, allerdings waren diese kleiner als der von ihr festgelegte kritische Betrag von fünf t-werten und somit nicht von inhaltlicher Relevanz. Als Vorteil der Gruppentestung ist vor allem die Ökonomie zu nennen. In der gleichen Zeit könnte statt einem Kind eine ganze Klasse getestet werden. In der vorliegenden Studie wurden 147 Schüler der dritten, fünften und achten Schulstufe randomisiert entweder der Bedingung Einzel- oder Gruppenvorgabe zugewiesen. Die Ergebnisse zeigen, dass zwischen den Leistungen der Individual- und der Gruppenbedingung in den Untertests *Antonyme* und *Synonyme Finden* keine signifikanten Unterschiede bestehen. Diese beiden Untertests sind also für den Einsatz als Gruppentest verwendbar. Es zeigten sich allerdings signifikante Leistungsunterschiede zwischen den Vorgabemodi im Untertest *Funktionen Abstrahieren*. Worauf diese Leistungsunterschiede zurückzuführen sind konnte im Rahmen der Studie nicht beantwortet werden. Er ist allerdings auszuschließen, dass sie durch den Vorteil des Nachfragens in der Individualbedingung seitens der Testleiterinnen zustande kommen. Die Gruppenzusammensetzung der Individual- und der Gruppenbedingung wurden bezüglich der erhobenen soziodemographischen Variablen verglichen und auch hier zeigte sich kein signifikanter Unterschied zwischen der Individual- und der Gruppenbedingung. Da die vorliegende Studie nur ein kleines Datensample und nur drei Untertests des AID 3 umfasst

wären in Zukunft weitere Forschungen zu diesem Thema wünschenswert, da durch eine Gruppenvorgabe die Ökonomie der Testungen immens gesteigert werden kann.

## IV. Literaturverzeichnis

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

Arnold, W., Eysenck, H.J. & Meili, R. (Hrsg.). (1997). *Lexikon der Psychologie*. Augsburg: Bechtermünz.

Bortz, J. und Döring, N. (2006), *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.) Berlin: Springer,

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*; 7; 117.

Fröhlich, W.D. (2005). *Wörterbuch Psychologie*. München: Deutscher Taschenbuch Verlag GmbH & Co KG.

Gittler, G.& Wild, B., (1988). Der Einsatz des LLTM bei der Konstruktion eines Itempools für das adaptive Testen. In K.D. Kubinger (Hrsg.), *Moderne Testtheorie – Ein Abriß samt neuesten Beiträgen* (S. 115-139). München: Psychologie Verlags Union.

Häcker, H. & Stapf, K.-H. (Hrsg.) (2004). *Dorsch Psychologisches Wörterbuch*. Bern: Hans Huber.

Hahne, J. (2003) Äquivalenzprüfung. In K.D. Kubinger & R.S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik*. (S.32-34). Weinheim: Beltz.

Herkner, W. (2001). *Lehrbuch Sozialpsychologie*. Bern:Hans Huber.

Huber, H. P. (1988). Einzelfalldiagnostik. In: R. S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch*. München: Psychologie-Verlags-Union, 162-169.

Jäger, R.S. (1986). Der diagnostische Prozeß. *Eine Diskussion psychologischer und methodischer Randbedingungen*. Göttingen: Hogrefe.

Karau, S. J. & Williams, K.D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology* , 65(4), 681-706.

Kubinger, K. D. & Wild, B., (1988). Die Optimierung der Meßgenauigkeit beim „branched“-adaptiven Testen. In K.D. Kubinger (Hrsg.), *Moderne Testtheorie – Ein Abriß samt neuesten Beiträgen* (S. 187-217). München: Psychologie Verlags Union.

Kubinger, K.D. (2009a). *Adaptives Intelligenzdiagnostikum - Version 2.2 (AID 2)*. Göttingen: Beltz.

Kubinger, K.D. (2009b). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens* (2., überarb. u. erweiterte Aufl.). Göttingen: Hogrefe.

Kühn, R. (1988). Gruppendiagnostik. In: R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch*. München: Psychologie-Verlags-Union, 169-174.

Latané, B., Williams, K. & Harkins, S.: Many Hands Make Light the Work: The Causes and Consequences of Social Loafing. *Journal of Personality and Social Psychology* 37:822-832.

Ohlert, J. (2009). *Teamleistung. Social Loafing in der Vorbereitung auf eine Gruppenaufgabe*. Hamburg: Dr. Kovac

Pessin, J. (1933). The comparative effects of social and mechanical stimulation on memorizing. *American Journal of Psychology*, 132, 103-119.

Reindl, A. (1998). *Einzel- vs. Gruppenvorgabe: Niveauunterschiede im Leistungsbereich*. Unveröffentlichte Diplomarbeit. Universität Wien.

Rost, J. (2004). *Lehrbuch Testtheorie, Testkonstruktion* (2. Aufl.). Göttingen: Hans Huber.

Sanders, G.S., Baron, R.S. & Moore, D.L. (1977). Distraction and social comparison as mediators of social facilitation effects. *Journal of experimental social Psychology*, 14, 291-303.

Scheuer. (1988). Diagnostik als Testung. In: R.S. Jäger (Hrsg.), *Psychologische Diagnostik. Ein Lehrbuch*. München: Psychologie-Verlags-Union, 207-212.

Tewes, U. (1994). *HAWIE-R: Hamburg Wechsler Test für Erwachsene, Revision 1991*. Bern: Huber.

Tewes, U., Rossmann, P. & Schallberger, U. (2002). *Hamburg-Wechsler-Intelligenztest für Kinder - dritte Auflage: HAWIK-III*. Bern: Huber.

Wagner- Menghin, M. (2003) Äquivalenzprüfung. In K.D. Kubinger & R.S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik*. (S.32-34). Weinheim: Beltz.

Weber, B., (2011). *Konstruktion des sprachlichen Untertests „Antonyme finden“ für die Intelligenztestbatterie AID 3*. Unveröffentlichte Diplomarbeit. Universität Wien.

Zajonc; R. B. (1965). Social facilitation. *Science*, 149, 269-274.



# V. Anhang 1

## 1. Ansuchen um die Genehmigung einer Untersuchung in oberösterreichischen Schulen

Jana Goldmann  
Diplomandin  
Fakultät für Psychologie der Universität Wien  
Institut für Entwicklungspsychologie und  
Psychologische Diagnostik  
Tel: xxxx/xxx xx xx  
Email: jana.goldmann@gmx.at



---

Wien, März 2010

### **Ansuchen um die Genehmigung einer Untersuchung in oberösterreichischen Schulen**

#### 1. Inhalt der Untersuchung

In der geplanten Studie im Rahmen einer Diplomarbeit an der Universität Wien, Fakultät für Psychologie, Arbeitsbereich Psychologische Diagnostik, soll der Einsatz von drei Untertests der aktualisierten und noch unveröffentlichten Version der Intelligenz-Testbatterie AID 2 (*Adaptives Intelligenz Diagnostikum: in der aktuellen Fassung AID 2, Version 2.2, Kubinger, 2009*) auf die Tauglichkeit zur Gruppentestung überprüft werden. Die drei Untertests (*Antonyme, Synonyme Finden* und *Funktionen Abstrahieren*) erfassen die sprachliche Intelligenz und werden bei Kindern und Jugendlichen zwischen 6 und 16 Jahren zur allgemeinen Intelligenzdiagnostik sowie zur Diagnostik von Teilleistungsschwächen eingesetzt. Ziel der geplanten Studie ist es, herauszufinden, unter welchen Rahmenbedingungen solche psychologisch-diagnostischen Testungen stattfinden müssen, damit Kinder und Jugendliche ihr Leistungsoptimum zeigen können. Forschungen dieser Art sind wichtig, um die Schullaufbahnberatung und die Schulpsychologie laufend zu verbessern und dem neuesten Stand der Wissenschaft anpassen zu können.

#### 2. Versuchsplan

## 2.1 Stichprobe

In der geplanten Studie sollen insgesamt 120 Kinder im Alter von 8 bis 15 Jahren in den Schulstufen 3, 5 und 8 getestet werden. Je Schulstufe werden 40 Kinder in die Studie einbezogen, wobei sich die Stichprobe in den Schulstufen 5 und 8 gleichmäßig aus GymnasiastInnen und HauptschülerInnen zusammensetzt. Die Hälfte der Kinder wird im Klassenverband im Rahmen einer Gruppentestung, die andere Hälfte der Kinder jeweils einzeln im Rahmen einer Einzeltestung getestet.

## 2.2 Verwendete Materialien und Befragungsdauer

Alle Kinder werden mit drei Untertests (je 15 Aufgaben) der aktualisierten Version der Intelligenz-Testbatterie AID 2 getestet, wobei in der Gruppentestung eine Papier- Bleistiftversion von den Schülerinnen bearbeitet wird, und in der Einzeltestung dieselben Aufgaben mündlich vorgegeben und beantwortet werden. Die Untersuchung findet während der Schulzeit statt und dauert ca. eine Schulstunde.

## 3. Teilnehmende Schulen

Die Direktoren(innen) folgender Schulen haben ihr prinzipielles, mündliches Einverständnis zur Durchführung der Studie gegeben:

Volksschule:

- Europaschule - Linz - Praxisvolksschule der Pädagogischen Hochschule Oberösterreich; Lederergasse 35, 4020 Linz;

Hauptschule:

- Europaschule - Linz - Praxisvolksschule der Pädagogischen Hochschule Oberösterreich; Lederergasse 35, 4020 Linz;

Gymnasium:

- Wirtschaftskundliches Realgymnasium der Franziskanerinnen Vogelweiderstraße 2-4, 4600 Wels



## 2. Erteilte Genehmigung

# LANDESSCHULRAT FÜR OBERÖSTERREICH



A-4040 LINZ, SONNENSTEINSTRASSE 20

Frau  
Jana Goldmann  
Jana.Goldmann@gmx.at

Bearbeiterin:  
U. Wagner  
Tel: 0732/7071-2321  
Fax: 0732/7071-2330  
E-mail:lsr@lsr-ooe.gv.at

Ihr Zeichen	vom	Unser Zeichen	vom
---	2. 4. 2010	B5 – 14/17 – 2010	6. 4. 2010

**Untersuchung im Rahmen Ihrer Diplomarbeit  
zum Thema "Einzel- versus Gruppenvorgabe  
der Intelligenztestbatterie AID 2"**

Sehr geehrte Frau Goldmann!

Sie haben beim Landesschulrat für OÖ um Genehmigung der gegenständlichen Untersuchung an Volksschulen, Hauptschulen und allgemeinbildenden höheren Schulen in Oberösterreich angesucht.

Nach Prüfung der Unterlagen genehmigt der Landesschulrat für OÖ Ihre Erhebung unter den üblichen Bedingungen:

- Freiwilligkeit der Teilnahme
- Information der Erziehungsberechtigten
- Einhaltung der Datenschutzbestimmungen
- Übermittlung eines Kurzberichtes an den Landesschulrat für OÖ.

Bei Ihrer Kontaktaufnahme mit den Schulen verweisen Sie bitte auf diese Genehmigung.

Mit freundlichen Grüßen

Für den Amtsführenden Präsidenten

Dr. Lang eh.

### 3. Elternbrief

Jana Goldmann  
Diplomandin  
Fakultät für Psychologie der Universität Wien  
Institut für Entwicklungspsychologie und  
Psychologische Diagnostik  
Tel: xxxx/xx xx xx  
Email: jana.goldmann@gmx.at



---

Wien, März 2010

Liebe Eltern!

Viele Kinder werden im Laufe ihrer schulischen Karriere aus den unterschiedlichsten Gründen mit psychologischen Tests untersucht. Bei schulpsychologischen Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Schullaufbahnberatungen, Verhaltensprobleme im schulischen Kontext, Aufmerksamkeitsprobleme, Abklärung einer möglichen Legasthenie/Dyskalkulie etc. kommt dabei meist ein Intelligenztest zum Einsatz, um die intellektuellen Stärken und Schwächen des Kindes abschätzen zu können.

Im Rahmen meiner Diplomarbeit an der Fakultät für Psychologie, Universität Wien (betreut von Prof. Dr. Klaus Kubinger), führe ich nun in den nächsten Wochen mit Einverständnis der Direktion und der Lehrkräfte in der Klasse Ihres Kindes eine Testung mit einigen Aufgaben aus dem Intelligenztest AID 2 (*Adaptive Intelligenz Diagnostikum AID – aktuelle Version 2.2, Kubinger, 2009*) durch, die den Wortschatz der Kinder erfassen.

Ziel dieser Testung ist es, herauszufinden, unter welchen Rahmenbedingungen solche Testungen stattfinden müssen, damit Kinder und Jugendliche ihr Leistungsoptimum zeigen können.

Forschungen dieser Art sind wichtig, um die Schullaufbahnberatung und die Schulpsychologie laufend zu verbessern und dem neuesten Stand der Wissenschaft anpassen zu können.

Ich wende mich daher mit der Bitte an Sie, Ihr Kind an dieser Untersuchung teilnehmen zu lassen, vorausgesetzt natürlich, Ihr Kind ist einverstanden. Die Untersuchung findet während der Schulzeit statt und dauert ca. eine Schulstunde. Erfahrungsgemäß macht den Kindern die Mitarbeit an den Aufgaben viel Spaß.

Für eventuelle Rückfragen erreichen Sie mich gerne telefonisch oder per e- Mail (Tel: xxxx/xx xx xx, email: jana.goldmann@gmx.at).

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

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

Mit freundlichen Grüßen und herzlichen Dank im Voraus!

Jana Goldmann

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

Ich erkläre mich mit der Teilnahme meiner Tochter/meines

Sohnes \_\_\_\_\_, geboren am \_\_\_\_\_, an der  
*Testung für die Diplomarbeit von Jana Goldmann* einverstanden.

.....  
Datum, Unterschrift

## 4. Lehrerbrief

Jana Goldmann  
Diplomandin  
Fakultät für Psychologie der Universität Wien  
Institut für Entwicklungspsychologie und  
Psychologische Diagnostik  
Tel: xxxx/xx xx xx  
Email: jana.goldmann@gmx.at



---

Wien, März 2010

Liebe Lehrende!

Viele Kinder werden im Laufe ihrer schulischen Karriere aus den unterschiedlichsten Gründen mit psychologischen Tests untersucht. Bei schulpsychologischen Fragestellungen wie z.B. schulische Unter- bzw. Überforderung, Schullaufbahnberatungen, Verhaltensprobleme im schulischen Kontext, Aufmerksamkeitsprobleme, Abklärung einer möglichen Legasthenie/Dyskalkulie etc. kommt dabei meist ein Intelligenztest zum Einsatz, um die intellektuellen Stärken und Schwächen des Kindes abschätzen zu können.

Im Rahmen meiner Diplomarbeit an der Fakultät für Psychologie, Universität Wien (betreut von Prof. Dr. Klaus Kubinger), führe ich nun in den nächsten Wochen mit Einverständnis der Direktion in Ihrer Schule Testung mit einigen Aufgaben aus dem Intelligenztest AID 2 (*Adaptive Intelligenz Diagnostikum AID – aktuelle Version 2.2, Kubinger, 2009*) durch, die den Wortschatz der Kinder erfassen.

Ziel dieser Testung ist es, herauszufinden, unter welchen Rahmenbedingungen solche Testungen stattfinden müssen, damit Kinder und Jugendliche ihr Leistungsoptimum zeigen können.

Forschungen dieser Art sind wichtig, um die Schullaufbahnberatung und die Schulpsychologie laufend zu verbessern und dem neuesten Stand der Wissenschaft anpassen zu können.

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

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

Für eventuelle Rückfragen erreichen Sie mich gerne telefonisch oder per e- Mail (Tel: xxxx/xx xx xx, email: jana.goldmann@gmx.at).

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

Sollten Sie während der Testung manche Aufgaben sehen, so möchten wir Sie darum bitte absolute Verschwiegenheit über die Testaufgaben einzuhalten, da die Studie mit einem urheberrechtlich geschützten Test stattfindet und die Verbreitung und Weitergabe der Items ist gesetzlich untersagt ist.

---

Datum, Unterschrift der Lehrkraft

Mit der Bitte um Unterstützung, freundlichen Grüßen und herzlichen Dank im Voraus!

Jana Goldmann

## VI. Anhang 2

### 1. t - Test zwischen t1 und t3

#### 1.1. Deskriptive Statistiken

	Zeitpunkt	N	MW	SD	Standardfehler des MW
Untertest C	t1	40	,257	1,315	,208
	3	36	,564	1,422	,237
Untertest F	t1	40	-,85	1,511	,239
	t3	36	-,50	1,748	,291
Untertest H	t1	40	,28	1,467	,232
	t3	36	,67	1,414	,236

#### 1.2. Ergebnisse des t - Test inkl. Levene-Test

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Sig.	T	df	Sig.	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
								Untere	Obere	
C	Varianzen sind gleich	,000	,991	-,976	74	,332	-,307	,314	-,932	,319
	Varianzen sind nicht gleich			-,972	71,570	,334	-,307	,315	-,935	,322
F	Varianzen sind gleich	1,066	,305	-,936	74	,352	-,350	,374	-1,095	,395
	Varianzen sind nicht gleich			-,929	69,641	,356	-,350	,377	-1,102	,402
H	Varianzen sind gleich	,145	,704	-1,182	74	,241	-,392	,331	-1,052	,269
	Varianzen sind nicht gleich			-1,184	73,640	,240	-,392	,331	-1,051	,267

## 2. Multivariate Varianzanalyse

### 2.1. Deskriptive Statistiken

		<b>MW</b>	<b>SD</b>	<b>N</b>
Untertest C	Einzel	,402	1,366	76
	Gruppe	,259	1,899	71
	Gesamt	,333	1,641	147
Untertest F	Einzel	-,68	1,627	76
	Gruppe	-,65	2,249	71
	Gesamt	-,67	1,946	147
Untertest H	Einzel	,46	1,446	76
	Gruppe	-,75	1,679	71
	Gesamt	-,12	1,671	147

### 2.2. Levene-Test auf Gleichheit der Fehlervarianzen

	<b>F</b>	<b>df1</b>	<b>df2</b>	<b>Sig.</b>
Untertest C	9,227	1	145	,003
Untertest F	7,824	1	145	,006
Untertest H	2,145	1	145	,145



### 2.3. Box-Test auf Gleichheit der Kovarianzenmatrizen

Box-M-Test	15,830
F	2,579
df1	6
df2	150411,555
Sig.	,017

Prüft die Nullhypothese, daß die beobachteten Kovarianzen- matrizen der abhängigen Variablen über die Gruppen gleich sind.

a. Design: Konstanter Term + indiv\_gruppe

### 2.4. Multivariate Tests

Effekt		Wert	F	df	Fehler df	Sig.
Konstanter Term	Pillai-Spur	,450	38,982 <sup>a</sup>	3,0	143,0	0,0
	Wilks-Lambda	,550	38,982 <sup>a</sup>	3,0	143,0	0,0
	Hotelling-Spur	,818	38,982 <sup>a</sup>	3,0	143,0	0,0
	Größte charakteristische	,818	38,982 <sup>a</sup>	3,0	143,0	0,0
	Wurzel nach Roy					
indiv_gruppe	Pillai-Spur	,274	17,972 <sup>a</sup>	3,0	143,0	0,0
	Wilks-Lambda	,726	17,972 <sup>a</sup>	3,0	143,0	0,0
	Hotelling-Spur	,377	17,972 <sup>a</sup>	3,0	143,0	0,0
	Größte charakteristische	,377	17,972 <sup>a</sup>	3,0	143,0	0,0
	Wurzel nach Roy					

a. Exakte Statistik

b. Design: Konstanter Term + indiv\_gruppe

### 3. Kreuztabellen

#### 3.1. Schulstufe

		<b>Bemerkung * Schulstufe</b>				
		<b>Schulstufe</b>				
		<b>3.VS</b>	<b>1.HS/G</b>	<b>4.HS/G</b>	<b>Gesamt</b>	
Bemerkung	Einzel	Anzahl	24	23	29	76
		% innerhalb von Bemerkung	31,6%	30,3%	38,2%	100,0%
	Gruppe	Anzahl	21	25	25	71
		% innerhalb von Bemerkung	29,6%	35,2%	35,2%	100,0%
<b>Gesamt</b>		Anzahl	45	48	54	147
		% innerhalb von Bemerkung	30,6%	32,7%	36,7%	100,0%

#### Chi-Quadrat-Tests

	<b>Wert</b>	<b>df</b>	<b>Asymptotische Signifikanz</b>
Chi-Quadrat nach Pearson	,410 <sup>a</sup>	2	,815
Likelihood-Quotient	,410	2	,815
Anzahl der gültigen Fälle	147		

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 21,73.

#### 3.2. Schulform

		<b>Bemerkung * Schulform Kreuztabelle</b>				
		<b>Schulform</b>				
		<b>Volksschule</b>	<b>Hauptschule</b>	<b>AHS</b>	<b>Gesamt</b>	
Bemerkung	Einzel	Anzahl	24	27	25	76
		% innerhalb von Bemerkung	31,6%	35,5%	32,9%	100,0%
	Gruppe	Anzahl	21	23	27	71
		% innerhalb von Bemerkung	29,6%	32,4%	38,0%	100,0%

<b>Gesamt</b>	Anzahl	45	50	52	147
	% innerhalb von Bemerkung	30,6%	34,0%	35,4%	100,0%

### Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz
Chi-Quadrat nach Pearson	,427 <sup>a</sup>	2	,808
Likelihood-Quotient	,427	2	,808
Anzahl der gültigen Fälle	147		

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 21,73.

### 3.3. Geschlecht

#### Bemerkung \* Geschlecht Kreuztabelle

			Geschlecht		
			männlich	weiblich	Gesamt
Bemerkung	Einzel	Anzahl	21	55	76
		% innerhalb von Bemerkung	27,6%	72,4%	100,0%
	Gruppe	Anzahl	28	43	71
		% innerhalb von Bemerkung	39,4%	60,6%	100,0%
<b>Gesamt</b>	Anzahl		49	98	147
	% innerhalb von Bemerkung		33,3%	66,7%	100,0%

### Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz	Exakte Signifikanz	Exakte Signifikanz
Chi-Quadrat nach Pearson	2,302 <sup>a</sup>	1	,129		
Kontinuitätskorrektur <sup>b</sup>	1,801	1	,180		
Likelihood-Quotient	2,306	1	,129		
Exakter Test nach Fisher				,162	,090
Anzahl der gültigen Fälle	147				

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 23,67.

b. Wird nur für eine 2x2-Tabelle berechnet

### 3.4. Muttersprache

**Muttersprache \* Bemerkung Kreuztabelle**

			<b>Bemerkung</b>		
			<b>Einzel</b>	<b>Gruppe</b>	<b>Gesamt</b>
Muttersprache	Deutsch	Anzahl	63	58	121
		% innerhalb von Muttersprache	52,1%	47,9%	100,0%
	andere	Anzahl	13	13	26
		% innerhalb von Muttersprache	50,0%	50,0%	100,0%
<b>Gesamt</b>	Anzahl		76	71	147
	% innerhalb von Muttersprache		51,7%	48,3%	100,0%

**Chi-Quadrat-Tests**

	<b>Wert</b>	<b>df</b>	<b>Asymptotische Signifikanz</b>	<b>Exakte Signifikanz</b>	<b>Exakte Signifikanz</b>
Chi-Quadrat nach Pearson	,037a	1	,848		
Kontinuitätskorrekturb	,000	1	1,000		
Likelihood-Quotient	,037	1	,848		
Exakter Test nach Fisher				1,000	,509
Anzahl der gültigen Fälle	147				

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 12,56.

## 4. Qualitative Untersuchung des Untertest Funktionen Abstrahieren

**Gruppenstatistik des Untertests Funktionen Abstrahieren ohne Ausschluss**

<b>Gruppe</b>	<b>N</b>	<b>MW</b>	<b>SD</b>	<b>Standardfehler des MW</b>
---------------	----------	-----------	-----------	------------------------------

PersPUH	Einzel	76	,46	1,446	,166
	Gruppe	71	-,75	1,679	,199

**Gruppenstatistik des Untertests Funktionen Abstrahieren MIT Ausschluss jener  
Schüler bei denen öfter als fünf Mal nachgefragt worden wäre**

Gruppe		N	MW	SD	Standardfehler des MW
PersPUH	Einzel	76	,46	1,416	,162
	Gruppe	49	-,78	1,738	,248

**Robuste Testverfahren zur Prüfung  
auf Gleichheit der Mittelwerte**

PersPUH				
	Statistik <sup>a</sup>	df1	df2	Sig.
Welch- Test	17,463	1	87,630	,000

a. Asymptotisch F-verteilt



## VII. Abbildungsverzeichnis

Abbildung 1.1 Darstellung der randomisierten Zuteilung der einzelnen Schulklassen zu den Testbedingungen .....	26
Abbildung 4.1 Vergleich der Mittelwerte in den Testergebnissen zum ersten und dritten Testzeitpunkt .....	40
Abbildung 4.2 Vergleich der Mittelwerte der Testergebnisse zwischen der Einzel- und der Gruppentestbedingung .....	42

## VIII. Tabellenverzeichnis

Tabelle 2.1 Darstellung der den Altersstufen entsprechenden Aufgaben-blöcke und deren Überschneidungsbereiche der drei Untertests <i>Synonyme finden</i> und <i>Funktionen Abstrahieren</i> . .....	31
Tabelle 2.2 Paralleltestformen für die dritte Schulstufe.....	32
Tabelle 2.3 Paralleltestformen für die fünfte Schulstufe.....	32
Tabelle 2.4 Paralleltestformen für die achte Schulstufe.....	33
Tabelle 4.1 Ergebnisse der multiplen univariaten Varianzanalyse .....	41



## Curriculum Vitae

### Persönliche Daten

Name: Jana Goldmann  
Geboren: Wien, 12.10.1984  
Staatsbürgerschaft: Österreich  
E-Mail: [Jana.Goldmann@gmx.at](mailto:Jana.Goldmann@gmx.at)

### Schulische Ausbildung

2003 Matura Gymnasium, 1020 Wien Kleine Sperlasse  
Seit 2003 Studium Psychologie, 1010 Wien Universität Wien

### Berufserfahrung

2003 – 2006 *Vindobona Kabarett – Restaurant, 1200 Wien, Kellnerin*  
2006 *Cafe Latte, 1070 Wien, Kellnerin*  
2006 – 2007 *City-Rooms-Dennstedt, 1080 Wien, Rezeptionistin*  
SS 2008 Beendigung des ersten Studienabschnitts  
Juli 2008 Pflichtpraktikum am *Institut für Psychosomatik*, Leitung Dr. Margot Schmitz  
Seit August 2008 Festanstellung am *Institut für Psychosomatik* im Umfang von ca.30 Stunden pro Woche, Aufgabenbereich: Mitbetreuung von klinischen Studien im Bereich: Major Depression, Schizophrenie, Bipolare Störungen, ADHD, Demenz und MCI