



universität
wien

Diplomarbeit

Titel der Arbeit

„Kontrollstudie zu Reindl (1998): Einzel- vs.
Gruppenvorgabe: Niveauunterschiede im Leistungsbereich“

Verfasserin

Raffaella Neustifter

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im September 2011

Studienkennzahl: 298

Studienrichtung: Psychologie

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

INHALTSVERZEICHNIS

1. EINLEITUNG.....	5
2. THEORETISCHER TEIL	9
2.1 ÜBERBLICK ÜBER DIE PSYCHOLOGISCHE DIAGNOSTIK.....	11
2.1.1 <i>Intelligenztests und Leistungsdimensionen.....</i>	<i>12</i>
2.1.2 <i>Normierung psychologischer Tests</i>	<i>13</i>
2.1.3 <i>Die Relevanz der Leistungsdiagnostik im speziellen für Kinder</i>	<i>15</i>
2.2 EINZEL- VS. GRUPPENTESTUNG	16
2.2.1 <i>Vorteile der Einzeltestung</i>	<i>17</i>
a) Einsatz besonderer Testmaterialien	17
b) besseres Eingehen auf die Testperson/en	17
c) weniger Störeinflüsse	17
d) keine Möglichkeit ab zu Schreiben	18
e) keine soziale Hemmung.....	18
2.2.2 <i>Vorteile der Gruppentestung</i>	<i>19</i>
a) größere Testleiterunabhängigkeit	19
b) höhere Ökonomie	19
c) Anonymität gegenüber dem Testleiter	19
d) leichter Zugang zu Normierungsdaten	20
2.2.3 <i>Gründe für mögliche Unterschiede im Leitungsniveau zwischen Einzel- und Gruppenvorgabe.....</i>	<i>20</i>
a) Der Testleiter als Störvariable	21
b) Die Auswirkung des Testpersonenbefindens	23
2.3 DIPLOMARBEIT VON REINDL 1996: <i>EINZEL- VS. GRUPPENVORGABE:</i> <i>NIVEAUUNTERSCHIEDE IM LEISTUNGSBEREICH.....</i>	25
3. EMPIRISCHER TEIL	27
3.1 ZIEL DER UNTERSUCHUNG (HYPOTHESEN).....	28
3.2 UNTERSUCHUNGSPLAN.....	30
3.3 STICHPROBE	33
3.4 ERHEBUNGSINSTRUMENTE	36
3.4.1 <i>WIT – 2</i>	<i>36</i>

3.4.2	<i>Befindensfragebogen</i>	38
3.5	TESTLEITERINNEN.....	40
3.6	DURCHFÜHRUNG DER UNTERSUCHUNG	42
3.7	ERGEBNISSE	43
4.	DISKUSSION UND AUSBLICK	50
5.	ZUSAMMENFASSUNG	53
	ABSTRACT	56
	LITERATUR	57
	ANHANG	61

1. Einleitung

Bei der hier vorliegenden Diplomarbeit, handelt es sich um eine Kontrollstudie zur Diplomarbeit von Andrea Reindl aus dem Jahr 1998. Damals wurde der Frage nachgegangen, ob es legitim sei, anzunehmen, dass Leistungen von Testpersonen bei einem psychologischen Test nicht vom Vorgabemodus (Einzel- vs. Gruppentestung) abhängig sind. Die Praxis zeigt, dass viele Testautoren diesen Umstand ohne ausreichende Voruntersuchungen als gegeben annehmen und keine getrennten Normen für die Einzel- und Gruppenvorgabe zur Interpretation der Testwerte anbieten. Die Frage ob Testpersonen, welche mittels eines Testes im Einzelsetting getestet wurden, vergleichbare Ergebnisse erzielen, wie Testpersonen im Gruppensetting, stand damals im Raum. In der Literatur zum Thema Verwendung von Einzel- oder Gruppenverfahren lassen sich Sätze wie: „...- natürlich ist in einem Gruppenverfahren auch die Einzelvorgabe möglich.“ (Kubinger, 2006) finden. Die bisherigen Ergebnisse zum Thema sprechen teilweise sogar eher für unterschiedliche Testleistungen je nach Vorgabemodi (siehe zum Beispiel Pessin, 1933, zitiert nach Herkner, 2001). Zumindest kann die These aufgrund der bisherigen Forschung noch nicht ganz verworfen werden.

Die Konsequenz unterschiedlicher Testergebnisse wäre für die Normierung von psychologischen Leistungstests (siehe Kapitel 2.1.2) von großer Bedeutung. Sollte man davon ausgehen müssen, dass Testpersonen in den verschiedenen Vorgabemodi relevant andere Leistungen erbringen, müssten die Testautoren getrennte Normen für die Einzel- und die Gruppenvariante bereitstellen. Da des Weiteren Kinder in vielen Bereichen nicht mit Erwachsenen verglichen werden können, soll sich die Fragestellung insbesondere auf das Kindesalter beziehen. Psychologische Leistungstestung berührt Kinder in so vielen Bereichen und ist somit von essentieller Bedeutung. Aus diesem Grund sollte der Thematik Beachtung geschenkt werden (siehe Heller, 1984). Ein paar Bereiche seien zum Verständnis angeführt: Schulleistungstests, Hochbegabungsdiagnostik, Schullaufbahnberatung, Diagnostik von Teilleistungsstörungen, Weiterentwicklung des Schulsystems und viele mehr.

Die Diplomarbeit von Reindl (1998), sollte dazu einen ersten Anhaltspunkt bieten. Die Ergebnisse von 1998 bezogen sich auf eine Kurzform des BIS-T (Berliner Intelligenz Struktur Test) und zeigten, dass sich im Rahmen dieser Testungen keine Unterschiede zwischen den beiden Vorgabemodi finden ließen und somit das Verfahren ohne getrennte Normierung im Rahmen einer Einzel- als auch Gruppenvorgabe zu verwenden ist (siehe Kapitel 2.4). Auf Grund dieser Studie ist an zu nehmen, dass sich das Leistungsniveau in Einzeltestsituationen nicht von dem in Gruppentestsituationen unterscheidet. Von Reindl (1998) wurde neben dem alleinigen Unterschied bezüglich des Vorgabemodus, auch noch ein möglicher Einfluss der Testleiter getestet, da in der Literatur häufig das Problem der Testleitereffekte beschrieben

wird. Dieses Phänomen soll auch in der vorliegenden Studie mitbehandelt werden. Nicht nur der Unterschied zwischen den Testleistungen in der Gruppen- oder der Einzeltestung ist von Interesse, sondern auch, ob diese möglichen Unterschiede durch Testleitereffekte hervorgerufen werden (hier wird im speziellen auf den Testleitererwartungseffekt Bezug genommen). Nach der Studie von Reindl (1998) gab es noch weitere ähnliche Versuchsreihen, welche das Thema „Einzel- vs. Gruppenvorgabe“ in verschiedenster Weise thematisierten. Beispielsweise soll Kiesel (2009) erwähnt werden. Hier wurde der Einfluss von Persönlichkeitseigenschaften auf die Niveauunterschiede bei Einzel- vs. Gruppenvorgabe untersucht.

Nun soll der Frage nachgegangen werden, ob sich die Ergebnisse von 1998 - mehr als 10 Jahre später - mit einem anderen Test, anderen Testpersonen und anderen Testleitern replizieren lassen.

Im Vergleich zum Versuchsdesign von 1998, wurden aktuell teilweise Änderungen vorgenommen, andere Vorgehensweisen wurden wiederum zwecks der Vergleichbarkeit übernommen.

Getestet wurde mit einem Subtest des WIT-2 (Wilde Intelligenz Test) von Kersting, Althoff und Jäger (2008). Der Subtest „Abwicklungen“ testet räumliches und schlussfolgerndes Denken. Es wurde gezielt dieser Subtest aus der Testbatterie ausgewählt, da es zum Einen ein Test sein sollte, welcher die Fähigkeit zum schlussfolgernden Denken erfasst und zum Anderen ein aktuelles Verfahren darstellen soll, welches laut Autoren *explizit* sowohl in einer Gruppentestsituation als auch in einer Einzeltestung einsetzbar ist. Als Testpersonen fungierten Kinder aus 3 vergleichbaren Wiener Schulen (um den Schuleinfluss konstant zu halten) im Alter zwischen 15 und 19 Jahren. Um den Einfluss der Versuchsleiter gezielt zu untersuchen, wie es schon in der Studie von Reindl (1998) gehandhabt wurde, wurden die Tests von vier ausgewählten und in ihrer Einstellung manipulierten Testleiterinnen durchgeführt. Da sich in der Literatur neben dem Einfluss der Testleiter auf die Testpersonen auch noch weitere mögliche Einflüsse finden ließen, die in den beiden unterschiedlichen Testvorgabesituationen unterschiedlich stark auftreten können (z.B. Störeinflüsse, Abschreiben, Anonymität in der Gruppe usw.), wurde zusätzlich ein Kurzfragebogen vorgegeben, in welchem Einflussfaktoren, die die Testsituation betreffen, von den Probanden bewertet werden sollen. Dieser soll später dazu dienen, mögliche Leistungsunterschiede durch die Befindlichkeit der Testpersonen zu erklären.

Die vorliegende Diplomarbeit soll durch wiederholte Untersuchung des Themas einen Beitrag zur Absicherung der bis dato getroffenen Annahmen liefern.

2. Theoretischer Teil

In Kapitel 2 dieser Diplomarbeit wird der theoretische Hintergrund der nachfolgenden Studie beleuchtet. Neben dem kurzen Überblick über die Psychologische Diagnostik, in welchem eine Systematik psychologischer Verfahren vorgestellt werden soll – um zu erläutern in welchem Rahmen die Fragestellung und die Untersuchungsplanung eingebettet ist –, widmet sich das Kapitel der genaueren Beschreibung des Begriffes *Intelligenztestung*. Da die Ergebnisse der folgenden Studie Einfluss auf die zukünftige Testnormierung haben könnte, wird auch diese in einem Kapitel (2.1.2) Erwähnung finden. Schon in den anschließenden Abschnitten wird von Einzel- vs. Gruppenvorgabe die Rede sein. Was darunter verstanden wird, wodurch sich die zwei Vorgabemodi unterscheiden und was ihre Vor- und Nachteile sind, findet sich in Kapitel 2.2 wieder. Des Weiteren soll in Kapitel 2.3 erörtert werden, worauf mögliche Unterschiede in den Testleistungen im Vergleich von Einzel- und Gruppenvorgabe rückführbar sind. Da es sich bei dieser Diplomarbeit um eine Wiederholungsstudie von Reindl (1998) handelt, wird in Kapitel 2.4 die Ursprungsarbeit vorgestellt werden.

2.1 Überblick über die Psychologische Diagnostik

Der Begriff Diagnostik umfasst im Fachbereich Psychologie einige Konstrukte, Definitionen und Beschreibungen. Es soll im Folgenden ein kurzer Überblick über die für die nachfolgende Untersuchung relevanten Bereiche der Leistungsdiagnostik gegeben werden, um nachvollziehbar zu machen was hier unter „Leistungstests“ und der daraus resultierenden Ergebnissen verstanden wird.

Nach Kubinger (2009) bezieht psychologisches Diagnostizieren folgendes mit ein:

- Klärung der Fragestellung
- Auswahl der einzusetzenden Verfahren
- Anwendung und Auswertung der Verfahren
- Interpretation und Gutachtenerstellung
- Festsetzen der Intervention

Zur Beantwortung der zu Beginn geklärten Fragestellung, müssen geeignete Erhebungsverfahren ausgewählt werden. Traditionell unterscheidet man in der psychologischen Diagnostik im Bereich der Testung (Beobachten und Befragen wären neben der Testung weitere zu erwähnende Erhebungsmethoden) *Leistungstests* von *Persönlichkeitstests* (bzw. *-fragebögen*). Kapitel 2.1.1 wird sich dem Thema der Leistungstests genauer widmen, da die Testung der hier nachgegangenen Hypothese mit Selbigen erfolgt. Der bei dieser Diplomarbeit relevante Unterschied der beiden Verfahren liegt darin, dass in der Leistungstestung Anstrengung für gute Leistungen erfolgen muss. In der Persönlichkeitsdiagnostik hingegen, sollen typische Verhaltensweisen beobachtet/erfasst werden. Die Störanfälligkeit von Leistungsergebnissen ist somit höher als die Ergebnisse eines Persönlichkeitstests/-fragebogens. Aus diesem Grund ist gerade bei psychologischen Leistungstests auf die Umgebungsfaktoren zu achten, wozu eben auch der Vorgabemodus (Einzel- oder Gruppentestung) zählt.

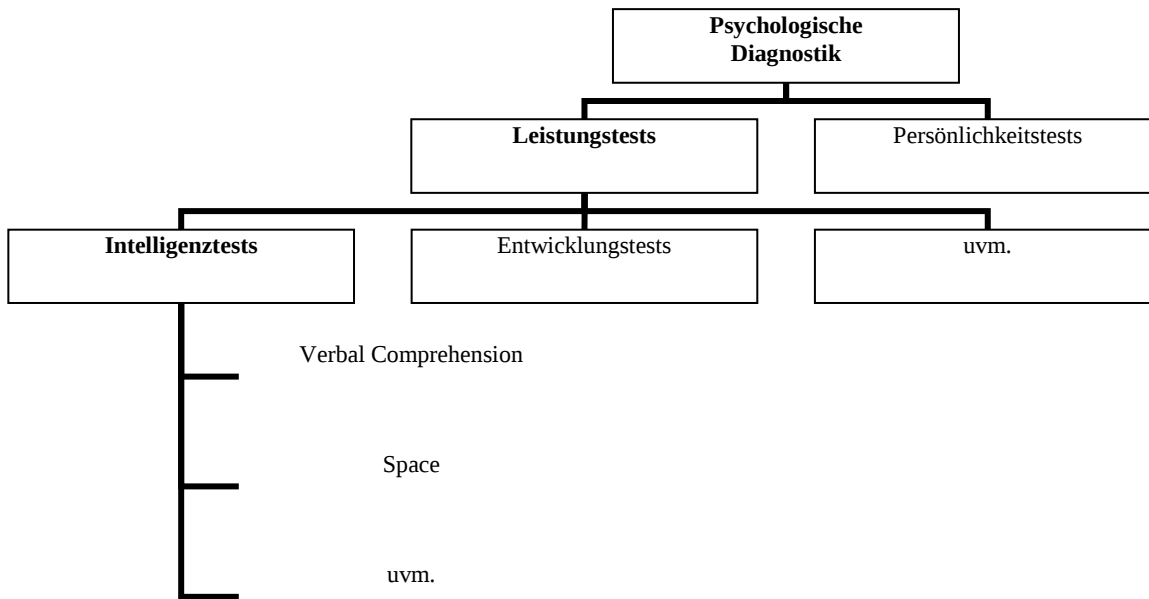


Abbildung 1 Überblick über psychologisch-diagnostische Verfahren

2.1.1 Intelligenztests und Leistungsdimensionen

Der Bereich der Leistungstests (neben den Persönlichkeitsfragebögen) umfasst *allgemeine Leistungstests*, *Intelligenztests*, *Entwicklungstests*, *Schultests* und *spezielle Funktionsprüfungstests (Eignungstests)*.

Im Weiteren wird der Fokus im Speziellen auf den *Intelligenztests* liegen, welcher durch die Durchführbarkeit der Studie bedingt ist, welche im Abschnitt 3.2 (Untersuchungsplan) noch näher erläutert wird. Aus diesem Grund wird hier nur auf den Bereich der Intelligenztests genauer eingegangen.

Intelligenztests können als Intelligenztestbatterien (Zusammenstellung verschiedener Untertests) oder als spezielle Leistungstests (Messung einer einzigen Dimension) vorgegeben werden und sollen die kognitiven Fähigkeiten eines Menschen erfassen.

Eine anschauliche Systematik der Leistungsdimensionen bei Intelligenztests (eigentlich auch Intelligenzfaktoren, da sich die Gesamtintelligenz aus einzelnen Komponenten zusammensetzt) baut im Wesentlichen auf den Intelligenzfaktoren von Thurstone auf (nach Kubinger, 2009):

- Verbal Comprehension (Sprachverständnis)
- Space (Raumvorstellung)
- Memory (Merkfähigkeit)

- Perceptual Speed (Verarbeitungsgeschwindigkeit)
- Reasoning (schlussfolgerndes Denken)
- Lernfähigkeit
- Aufmerksamkeit und Konzentration
- Technisches Verständnis

Jedoch gibt es eine Vielzahl von Intelligenztheorien mit unterschiedlichen Auffassungen von Intelligenz und derer Dimensionen. Als Exkurs soll die obige Aufstellung jedoch genügen.

Ein als von vielen Autoren wichtiges Konstrukt wird *reasoning* angesehen. Kubinger und Jäger (2003) meinen: “Als eine der wichtigsten solcher allgemeinen geistigen Fähigkeiten wird seit langem die Fähigkeit zum schlussfolgernden Denken betrachtet, die daher auch in den meisten Intelligenztests besonders beachtet wird.“ Aus dem Grund der Ökonomie konnte in der vorliegenden Untersuchung keine Intelligenztestbatterie verwendet werden, sogar die Vorgabe von mehr als einem speziellen Leistungstest wäre nicht möglich gewesen. Aus den oben angeführten Gründen fiel bei der Wahl des Erhebungsinstrumentes die Entscheidung auf die Testung des Konstruktes *reasoning* (siehe Kapitel 3.4.1 Wilde-Intelligenz-Test 2).

2.1.2 Normierung psychologischer Tests

Die Normierung von Testwerten dient der Vergleichbarkeit von Personen und ermöglicht die Interpretation von Testleistungen (oder auch von Verhaltensweisen). Sie liefert sozusagen ein Bezugssystem. Die Testwerte werden in Bezug zu einer *Standardisierungsstichprobe* gesetzt. Am Beispiel eines 15-jährigen männlichen AHS-Schülers ausgedrückt: Wie sind die erbrachten Leistungen im Test XY im Vergleich zu gleichaltrigen männlichen AHS-Schülern? Wie viel Prozent vergleichbarer Kinder erzielten schlechtere Leistungen?

Die Normierung ist eines der 10 Gütekriterien psychometrischer Tests (vgl. Kubinger, 2009), welche die Eigenschaften psychologischer Verfahren beschreiben.

Wichtig bei einer Norm ist nach den „Standards for Educational and Psychological Testing“ der APA (nach Kubinger und Jäger 2003):

- Aktualität der Normen
- Geltungsbereich der Normen
- Grad der Repräsentativität der Normierungsstichproben

- Umfang der Normierungsstichprobe
- Messgenauigkeit zur Bestimmung des Messfehlers

Um Normen aktuell zu halten, müssen immer wieder neue Normierungsstichproben mit einbezogen werden oder alte gar gänzlich verworfen werden. Die DIN 33430 (DIN Deutsches Institut für Normung e. V., 2002, S. 7) schreibt vor, Normwerte spätestens alle 8 Jahre prüfen zu lassen. Außer die Norm verliert schon vorher ihren Geltungsbereich oder die Repräsentativität. Und genau dies könnte der Fall sein, sollten sich Leistungsunterschiede zwischen den Vorgabemodi eines Intelligenztests finden lassen. In gängigen Leistungstests lassen sich keine getrennten Normen für die Vorgabe im Einzel- oder Gruppensetting finden, obwohl von den Testautoren oft eine Verwendung in beiden Situationen anbieten. Ob das gerechtfertigt ist, soll die vorliegende empirische Studie klären.

Im Folgenden sollen zwei Tests bezüglich ihrer Verwendungsmöglichkeit und Normierung beispielweise angeführt werden.

Im viel verwendeten *FAIR* (Frankfurter Aufmerksamkeits-Inventar; Moosbrugger und Oehlschlägel, 1996) wird von den Autoren ausdrücklich darauf hingewiesen, dass das Verfahren als Einzel- oder als Gruppentest durchgeführt werden kann. Die Normierungstabellen beinhalten jedoch lediglich Normen welche nach Alter getrennt angegeben sind und eine Gesamtnorm.

Der *Test d2* von Brickenkamp (1994) wird ebenfalls als Einzel- und Gruppentest (siehe <http://www.testzentrale.de/programm/d2.html>) beschrieben. Die Normentabellen unterscheiden sich wiederum nur durch verschiedene Altersgruppen.

Andere Testautoren beschreiben ihr Verfahren wiederum als „Gruppentest“, ohne die Möglichkeit einer Einzelvorgabe zu erwähnen. Kann und darf man dieses Verfahren in diesem Falle auch nur als Gruppentest anwenden und nicht in der Einzelvorgabe? Es kann jedoch angenommen werden, dass dies von den Testautoren meist nicht in dieser Weise verstanden wird. Gängig ist eher die Annahme wie sie Kubinger (2009) tätigt, dass bei einem Gruppenverfahren *natürlich* auch die Einzelvorgabe möglich ist. Obwohl keine aussagekräftigen Voruntersuchungen existieren, geht eine Vielzahl der Testautoren davon aus, dass sich die Testleistungen der zwei Vorgabemodi nicht relevant unterscheiden und bietet somit keine getrennten Vergleichsstichproben an. Sollten die oben erwähnten Unterschiede jedoch empirisch zu finden sein, würde dies bedeuten, dass eine neuerliche Normierung

notwendig wäre. Es müsste dann möglich sein, die Leistungen des getesteten Testpersonen (als Beispiel in einer Gruppe getestet) mit den Leistungen der passenden Vergleichsstichprobe (ebenfalls in einer Gruppe getestet) zu vergleichen.

2.1.3 Die Relevanz der Leistungsdiagnostik im speziellen für Kinder

In diesem Absatz soll die Relevanz einer korrekten Leistungsdiagnostik im speziellen für Kinder verdeutlicht werden, da hier die Auswirkungen „schlechter“ psychologischer Tests gravierend wären. Aufgrund neuer Forschungsergebnisse und neuer Erkenntnisse sollten Tests immer weiter angepasst und verändert werden, so dass die Gütekriterien eingehalten werden können.

So tangieren psychologische Tests Kinder beispielsweise bei Aufnahmen in Schulen, etwa im Rahmen von Schuleignungsüberprüfungen oder Schulreifeuntersuchungen. Auch bei Teilleistungsstörungen (z. B.: Lese-Rechtschreibstörung) ist eine detaillierte und vor allem korrekt durchgeführte und somit aussagekräftige Diagnostik notwendig, da diese dem Kind Zugang zu gezielter Förderung ermöglicht. Es seien auch noch die Bereiche Erziehungsprobleme, Verhaltensstörungen und Berufswahl zu erwähnen.

Auch bei Hochbegabungen hat die Leistungsdiagnostik eine wichtige Funktion. Diese Kinder leiden meist an Unterforderung, was wiederum oft zu Verhaltensauffälligkeiten und Versagen in der Schule führt. All diese Bereiche beeinflussen ein gezieltes Weiterkommen in der schulischen Laufbahn, was wiederum Auswirkung auf die späteren Berufschancen, aber auch die Lebenszufriedenheit im Allgemeinen haben kann.

2.2 Einzel- vs. Gruppentestung

Psychologische Tests lassen sich in verschiedenster Weise gestalten. In welchen formalen Rahmen ein Test eingebettet wird lässt sich durch folgende Gestaltungsweisen zeigen:

- Einzel- vs. Gruppentest
- Papier-Bleistift - vs. Computerverfahren
- Power- vs. Speed-and-Power-Test
- Freies Antwortformat vs. Multiple Choice

Jede Gestaltungsweise hat seine Vor- und Nachteile. Je nach Testaufbau, getesteter Leistungsdimension und Verwendungszweck sollte die Entscheidung auf die eine oder andere Variante fallen. Schon in der Testentwicklung und –gestaltung müssen diverse Überlegungen angestellt werden. Um zu entscheiden wie das Verfahren aufgebaut werden soll oder aufgebaut sein muss um es für die benötigten Zwecke verwenden zu können braucht man detailliertes Wissen über pro und contra.

Da in dieser Studie vor allem die Unterscheidung zwischen Einzel- und Gruppentest wichtig ist, soll in Kapitel 2.2.1 und 2.2.2 auf die Vor- und Nachteile der zwei Varianten eingegangen werden. Im Folgenden werden die Begriffe *Einzelvorgabe* und *Gruppenvorgabe* erläutert.

Prinzipiell versteht man in der psychologischen Leistungsdiagnostik unter *Einzeltestung* das Testen nur einer Testperson. Die Vorgabe und Instruktion erfolgt durch einen Testleiter. In einer *Gruppentestung* werden von meist einem Testleiter gleichzeitig mehrere Personen in einem Raum getestet, ähnlich einer Schularbeitssituation. Wie schon erkennbar ist, spielt der Testleiter eine wesentliche Rolle in einer psychologischen Testsituation. In der einen Bedingung kommt auf eine Testperson ein Testleiter. In der anderen beschäftigt sich ein Testleiter mit meist um die 10 Testpersonen gleichzeitig. In Kapitel 2.3.1 wird der Einfluss des Testleiters genauer erläutert. Vorab ist jedoch hervor zu heben, dass die Einwirkung des Testleiters in den zwei Testbedingungen eine deutlich unterschiedliche ist. Ob dieser Unterschied jedoch auch Auswirkungen auf die Testleistungen der Testpersonen hat, kann noch nicht ausreichend beantwortet werden und bedarf weiterer Untersuchungen. Eine weitere Begründung für mögliche Leistungsdifferenzen zwischen Einzel- und Gruppenvorgabe findet sich in Kapitel 2.3.2.

2.2.1 Vorteile der Einzeltestung

a) Einsatz besonderer Testmaterialien

Die Einzeltestung ermöglicht den Einsatz besonderer Testmaterialien. Zum Beispiel verwendet das *Adaptive Intelligenz Diagnostikum -Version 2.1 (AID 2;* Kubinger & Wurst, 2000) verschiedenartiger Materialien, wie Würfel, Karten, Bildtafeln und vieles mehr, welche vom Testleiter je nach Untertest dem Testpersonen vorgelegt oder gezeigt werden. Diese Vorgabe ist jedoch nur in der Individualtestung möglich und nicht wenn der Testleiter mehrere Testpersonen gleichzeitig zu testen hat.

b) besseres Eingehen auf die Testperson/en

Der Testleiter hat im Individualverfahren die Möglichkeit, speziell auf die Testperson einzugehen. So kann die Instruktion angemessen erfolgen. Selbst bei standardisierten Testvorgaben und Instruktionen, bei denen die Handlungsschritte und die genaue sprachliche Anweisung vorgegeben ist und keinen Spielraum zu lassen, ist es möglich, Tonfall, Geschwindigkeit und nonverbale Kommunikation zu variieren. Auch wird bei Gruppenvorgaben kritisiert (siehe Kubinger & Jäger, 2003), dass bestimmte Personengruppen benachteiligt werden. So sind meist gute Lesefertigkeiten von Nöten. Bei der Einzeltestung haben Testpersonen mit diesem Handikap keinen Nachteil.

c) weniger Störeinflüsse

Störung wirkt sich leistungsmindernd aus, da nach Schneider, Wegge und Konradt (1993) Störeinflüsse Kapazitäten verbrauchen, welche dann nicht mehr für die Bearbeitung des Tests bereitstehen. Auch in der Einzeltestung kann es Störungen von außen geben (Straßenlärm, jemand betritt den Raum, Lichtverhältnisse etc.), jedoch in viel geringerem Maße als es in der Gruppentestung der Fall ist.

d) keine Möglichkeit ab zu Schreiben

Der Testleiter wird auch bei Gruppenverfahren meist dazu angehalten, ein Abschreiben der Testpersonen untereinander zu verhindern. Oft bieten jedoch die Räumlichkeiten oder andere Umstände nicht die Möglichkeit ein Abschreiben zu verhindern. Für eine Gruppe von 10 Personen ist meist ein Testleiter/eine Testleiterin ausreichend. Bei mehreren Testpersonen sollten entsprechend mehr eingesetzt werden. Bei so einem oder einem ähnlichen Verhältnis von Testleiter zu Testpersonen ist es allerdings nur schwer möglich alle Testpersonen im Blick zu haben. Gerade wenn die Vorgabe am Computer stattfindet, sind die räumlichen Bedingungen im Voraus festgelegt und können meist nicht auf eine psychologische Testsituation angepasst werden. Im Einzelsetting kann es zu keinen Verfälschungen durch Abschreiben kommen.

e) keine soziale Hemmung

Der Begriff *soziale Leistungsaktivierung* ist in der Sozialpsychologie schon lange bekannt. Die Frage, ob die Anwesenheit anderer die Leistung steigert oder behindert ist allerdings nicht so leicht beantwortbar. Gesagt werden kann, dass die Anwesenheit anderer zu einer gesteigerten Antriebsniveau führt, welches sich positiv oder negativ auswirken kann. Die Lernpsychologie zeigt uns, dass in Zuständen erhöhter Aktiviertheit die dominante Handlungsalternative ausgeführt wird (meist stehen mehrere Alternativen zur Verfügung). Daraus leitet Zajonc (1965, nach Herkner 2003) folgende zwei Hypothesen ab: (1) Lernprozesse werden durch die Anwesenheit anderer Personen gestört. (2) Gut gelernte und geübte Tätigkeiten jedoch werden dadurch gefördert. Da eine psychologische Testsituation mit meist unbekannten Tests eine neue Situation ist, kann davon ausgegangen werden, dass hier die Anwesenheit anderer Personen die Leistung minimiert und dass eine psychologische Testung im Einzelsetting in diesem Fall bessere Ergebnisse provoziert.

2.2.2 Vorteile der Gruppentestung

a) größere Testleiterunabhängigkeit

Der in Kapitel 2.2.1 erwähnte Vorteil der Einzeltestung, nämlich die Möglichkeit des individuellen Eingehens auf die Testperson, ist unter dem Gesichtspunkt des Gütekriteriums der Objektivität, genauer der Testleiterunabhängigkeit, ein grober Nachteil. Testleiterunabhängigkeit ist dann gegeben, wenn ein Testergebnis unabhängig von zufälligen oder systematischen Verhaltensvariationen des Testleiters ist. Um diesen Effekt aus zu schalten wird eine maximale Standardisierung der Testsituation angestrebt. Dies ist bei einem Gruppentest auch leicht möglich, da alle Gruppenmitglieder dieselbe Instruktion erhalten. Gerade Einflüsse wie der Testleitererwartungseffekt (siehe Kapitel 2.3.1) würden sich gerade in der Einzeltestung besonders manifestieren. Hier hat der Testleiter mehr Möglichkeit sich eine Einstellung gegenüber der Testperson aus zu bilden. In einer Gruppentestung hingegen ist es für den Testleiter nur schwer möglich sich von jedem Einzelnen/jeder Einzelnen ein Bild zu machen.

Zu erwähnen ist allerdings, dass von manchen Testautoren der Einfluss des Testleiters und somit die Interaktion zwischen Testperson und Testleiter sogar gewünscht wird. Dies stellt allerdings eine Ausnahme dar und soll hier nicht näher beleuchtet werden, da diesem Thema genauere Aufmerksamkeit geschenkt werden sollte. Im Rahmen dieser Studie ist dies jedoch nicht Möglich.

b) höhere Ökonomie

Auf der Hand liegt die höhere Ökonomie von Gruppenverfahren im Vergleich zu Einzelverfahren. Erstens benötigt man für die Testung mehrerer Personen gleichzeitig lediglich einen Testleiter. Zweitens kann dieser alle Testpersonen gleichzeitig instruieren, was wiederum Zeit spart. Der Zeitaufwand reduziert sich so auf ein Minimum.

c) Anonymität gegenüber dem Testleiter

Die Gruppe bietet Schutz. Ist man Teil einer Gruppe, fällt man als Einzelner nicht auf. Laut Kubinger (2009) fühlen sich genau aus diesem Grund manche Personen in einer Gruppentestung wohler. Das Gefühl des „Beobachtetseins“ tritt nicht so stark in den

Vordergrund. Wie schon im Kapitel 2.2.1 c) beschrieben wurde, binden Störeinflüsse kognitive Ressourcen, welche dann nicht mehr für die Bearbeitung des Tests zur Verfügung stehen. Auch ein Unwohlsein durch das zu nahe Verhältnis mit dem Testleiter in der Testsituation bei Einzeltests, welches auch als Störeinfluss gesehen werden kann, kann solche kognitiven Ressourcen „belegen“.

d) leichter Zugang zu Normierungsdaten

Durch die ökonomischere Möglichkeit, mit der Gruppentestung eine größere Menge an Personen zu testen, ist anzunehmen, dass die Normierungsstichprobe bei Gruppentests höher sein wird als bei Tests, welche für die alleinige Verwendung als Einzeltest konzipiert sind.

2.2.3 Gründe für mögliche Unterschiede im Leitungsniveau zwischen Einzel- und Gruppenvorgabe

Schon zu Beginn soll der Frage nachgegangen werden, welche Erklärungen für möglichen Unterschiede im Leistungsniveau zwischen Einzel- und Gruppenvorgabe gefunden werden können, um die hier angestrebte Untersuchung im Voraus korrekt planen zu können. Aufgrund dieser Recherche wird in der folgenden Untersuchung neben der gängigen Intelligenztestung noch ein Befindensfragebogen von den Testpersonen am Ende der Testung ausgefüllt werden (Kapitel 3.2.3) und die Untersuchung wird experimentellen Charakter haben, um mögliche Unterschiede auf Testleitereffekte (vor allem den Versuchsleitererwartungseffekt) zurückführen zu können.

Wie schon in Kapitel 2.2 (Einzel- vs. Gruppentestung) beschrieben, provoziert alleine die Versuchsbedingung unterschiedliche Vor- und Nachteile. So etwa könnten Kinder, welche im Klassenverbund getestet wurden, bessere Leistungen erzielen, da sie gegenseitig abschreiben könnten. Diese Möglichkeit bietet sich Kindern, welche alleine mit einem Testleiter im Raum sind und somit unter ständiger Beobachtung stehen, nicht. Umgekehrt könnte sich, wie oben erwähnt, eine Testperson vielleicht besser konzentrieren, da Störeinflüsse wie andere - Lärm verursachende - Personen fehlen.

Sollten sich in der vorliegenden Untersuchung Unterschiede im Leistungsniveau zeigen, wirft dies die Frage auf, welche Gründe dafür gefunden werden können, um in zukünftigen

diagnostischen Prozessen diese Störeffekte gezielt aus zu schalten und die Unterschiede in die Normierung von Testwerten mit auf zu nehmen.

Wie oben schon angedeutet, lassen sich Einflüsse des Testleiters und des Befindens der Testpersonen/Probandinnen während der Testung auf die Testleistungen vermuten, welche in den nachfolgenden zwei Kapiteln beschrieben werden.

Exkurs:

Anzumerken ist des Weiteren, dass auch die Persönlichkeit der Testperson eine Rolle spielen könnte. Kiesel (2009) ging der Frage nach, auf welche Persönlichkeitseigenschaften Niveauunterschiede bei Einzel- und Gruppenvorgaben rückführbar sind (Anm.: sofern welche existieren). Welche Persönlichkeitseigenschaften begünstigen das bessere Abschneiden in der einen oder anderen Testsituation? Die Ergebnisse deuten darauf hin, dass Unterschiede existieren. Sich als *verträglich* beschreibende Personen schneiden beispielsweise bei Einzeltests besser ab. Bei figuralen Aufgaben konnten, sich als *offen* beschreibende, bessere Ergebnisse bei Leistungstests erzielen. Ein Unterschied bei weiteren Persönlichkeitseigenschaften konnte nicht gefunden werden.

Da die Testpersonen aufgrund ihrer verschiedenen Persönlichkeiten, in Bezug auf Einzel- vs. Gruppentestung, andere Ergebnisse erbringen, ist anzunehmen, dass auch ihr subjektives Befinden anders sein wird. Zum Beispiel: Eine wenig verträgliche Person wird sich im Gruppensetting auch weniger wohl fühlen. Aus diesem Grund fließt der Faktor *Persönlichkeit* in der vorliegenden Studie bei der Datenerhebung im Punkt *Befindensfragebogen* (Kapitel 3.4.2) mit ein und wird nicht außer Acht gelassen.

a) Der Testleiter als Störvariable

Es darf nicht vergessen werden, dass auch Testleiter Personen mit ganz eigenen Erwartungen, Gefühlen und Motiven sind. Deshalb kann es vorkommen, dass es dem Testleiter/der Testleiterin schwer fällt gänzlich in die vorgegebene und erwartete Rolle des unvoreingenommenen Testleiters/Testleiterin zu schlüpfen. Oft werden in Testmanualen ganz genaue Anweisungen und Anleitungen gegeben wie sich der Testleiter/die Testleiterin in den verschiedenen Situationen zu verhalten hat, um die „Gleichheit“ des TestleiterInnenverhaltens zu gewährleisten.

Seitens des Versuchsleiters unterscheidet Bredenkamp (1969, nach Gniech 1976) bezogen auf psychologische Experimente folgende, vom Versuchsleiter verursachte, *nicht beabsichtigte* Einflüsse:

- Experimentator-Bias
 - Einflüsse welche durch *Erwartungen* des Versuchsleiters geprägt sind
- Experimentator-Effekte
 - *biosoziale und psychosoziale Aspekte* des Versuchsleiters als Person

Da diese Einflüsse aber nicht nur für das Experiment gelten, sondern für jede wissenschaftliche Untersuchung und auch für die Interaktion im Rahmen einer psychologischen Testung, werden wir hier von *Testleiter-Bias* (oder auch *Testleitererwartungseffekt*) und *Testleiter-Effekten* sprechen, da dies in diesem Rahmen der angemessenere Terminus ist.

Zu den biosozialen und psychosozialen Aspekten (Testleitereffekte) zählen zum Beispiel das Geschlecht, das Auftreten, das Verhalten oder die Ausdrucksweise des Testleiters.

Der Versuchsleitererwartungseffekt entsteht dadurch, dass der Versuchsleiter in den meisten Fällen eine eigene Theorie bezüglich der Testperson oder der Untersuchungsergebnisse hat und diese durch nicht beabsichtigte Verhaltensweisen den Testpersonen nonverbal mitteilt. Auch in Form einer Wahrnehmungsverzerrung der Testperson gegenüber (z.B. bei der Verhaltensbeobachtung) kann sich der Versuchsleitererwartungseffekt manifestieren.

In Kapitel 2.2.2 (a) wurde schon erwähnt, dass in einer Einzeltestung der Testleitererwartungseffekt besonders zu tragen kommen würde, da hier eine intensivere Interaktion zwischen Testleiter und Testperson stattfindet und die Einstellungsbildung in einem höheren Grad möglich ist als in der Gruppentestung. Eine weitere Einstellung des Testleiters/der Testleiterin, kann sich auf das Setting in welchem getestet wird beziehen. Es soll ein fiktives Beispiel skizziert werden: Ein Testleiter testet ein Kind in der häuslichen Umgebung und ist fest der Ansicht, dass Kinder im eigenen Wohnumfeld sich nicht so gut konzentrieren können und deshalb schlechtere Leistungen bringen als Kinder welche in einem Testraum getestet werden. Der Testleiter/die Testleiterin erwartet also je nach Testperson und Testsituation ein bestimmtes Ergebnis. Durch diese Erwartung kann es passieren, dass der Testleiter/die Testleiterin die Testperson unbeabsichtigt und durch nonverbales Verhalten beeinflusst. Schon Rosenthal (1966; nach Herkner, 2003) hat den Einfluss der Erwartung von Versuchsleitern in Lernexperimenten mit Ratten bestätigt. So wurde einem Teil der Testleiter

erzählt sie hätten es mit besonders schlaun Ratten zu tun, dem anderen Teil es wären Ratten einer außerordentlich dummen Population. Das Ergebnis entsprach den Erwartungen der Versuchsleiter. Die angeblich intelligenten Ratten schnitten bei den Lernaufgaben deutlich besser ab, als die Ratten der Versuchsleiter, die dachten ihre Ratten seien wenig intelligent. Diese Ergebnisse lassen sich dadurch erklären, dass Versuchs- oder Testleiter sich mit den Versuchs- oder Testpersonen in einer anderen Weise beschäftigen. Im Falle des Experimentes mit den Ratten, waren die Versuchsleiter den „intelligenten“ Ratten viel mehr zugewandt, hatten mehr Kontakt zu ihnen, die Stimmlage war eine andere und waren geduldiger. Es ist anzunehmen, dass auch Testleiter in psychologischen Testsituationen ähnliches, nonverbales Verhalten zeigen.

Durch Standardisierung der Testinstruktion und der Testvorgabe, können verzerrende Effekte, welche durch den Testleiter hervorgerufen werden minimiert werden. Eine stark vorgegebene Instruktion beispielsweise könnte Abhilfe schaffen. Es bliebe dem Testleiter nur noch wenig Spielraum seine individuellen Ausführungen ein zu bringen. Ob ein Testleiter der Testperson jedoch sympathisch ist, ob ihm seine Stimmlage oder sein äußeres Erscheinungsbild zusagt oder nicht, kann auch durch eine Standardisierung des Testleiterverhaltens nicht beeinflusst werden. Eine Möglichkeit Testleitereffekte aus zu schalten ist die Anwendung von computer-gestützter Diagnostik. Hier fällt der Einfluss des Testleiters fast gänzlich aus. Dazu muss das PC-Verfahren klarerweise eine so verständliche und ausführliche Testanweisung enthalten, dass die Testperson tatsächlich ohne weitere Erklärungen durch den Testleiter/die Testleiterin den Ablauf ausreichend verstehen kann.

b) Die Auswirkung des Testpersonenbefindens

Nicht jeder fühlt sich in den gleichen Situationen wohl. Die einen arbeiten nicht gerne in Gruppen, die anderen mögen es nicht alleine zu sein. Auch wirken sich Störeinflüsse, wie etwa das Husten einer weiteren Testperson im Raum, bei jeder Person anders aus. Einer fühlt sich dermaßen gestört durch den Lärm, den andere verursachen, dass er sich auf die eigentliche Testung kaum noch konzentrieren kann, der Andere lässt sich auch dadurch nicht aus dem Konzept bringen. Die in Kapitel 2.2 beschriebenen Unterschiede in den Ausgangsbedingungen zwischen Einzel- und Gruppentestung schlagen sich vor allem im Befinden der Testpersonen nieder. Ob das Wohlbefinden in der Testsituation sich nun auch

auf die Testleistungen bei einer Intelligenztestaufgabe niederschlägt, soll die vorliegende Diplomarbeit klären. Sollten sich Leistungsunterschieden zwischen der Einzel- und der Gruppentestung finden lassen, könnten die unterschiedlichen Rahmenbedingungen der beiden Vorgabemodi, neben den Testleitereffekten, ein Erklärungsansatz sein.

2.3 Diplomarbeit von Reindl 1996: *Einzel- vs. Gruppenvorgabe: Niveauunterschiede im Leistungsbereich*

Da die Studie von Reindl (1998) als Grundlage der hier vorliegenden Diplomarbeit dient, soll sie kurz umrissen werden:

In der Studie wurde eine eigens erstellte Kurzform des Berliner-Intelligenzstrukturtest (BIS-T) (Jäger et al., 1997) vorgegeben. Es wurde Bearbeitungsgeschwindigkeit, Merkfähigkeit, Einfallsreichtum und Verarbeitungskapazität erfasst. Als Testpersonen fungierten 241 Schüler aus 3 Wiener Schulen im Alter zwischen 14 und 18 Jahren. Um den Testleiterbias nicht zu vernachlässigen, wurden 4 ausgewählte Testleiter in ihrer Einstellung bezüglich der erwarteten Testleistungen manipuliert.

Studienaufbau

Wie oben erwähnt, wurde mit einer eigens erstellten Kurzform des BIS-T getestet, welcher nur den Bereich der Verarbeitungskapazität mit ein bezog. Des Weiteren wurde vor der Bearbeitung des Testes ein Fragebogen vorgegeben, welcher zur Erfassung des Anregungsniveaus (Erhebung von Variablen wie Geschwisteranzahl, soziales Umfeld, Mitbestimmung und Freizeitgestaltung) und der sozioökonomischen Daten diente. Die Erhebung des Anregungsniveaus wurde durchgeführt, um grundlegende Unterschiede zwischen den getesteten Gruppen ausschließen zu können. Die im Fragebogen enthaltenen Dimensionen lauten *Hilfe und Förderung in der Schule*, *Soziale Kontakte* und *Selbstständigkeit und Selbstbestimmung*. Die Manipulation der Einstellung der 4 Testleiter erfolgte durch Meinungsbildung in einem Einzelgespräch. Reindl (1998) schreibt: „Zwei Testleiter, Nr.1 und Nr.4, erhielten die Informationen, dass Gruppenvorgaben sinnvoller sind, die beiden anderen Testleiter, Nr. 2 und Nr. 3, sollten glauben, dass Einzelvorgaben angebrachter wären.“ Es sollte also ein Teil so manipuliert werden, dass sie denken, die Testpersonen welche im Einzelsetting getestet werden, schneiden bei dem Leistungstest besser ab und der andere Teil soll die Einstellung vertreten, dass die Leistungen im Gruppentest besser ausfallen werden.

Ergebnisse

Um inhaltlich bedeutsame Unterschiede handelt es sich erst ab 5 T-Werten Unterschied zwischen den getesteten Gruppen. Dieser Wert wurde aufgrund des Messfehlers des Testes und inhaltlichen Überlegungen a priori festgelegt. Aus diesem Grund konnte statistisch festgestellt werden, dass alle 8 Nullhypothesen beibehalten werden müssen und sich keine Unterschiede zwischen den Gruppen finden ließen. Einige Ergebnisse fielen zwar signifikant aus, jedoch war der gefundene Unterschied zu klein, als dass sie inhaltlich relevant gewesen wären (Unterschiede des Leistungsniveaus weniger als die oben genannten 5 T-Werte).

Somit konnte die implizite Annahme, man könne den BIS-T sowohl als Einzel- sowie auch als Gruppenverfahren einsetzen, bestätigt werden. Des Weiteren wurden keine Testleitereffekte gefunden. Somit führt die Diplomarbeit von Reindl zu dem Schluss, dass es keine Niveauunterschiede im Leistungsbereich in Abhängigkeit vom Vorgabemodus gibt und somit für Gruppentestungen konzipierte Verfahren auch in der Einzeltestung verwendet werden können. Es kann erwartet werden, dass das Vorgabemodus nicht unterschiedliche Testergebnisse provoziert.

3. Empirischer Teil

3.1 Ziel der Untersuchung (Hypothesen)

In der vorliegenden Studie soll, wie auch schon Reindl (1998), neben dem grundsätzlichen Unterschied im Leistungsniveau zwischen Einzel- und Gruppenvorgabe, auf den Testleitererwartungseffekt eingegangen werden. In Ergänzung zu der Vorgängerstudie soll nun auch noch das Befinden der Testpersonen während der Testung erhoben werden um für mögliche Unterschiede Erklärungsansätze bieten zu können. Wie in Kapitel 2.2.3 beschrieben, könnte der Testleitererwartungseffekt sowie das unterschiedliche Befinden der Testpersonen in den verschiedenen Vorgabemodi eine Erklärung für eventuelle Niveauunterschiede im Leistungsbereich verantwortlich sein und soll aus diesem Grunde mit erhoben werden.

Hypothesen:

Haupthypothese

- H1₀: Es bestehen keine Niveauunterschiede im Leistungsbereich in Abhängigkeit vom Vorgabemodus (Einzel- vs. Gruppenvorgabe).
- H1₁: Es bestehen Leistungsunterschiede im Leistungsbereich in Abhängigkeit vom Vorgabemodus (Einzel- vs. Gruppenvorgabe).

Nebenhypothesen

- H2₀: Es bestehen keine Niveauunterschiede im Leistungsbereich in Abhängigkeit vom Testleiter.
- H2₁: Es bestehen Niveauunterschiede im Leistungsbereich in Abhängigkeit vom Testleiter.
- H3₀: Es bestehen keine Niveauunterschiede im Leistungsbereich in Abhängigkeit von der Einstellung des Testleiters und des Vorgabemodus (Wechselwirkung).
- H3₁: Es bestehen Niveauunterschiede im Leistungsbereich in Abhängigkeit von der Einstellung des Testleiters und des Vorgabemodus (Wechselwirkung).

H_{4_0} : Es lassen sich keine Unterschiede im Befinden der Testpersonen zwischen den beiden Vorgabemodi finden.

H_{4_1} : Es lassen sich Unterschiede im Befinden der Testpersonen zwischen den beiden Vorgabemodi finden.

3.2 Untersuchungsplan

In diesem Abschnitt sollen die Rahmenbedingungen der Untersuchung grob skizziert werden. Genaue Ausführungen finden sich in den weiterführenden Kapiteln.

Die Testung fand an 3 Wiener Schulen statt, welche in der Art und Weise ausgewählt wurden, dass es wahrscheinlich war, dass die Leistungen der SchülerInnen vergleichbar waren und sich nicht in hohem Maße aufgrund der Rahmenbedingungen, Lehrmethoden und ethnische, kulturelle und soziale Herkunft der SchülerInnen unterscheiden würden. Aus diesem Grund wurden 3 von der Untersuchungsleiterin persönlich bekannte Schulen ausgewählt, um die oben angeführten Variablen geeignet abschätzen zu können. Als erster Schritt wurde, nach Festlegung der Hypothesen und des Testmaterials, das Einverständnis der Schuldirektoren/-direktorinnen eingeholt um das Ansuchen um Bewilligung einer wissenschaftlichen Erhebung beim Wiener Stadtschulrat zu stellen.

Die Testung fand in folgenden Schulen statt:

- G und RG DeLaSalle Schule Strebersdorf, 1210 Wien
Direktor: Mag. Peter Strass

- G und RG Bertha v. Suttner – Schulschiff, 1210 Wien
Direktorin: Mag. Judith Kovacic

- G und RG Franklinstraße 26, 1210 Wien
Direktorin: Mag. Eva Reschke

Nach Erhalt der Bewilligung folgte das Einholen des Einverständnisses der Eltern und KlassenlehrerInnen. Dies geschah mittels eines Eltern- und eines LehrerInnenschreibens. Es nahmen nur diejenigen Kinder an der Testung teil, welche ein schriftliches Einverständnis der Eltern erbrachten. Zu diesem Zwecke konnte ein, dem Schreiben angefügter, Abschnitt unterschrieben abgegeben werden.

Benötigt würde eine Stichprobe von 34 Kindern pro Vergleichsgruppe (Einzel- und Gruppenvorgabe). Die Stichprobengröße wurde mittels G-Power¹ für einen auf zu deckenden Effekt von 0,7 (bei $\alpha=0,05$ und $1-\beta=0,8$ für einen t-Test) berechnet, welcher von der Untersuchungsleiterin festgelegt wurde. Die Gruppe der Teilnehmenden an der Gruppentestung würde aus untersuchungsinhaltlichen Gründen größer ausfallen als die vorab berechneten *insgesamt* 34 Personen, da die Effekte welche durch die Gruppendynamik zu erwarten sind erst ab einer Gruppengröße von vermutlich mindestens 10 Personen pro Gruppe auftreten würden (genaue Untersuchungen dazu fehlen) - da in weiterer Folge feststehen wird, dass 4 Gruppen getestet werden, sprechen wir also von erwarteten 40 getesteten Personen. Die unterschiedlichen Gruppengrößen sollen aber bei der Berechnung des Leistungsunterschiedes keine Rolle spielen (nähere Erläuterungen sind in Kapitel 3.3 zu finden). Es sollen SchülerInnen zwischen 15 und 19 Jahren in die Testung mit einbezogen werden. Das entspricht in den ausgewählten Schulformen des *Oberstufengymnasiums* (9. – 12. Schulstufe). Da damit zu rechnen war, dass ein Testleiter maximal bis zu 10 Kinder im Einzelsetting testen kann (zwecks zeitlicher Ökonomie), wurden 4 Testleiter akquiriert. Um den Einfluss des Geschlechtes des Testleiters auszuhalten, wurde diese Variable konstant gehalten und aufgrund der besseren Zugänglichkeit wurden weibliche Testleiterinnen ausgewählt. Auch wurde im Vergleich zur Vorstudie das Alter der Testleiterinnen konstant gehalten. Sie sollten als *jung* einzustufen zu sein. Wichtig sollte sein, dass sie über grundlegenden theoretischen und praktischen Kenntnissen über die Psychologische Diagnostik verfügen, jedoch keine vertiefenden Kenntnisse besitzen. Zur Feststellung eventueller Testleitererwartungseffekte wurde eine Einstellungsmanipulation durchgeführt. Die genaue Vorgehensweise ist in Kapitel 3.2.4 zu finden. Jede Testleiterin sollte etwa 10 Kinder im Einzelsetting testen und eine Schulklasse (ca. 15 SchülerInnen) im Gruppensetting. Die Einzeltestungen fanden nacheinander statt. Die Gruppentestung erfolgte mit allen teilnehmenden SchülerInnen auf einmal. Die 4 Testleiterinnen teilten sich auf die ausgewählten 3 Schulen wie folgt auf:

Schule 1: Testleiterin 1

Schule 2: Testleiterin 2

Schule 3: Testleiterin 3 und 4

Eine komplett randomisierte Zuteilung der Testpersonen auf die Untersuchungsbedingungen würde aus ökonomischen Gründen nicht möglich sein. So wurden die Testleiterinnen randomisiert auf die Schulen aufgeteilt und führten dort jeweils ihre gesamte Testung durch. Jeder Testleiterin wurden 2 Klassen zugeteilt, welche im Voraus vom Schuldirektor/von der

¹ Programm zur statistischen Analyse

Schuldirektorin zur Verfügung gestellt wurden. Aus einer Klasse wurden die Testpersonen für die Einzelvorgabe zufällig ausgewählt, die andere Klasse wurde komplett getestet um die Daten für den Vorgabemodus *Gruppentestung* zu gewinnen. In welcher der, der Testleiterin zugeteilten, zwei Klassen jeweils die Einzeltestungen und in welcher die Gruppentestungen durchgeführt wurde, entschied ebenfalls das Los.

Die Testung wurde mit einem Untertest des *Wilde Intelligenztest 2 (WIT-2)*² durchgeführt, welcher in Form eines Papier-Bleistift-Verfahrens angewandt wird. Der verwendete Untertest „Abwicklungen“ misst die Dimension *räumliches und schlussfolgerndes Denken* (Thurstones Bezeichnung: Reasoning; siehe Kapitel 2.1.1). Die Testung nimmt inklusive Instruktion und Bearbeitung des Befindensfragebogens in etwa 30 Minuten in Anspruch. Es handelt sich um ein Multiple-Choice-Antwortformat, bei dem aus 5 Antwortmöglichkeiten die Richtige ausgewählt werden soll. Die Schulkinder werden darauf aufmerksam gemacht, dass die Testung vollkommen anonym ist und dass die Testwerte nur für die vorliegende Diplomarbeit verwendet werden. Auch die Eltern, LehrerInnen und Direktoren/Direktorinnen wurden über diesen Umstand informiert. Die Testungen wurden in den Räumlichkeiten der jeweiligen Schule durchgeführt und erfolgten von Mitte Dezember 2010 bis Mitte Jänner 2011.

² Kersting, Althoff & Jäger; 2008

3.3 Stichprobe

Insgesamt wurden 111 Kinder im Alter von 15 bis 19 Jahren getestet. Es mussten keine Testpersonen ausgeschieden werden, somit konnten 111 Fälle in die Berechnung mit einbezogen werden.

Die Testpersonen stammen aus Schulen aus dem 21. und 22. Wiener Gemeindebezirk. Die Schulen zeichnen sich durch einen geringen Immigrantenteil aus, weshalb zu vermuten ist, dass die Instruktion von allen Kindern gleich gut verstanden werden konnte. Sprachverständnisprobleme könnten vor allem in der Gruppentestung zu Problemen führen. Des Weiteren erwähnen die Testautoren des verwendeten Verfahrens im Testmanual ausdrücklich, dass die perfekte Beherrschung der deutschen Sprache Voraussetzung für die Anwendbarkeit des Tests ist, weshalb die Wahl auf die oben genannten Schulen fiel.

33% der Testpersonen waren männlich (Häufigkeit: 37) und 67% waren weiblich (Häufigkeit: 74).

Von den 111 Getesteten, befanden sich 33 in der Einzeltestung und 78 in der Gruppentestung. Eine zuvor durchgeführte Berechnung der benötigten Stichprobe um einen mittleren Effekt auf zu decken ($d=0,7$) ergab folgendes Ergebnis³:

[3] -- Tuesday, May 03, 2011 -- 18:36:44

t tests – Means: Difference between two independent means (two groups)

Analysis: A priori: Compute required sample size

Input:	Tail(s)	=	Two
	Effect size d	=	0.7
	α err prob	=	0.05
	Power (1- β err prob)	=	0.80
	Allocation ratio N2/N1	=	1
Output:	Noncentrality parameter δ	=	2.8861739
	Critical t	=	1.9965644
	Df	=	66
	Sample size group 1	=	34
	Sample size group 2	=	34
	Total sample size	=	68
	Actual power	=	0.8116461

Es wurde versucht diese Stichprobengrößen ein zu halten (siehe Kapitel 3.2).

³ Berechnung mittels G-Power 3.1

Die Verteilung des Alters stellt sich wie folgt dar:

Alter		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	15	52	46,8	46,8	46,8
	16	35	31,5	31,5	78,4
	17	15	13,5	13,5	91,9
	18	8	7,2	7,2	99,1
	19	1	,9	,9	100,0
Gesamt		111	100,0	100,0	

Tabelle 1 Altersverteilung – Alter in Jahren

Der Mittelwert des Testpersonenalters beträgt 15,84 Jahre.

Die Anzahl der getesteten Kinder pro Testleiterin wird in Tabelle 2 wiedergegeben. Gut erkenntlich ist, dass die Anzahl der Testungen pro Testleiterin in etwa gleich groß ist.

		Häufigkeit	Prozent
Gültig	Testleiter 1	34	30,6
	Testleiter 2	21	18,9
	Testleiter 3	30	27,0
	Testleiter 4	26	23,4
	Gesamt	111	100,0

Tabelle 2 Anzahl der Testpersonen pro Testleiter

Weiters wird in Tabelle 3 die Anzahl der jeweiligen Einzel- und Gruppentestungen angegeben. Es wurde, wie schon erwähnt, versucht die Gruppengrößen gleich zu halten. Die Abweichung zwischen Einzel- und Gruppentestung soll aber bei der verwendeten, recht robusten Auswertmethode keine Rolle spielen.

	Einzeltestung	Gruppentestung
Testleiter 1	9	25
Testleiter 2	9	12
Testleiter 3	9	22
Testleiter 4	7	19

Tabelle 3 Aufteilung nach Testleiter und Vorgabemodus

3.4 Erhebungsinstrumente

3.4.1 WIT – 2

Die Testung wurde mit dem Wilde Intelligenz Test-2 (WIT-2; Kersting, Althoff & Jäger, 2008) durchgeführt. Er dient der differenzierten Erfassung kognitiver Fähigkeiten von Jugendlichen und Erwachsenen und ermöglicht die Testung folgender Dimensionen: (1) schlussfolgerndes Denken, (2) sprachliches Denken, (3) Rechnerisches Denken, (4) Räumliches Denken, (5) Merkfähigkeit, (6) Arbeitseffizienz, (7) Wissen und Wirtschaft und (8) Wissen Informationstechnologie. Es ist möglich zwischen leistungsstarken und leistungsschwachen Testpersonen zu differenzieren, da das Verfahren einfache, schwierigere und sehr schwierige Items enthält. Der WIT-2 ist eine Weiterentwicklung des WIT (Jäger & Althoff, 1994). Die theoretische Grundlage ist das Intelligenzmodell von Thurstone – Modell der Primary Mental Abilities (PMA) (vgl. Kersting et al., 2008). Die Intelligenzfaktoren wurden in Kapitel 2.1.1 skizziert.

Der WIT-2 ist nicht nur in der Gesamtversion, sondern auch Modulweise verwendbar. Die Module entsprechen den oben angeführten Dimensionen.

Die Auswertung erfolgt mittels Auswertschablone.

Die wesentliche Voraussetzung für das ausgesuchte Verfahren war, die *ausdrückliche* Verwendbarkeit als Einzel- und als Gruppenvorgabe, ohne Bereitstellung getrennter Normen für den jeweiligen Vorgabemodus. Beide werden vom WIT-2 erfüllt.

(a) Testmaterial

Das Testmaterial besteht aus einem Testheft und einem Antwortbogen, in welchem der zu bearbeitende Untertest (*Abwicklungen*) deutlich mit Leuchtstift markiert wurde. Da für die meisten Verfahren, welche in der Gruppe vorgegeben werden können, zwei Testformen existieren, wurde auch hier Form A und Form B verwendet um ein Abschreiben der Testpersonen untereinander zu verhindern. Zu Beginn des Testheftes befindet sich eine Einführung, wie der Test im Allgemeinen zu bearbeiten ist. Den Schulkindern wurde diese Einführung vom Testleiter vorgelesen, so wie es die Testautoren vorsehen. Die Testpersonen selbst hatten die Möglichkeit mit zu lesen. Für die TestleiterInnen stellen die Autoren ein

eigenes Instruktionsheft zur Verfügung, welches neben den Testaufgaben noch zusätzliche Ablauf- und Erklärungshinweise beinhaltet. Die 4 Testleiterinnen bekamen von der Untersuchungsleiterin noch zusätzliche Instruktionen vor der Testung, welche über die Anleitungen im Instruktionsheft hinausgingen – sowohl schriftlich als auch mündlich. Dies war notwendig da sich die Testung aufgrund der Vorgabe nur eines Untertestes von der üblichen Vorgabe unterschied. Die Instruktion erfolgte für alle gleich.

Es werden den Testpersonen 2 Beispielitems erklärt und auf Verständnis geprüft. Hier haben die TestleiterInnen die Möglichkeit, im Vergleich zur Instruktion welche vorgelesen werden soll und eigentlich keine Fragen offen lässt, ihren eigenen Stil und ihren eigenen Charakter in die Erklärung einfließen zu lassen, falls es von Schülerseite Fragen geben sollte.

(b) Untertests

Idealerweise sollte ein sprachfreies Verfahren und ein sprachgebundenes Verfahren zur Anwendung kommen. Da aus zeitlichen Gründen (sowohl von Direktoren-/Direktorinnenseite, als auch von Seiten der Testleiterinnen) jedoch nicht mehr als eine knappe halbe Stunde für die Testung zur Verfügung stand, war eine Kombination der beiden notwendigen Dimensionen nicht möglich. Da die Entscheidung auf lediglich einen Untertest fallen musste, entschied sich die Untersuchungsleiterin sich in der vorliegenden Diplomarbeit für die Testung der Dimension *schlussfolgerndes Denken*, da diese, wie schon in Kapitel 2.1.1 ausgeführt, als die grundlegende geistige Fähigkeit gesehen wird. Schlussfolgerndes Denken wird nach Thurstone die Fähigkeit bezeichnet, bestimmte logische Regeln und Gesetzmäßigkeiten zu erkennen und zweckentsprechend anwenden zu können. Der WIT-2 bietet dazu 3 Untertests an: (1) Analogien, (2) Abwicklungen und (3) Zahlenreihen. Da der Untertest *Abwicklungen* mit einer Vorgabedauer von laut Manual angegebenen 14 Minuten inklusive Instruktion genau in den Zeitplan passte, fiel die Entscheidung auf diesen.

(c) Untertest Abwicklungen

Der Untertest *Abwicklungen* besteht aus 20 Items. Der Testleiter/die Testleiterin bekommt im Instruktionsheft die Anweisung die Testung nach 9 Minuten zu beenden. Die Testpersonen bekommen im Voraus diese Information nicht. Für bis zu diesem Zeitpunkt nicht bearbeitet Items bekommt die Testperson keine Punkte. Die getestete Dimension ist räumliches und schlussfolgerndes Denken. Die Aufgabe besteht aus einer Faltvorlage mit mehreren Flächen.

Zu dieser Faltvorlage ist aus 5 vorgegebenen Körpern derjenige aus zu wählen, der sich aus der Faltvorlage herstellen lässt.

Die jeweils gewählte Antwort ist in dem beigelegten Antwortbogen an zu kreuzen. Wie dies geschehen soll wird im Laufe der Instruktion von der Testleiterin erläutert. Zur Veranschaulichung werden Beispiele an der Tafel oder in der Einzeltestung an einem Blatt Papier demonstriert.

(c) Auswertung

Die Auswertung der Testbögen erfolgte mittels im Testpaket mitgelieferten Auswerteschablonen. Auf diesen sind die jeweils richtigen Antworten markiert. Der Rohwert wurde, nicht wie für die konventionelle Durchführung vorgesehen, auf einem umfassenden Auswertbogen eingetragen, sondern lediglich auf dem Testbogen notiert. Es wurde diese Vorgehensweise bevorzugt, da der Auswertbogen für die Durchführung des Gesamttests oder einzelner Module konzipiert wurde und in diesem Rahmen schlicht nicht benötigt wurde, da es sich jeweils nur um einen einzelnen Testwert handelte. Dieser wurde mittels Normtabelle (siehe Anhang) für das Alter zwischen 14 und 17 und der Normtabelle für 18 Jahre in einen Standardwert (auch Z-Wert) transformiert. Es wurden Normtabellen ausgewählt welche für die Schulbildung nicht differenzieren. Für die spätere Berechnung notwendigerweise, wurde dieser Standardwert für jede Testperson in einen T-Wert umgewandelt. Dazu wurde das Normwertrechnerprogramm *Psychometrica* verwendet.

3.4.2 Befindensfragebogen

Wie in Abschnitt 2.2.3 b ausgeführt wurde, könnte sich das Befinden der Testpersonen auf die Leistung niederschlagen. Da sich Testpersonen durch die beiden Vorgabemodi (Einzel- vs. Gruppenvorgabe) in deutlich anderen Situationen befinden, ist an zu nehmen, dass sich auch die Testpersonen in unterschiedlichem Maße Wohlfühlen. Je nach Vorliebe der einen oder anderen Testsituation, könnte dies bessere oder schlechtere Testleistungen zur Folge haben. Um unter Umständen gefundene Leistungsunterschiede in den beiden Vorgabemodi erklären

zu können, wird nach der Testung von den Schulkindern der Befindensstatus erhoben. Es wurden zwei Kurzfragebogen gestaltet, die auf die jeweilige Vorgabeform angepasst ist. Der Fragebogen enthält jeweils 11 Fragen zu den Themengebieten, welche in Kapitel 2.2 angeführt wurden. Da beispielsweise vermutet wird, dass ein Nachteil der Gruppentestung die vermehrte Störung durch andere ist, wurde die Testperson gefragt, ob es störende Einflüsse während der Testung gab. Die Fragen wurden in den beiden Fragebogen so gestellt, dass sie einen Vergleich der beiden Bedingungen erlauben. So gibt es dem nach zu jeder Frage im Gruppenfragebogen eine gleichwertige im Einzelfragebogen.

Die Antwort ist auf einer 5-stufigen Skala von „trifft voll zu“ bis „trifft gar nicht“ zu an zu kreuzen. In der Instruktion wird deutlich darauf hingewiesen, dass es darum geht, wie es den Testpersonen während der vorangegangenen Testung gegangen ist und dass es keine richtigen und falschen Antworten gibt. Sie werden auch aufgefordert ehrlich zu antworten.

3.5 TestleiterInnen

In der vorliegenden Studie soll die Möglichkeit bestehen, etwaige Unterschiede im Leistungsniveau in Abhängigkeit vom Vorgabemodus, durch Testleitererwartungseffekte erklären zu können, muss die Testung von 4 unabhängigen TestleiterInnen durchgeführt werden. Eine Einstellungsmanipulation ist für das Aufdecken von Testleitererwartungseffekten notwendig. Die manipulierte Einstellung/Meinung der TestleiterInnen soll sich auf die erwarteten Testleistungen der SchülerInnen beziehen.

Die Anzahl von 4 TestleiterInnen kommt durch die benötigte Anzahl von getesteten SchülerInnen zustande (siehe Kapitel 3.3). Um den Zeitaufwand für die TestleiterInnen ökonomisch zu gestalten, sollten die Testungen für jeden Testleiter an einem Tag möglich sein. Da, wie vorab berechnet, 34 Testpersonen pro Vorgabemodus benötigt werden, wurden 4 TestleiterInnen eingesetzt um die oben erwähnte ökonomische Durchführung zu gewährleisten. Das bedeutet einen Zeitaufwand von 5,5 Stunden pro Testleiter (bei einer großzügigen Testdauer von 30 Minuten – 5 Stunden für Einzeltestungen und 30 Minuten für die Gruppentestung). Keine der Testleiterinnen kannte die Vorstudie von Reindl (1998). Sie haben auch, über die Standardausbildung hinausgehend keine vertiefende Kenntnis über Versuchsleitererwartungseffekte. Das Durchschnittsalter beträgt 25 Jahre.

Die Manipulation der Einstellung erfolgte wie schon in der Vorstudie durch Bereitstellung falscher Informationen. Das *elaboration-likelihood model* von Petty und Cacioppo (1981; zitiert nach Herkner; 2008) besagt, dass zwei Wege der Informationsverarbeitung existieren. Sie unterscheiden den zentralen Weg vom peripheren Weg der Verarbeitung. Die Informationen und Einstellungen die die Testleiterinnen bisher über die Leistungen von Testpersonen in den beiden Vorgabemodi gespeichert haben, werden anzunehmender Weise über den peripheren Weg geschehen sein, da angenommen werden kann, dass keine gründliche kognitive Auseinandersetzung mit dem Thema passiert ist. So entstandene Einstellungen sind leicht zu ändern - noch dazu, wenn die neue Information auf scheinbaren Fakten basiert. Zusätzlich wird die neue Information in diesem Fall eher auf dem zentralen Weg verarbeitet, da die Testleiter nun an der Mitteilung interessiert sind – sie sind aktiver Teil einer Untersuchung und haben dadurch persönlichen Bezug zum Thema. Des Weiteren werden den Testleitern mehrere Informationsquellen angegeben, welche alle dieselbe Meinung enthalten, somit gibt es nur wenig Grund die Richtigkeit der Informationen anzuzweifeln.

Zwei Testleiterinnen bekamen nun die Information, dass die Kinder im Gruppensetting bessere Ergebnisse erzielen werden. In weiterer Folge werden wir von der Einstellung „pro Gruppe“ sprechen. Den anderen zwei Testleiterinnen wurde mitgeteilt, dass die Ergebnisse der Kinder, welche einzeln getestet wurden, sicher besser sein werden („pro Einzel“). Vor der Testung wurden allen Testleiterinnen die Informationen in einem persönlichen Gespräch übermittelt. Ihnen wurde auch eine schriftliche Version zum Nachlesen mitgegeben (siehe Anhang).

Testleiterin 1	Pro Einzel
Testleiterin 2	Pro Gruppe
Testleiterin 3	Pro Einzel
Testleiterin 4	Pro Gruppe

Tabelle 4 Einteilung der Testleiterinnen zu den Manipulationsbedingungen

3.6 Durchführung der Untersuchung

Nachdem die von den Direktoren/Direktorinnen ausgesuchten SchülerInnen die Bestätigung der Erziehungsberechtigten über das Einverständnis für die Teilnahme an der Untersuchung erbracht haben, wurde von der Untersuchungsleiterin der Termin für die Durchführung mit den Testleiterinnen vereinbart. Die Testpersonen wurden über den Eltern- und Lehrerbrief schon vorab über die Untersuchung informiert. Nach Akquirierung der Testleiterinnen fand die Einstellungsmanipulation mit jeder Testleiterin einzeln statt. Ein Teil erfolgte im persönlichen Gespräch, der andere Teil erfolgte schriftlich in Form eines zusammenfassenden Schreibens. Am Tag der Testung bekamen die Testleiterinnen alle notwendigen Unterlagen ausgehändigt und wurden nochmals über den Ablauf instruiert. Die Untersuchungsleiterin war bei jeder Testung in der Schule anwesend und nahm nach Beendigung die Unterlagen wieder an sich. Je nach schuleseitiger Planung erfolgte zuerst die Gruppen- oder Einzeltestung. Die Gruppentestung fand in dem jeweiligen Klassenzimmer statt. Für die Einzeltestungen wurden ruhige Räume zur Verfügung gestellt. Die Testleiterinnen befanden sich jeweils alleine mit den SchülerInnen im Raum.

Die gesamte Durchführung fand laut Manual statt. Zuerst wurde instruiert und mit einem „Kontrollbeispiel“, welches von den Testpersonen selbst zu lösen war, wurde das Aufgabenverständnis geprüft. Für die Bearbeitung der restlichen 20 Items war 9 Minuten Zeit. Den Testpersonen wurde jedoch im Voraus keine Information über die Zeitbeschränkung erteilt. Nach Ablauf der neun Minuten wurde der Befindensfragebogen zum Ausfüllen ausgehändigt. In der Einzeltestung wurde jeder Testperson zu Beginn der Testung der Ablauf extra erklärt. In der Gruppentestung fand die Instruktion für die gesamte Gruppe auf einmal statt. Jeder Testdurchlauf dauerte an die 25 Minuten.

3.7 Ergebnisse

Im Folgenden sollen die Ergebnisse in Bezug zu den aufgestellten Hypothesen dargestellt werden. Die abhängige Variable (AV) ist bei jedem Signifikanztest die Variable „t-Wert“ und stellt somit das Testergebnis dar. Sämtliche Auswertungstabellen und Grafiken sind im Anhang zu finden.

Die Auswertung erfolgte mithilfe der Analyseprogramme *PASW® Statistics 18* und *GPow* 3.1.

H1: Es bestehen Leistungsunterschiede im Leistungsbereich in Abhängigkeit vom Vorgabemodus (Einzel- versus Gruppenvorgabe).

Die Auswertung der Mittelwertsunterschiede der Leistung in der Einzel- bzw. Gruppenvorgabe erfolgte mit dem Welch-Test – es wurde auf die Testung mittels t-Test verzichtet (siehe Rasch & Kubinger 2011) -, welcher ein asymptotisches $F(1;4,1) = 0,527$ ergibt und somit die H_0 nicht verworfen wird. Der Mittelwert der T-Werte in der Gruppenvorgabe beträgt 49,26 und in der Einzeltestung 51,06.

Es konnten daher kein signifikanter Unterschied im Leistungsniveau in Abhängigkeit vom Vorgabemodus gefunden werden.

H2: Es bestehen Niveauunterschiede im Leistungsbereich in Abhängigkeit vom Testleiter/der Testleiterin.

Die Ergebnisse gibt die zweifaktorielle ANOVA (Haupteffekt *TL*) wieder.

Die H_0 wird verworfen ($F(3,107)=3,431, p<.05$). Es lassen sich signifikante Unterschiede in der Leistung in Abhängigkeit vom Testleiter/der Testleiterin finden. Es wurde hierfür keine Aufteilung nach Einzel- und Gruppentestung vorgenommen. Die Berechnung bezieht alle vom jeweiligen Testleiter getesteten Personen mit ein.

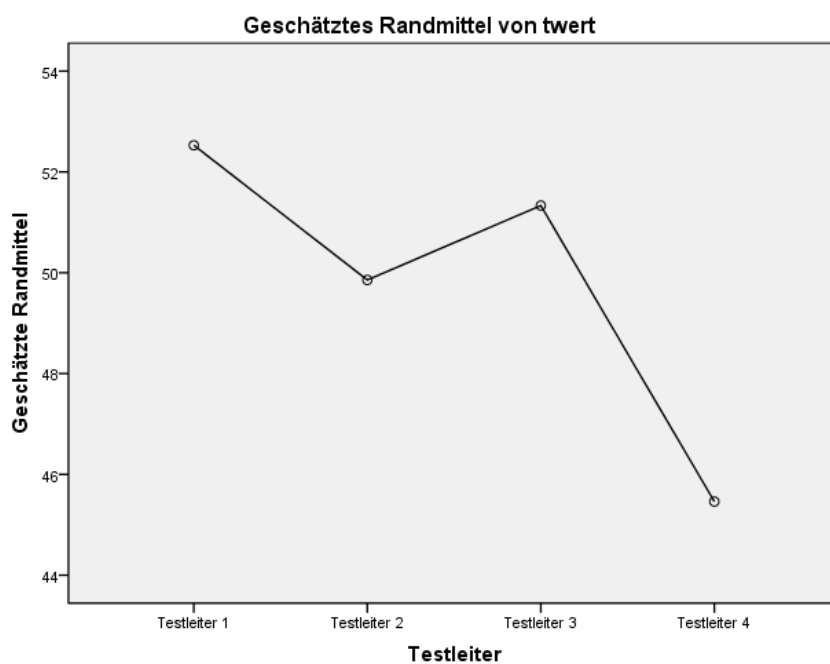
Die Mittelwerte zwischen den Testleiterinnen unterscheiden sich um 7 T-Werte. In Tabelle 5 finden sich die Mittelwerte der von den verschiedenen Testleiterinnen getesteten SchülerInnen wieder.

twert

	N	Mittelwert	Standardabweichung	Standardfehler
Testleiter 1	34	52,53	8,938	1,533
Testleiter 2	21	49,86	6,945	1,515
Testleiter 3	30	51,33	8,691	1,587
Testleiter 4	26	45,46	9,941	1,950
Gesamt	111	50,05	9,080	,862

Tabelle 5 Mittelwerte aufgeteilt nach Testleiter

Der Post-Hoc-Vergleich nach Scheffè zeigt einen signifikanten Unterschied der Testergebnissen der SchülerInnen zwischen Testleiterin 1 und Testleiterin 4.



H3₁: Es bestehen Niveauunterschiede im Leistungsbereich in Abhängigkeit von der Einstellung des Testleiters/der Testleiterin und des Vorgabemodus.

Die zweifaktorielle ANOVA (F1=Testleiter; F2=Vorgabemodus) ergibt keinen signifikanten Wechselwirkungseffekt $F(3,103) = 0,592$.

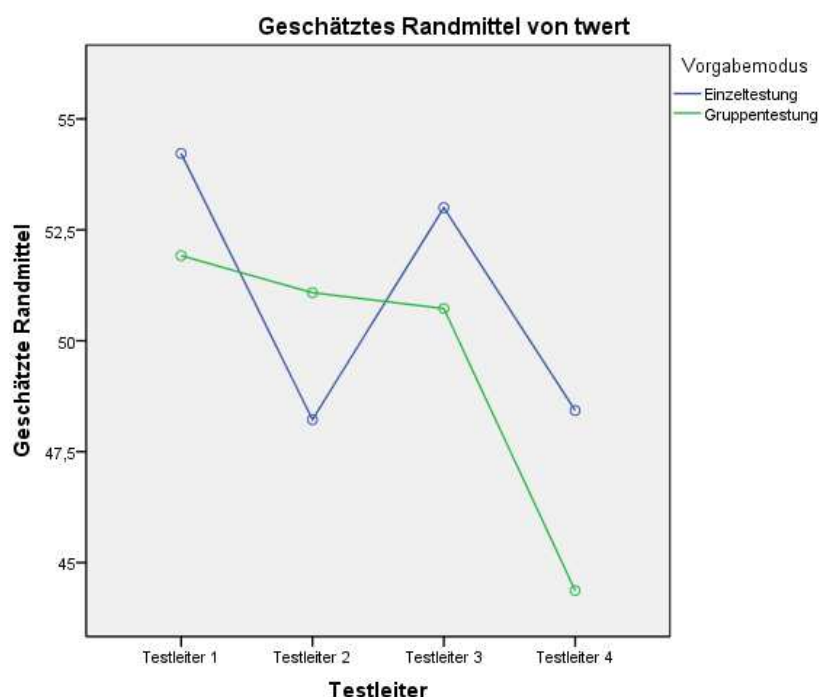
Die Voraussetzungen, um eine zweifaktorielle Varianzanalyse gültig berechnen zu können, wurden erfüllt.

Tests der Zwischensubjekteffekte

Abhängige Variable:twert

Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	998,408 ^a	7	142,630	1,820	,091
Konstanter Term	278000,225	1	278000,225	3548,045	,000
tl	806,605	3	268,868	3,431	,020
modus	52,711	1	52,711	,673	,414
tl * modus	139,092	3	46,364	,592	,622
Fehler	8070,367	103	78,353		
Gesamt	287069,000	111			
Korrigierte Gesamtvariation	9068,775	110			

a. R-Quadrat = ,110 (korrigiertes R-Quadrat = ,050)



H4₁: Es lassen sich Unterschiede im Befinden der Testpersonen in den beiden Vorgabemodi finden.

Die Ergebnisse wurden mittels Kreuztabellen deskriptiv dargestellt (siehe Anhang). In den Tabellen ist die Anzahl der gegebenen Antworten angegeben, sowie die jeweilige Spalten- und Zeilensummen. In der Kopfzeile ist mit „E“ und „G“ der Bezug zur Frage im Einzel- und im Gruppenfragebogen angegeben.

Je ein Item aus jedem der beiden Fragebögen musste ausgeschlossen werden (keine gültige Vergleichbarkeit gegeben). Somit resultieren 10 vergleichbare Items.

Pro Kreuztabelle ist untenstehend die jeweilige zusammenfassende Metaaussage angegeben, welche sich aus der Fragestellung im Fragebogen ergibt und eine Interpretation der Ergebnisse des Fragebogens. Ebenso die Interpretation der Auswertung in Bezug auf die Metaaussage.

Die interpretativen Aussagen beziehen sich nur auf die getestete Stichprobe.

Die Vorgabe des Fragebogens sollte dem Zweck dienen, mögliche Unterschiede zwischen den Vorgabemodi erklären zu können. In der vorliegenden Untersuchung konnten jedoch keine relevanten Niveauunterschiede ausgemacht werden, wodurch die Auswertung des Fragebogens nur vollständigkeithalber durchgeführt wurde.

(1) Metaaussage: Die Situation mit dem Testleiter alleine zu sein ist unangenehm

Interpretation der Ergebnisse:

Alleine mit dem Testleiter zu sein bereitet größtenteils keine Probleme/Unwohlsein.

(2) Metaaussage: In der jeweiligen Situation konnte man sich nicht konzentrieren, man hätte die andere Situation bevorzugt

Interpretation der Ergebnisse:

In der jeweils befindlichen Situation konnte man sich gut konzentrieren. In der subjektiven Konzentration zeigten sich keine Unterschiede zwischen den beiden Vorgabemodi.

(3) Metaaussage: Der Test war mir egal

Interpretation der Ergebnisse:

In der Gruppentestung wird der Test eher als egal empfunden. Möglicherweise wird der Testung im Einzelsetting eine größere Wichtigkeit beigemessen.

(4) Metaaussage: Es ist wichtig nicht beobachtet zu werden.

Interpretation der Ergebnisse:

„Schutz“ vor anderen wird nicht benötigt. Egal ob es der Schutz durch die Gruppe ist (Anonymität) oder der Schutz durch das Alleinsein im Einzelsetting.

(5) Metaaussage: Aktivitäten mit den Klassenkollegen/-kolleginnen werden gemocht

Interpretation der Ergebnisse:

Eine mögliche Abneigung gegen Aktivitäten mit der Klasse und somit auch eine Abneigung gegen eine Gruppentestung mit dieser Klasse muss nicht befürchtet werden.

(6) Metaaussage: Es konnte geschummelt/abgeschrieben werden

Interpretation der Ergebnisse:

Durchwegs stellt schummeln/abschreiben kein Problem dar, was jedoch nur mit Vorsicht aufgrund der Selbstangaben zu beurteilen ist.

(7) Metaaussage: Anwesenheit einer Ablenkung

Interpretation der Ergebnisse:

Weder in Einzel- noch in Gruppentestung fühlten sich die Testpersonen/Probandinnen abgelenkt.

(8) Metaaussage: Prinzipielles Wohlfühlen während der Testung

Interpretation der Ergebnisse:

Der Großteil der Testpersonen hat sich in der *jeweiligen* Situation wohlgefühlt.

(9) Metaaussage: Der Testleiter beschäftigt sich subjektiv besonders mit einem

Interpretation der Ergebnisse:

In der Einzeltestung finden die Testpersonen, dass sich der Testleiter besonders mit ihnen beschäftigt, was möglicherweise zu negativen Effekten bei unsympathischen Testleitern führen kann.

(10) Metaaussage: Ich arbeite gerne in der Gruppe (besondere Anregung) vs. ich arbeite gerne alleine (ohne Störung)

Interpretation der Ergebnisse:

Das Arbeiten alleine, weil es keine Störung gibt wird als positiv beurteilt, jedoch bietet die Gruppe keine zusätzliche Anregung zu besseren Leistungen. Dies zeigt eine Tendenz in Richtung, dass das Arbeiten in der Gruppe keinen „Zusatzgewinn“ bringt.

Zusammenfassung der empirischen Ergebnisse

Ein Unterschied zwischen den Vorgabemodi konnte nicht nachgewiesen werden (*Haupthypothese*).

Des Weiteren konnte ein statistisch signifikanter und inhaltlich bedeutsamer Einfluss des Testleiters/der Testleiterin auf die Testleistungen von Kindern gefunden werden (*H2*).

Der Testleitererwartungseffekt konnte jedoch nicht festgestellt werden (*H3*), da der Wechselwirkungseffekt zwischen den unabhängigen Variablen „Testleiter“ und „Vorgabemodus“ nicht signifikant ausfiel. Daher beeinflusste die positive oder negative Einstellung des Testleiters gegenüber der Gruppentestung oder der Einzeltestung (hervorgerufen durch die in Kapitel 3.5 beschriebene Einstellungsmanipulation) die Testleistungen der Testpersonen nicht. Betrachtet man die Mittelwerte der Testpersonen/Probandinnen unter den 4 Testleiterinnen, erkennt man zwar eine Tendenz in Richtung des Testleitererwartungseffektes, jedoch fallen die Testergebnisse welche unter der Testleiterin 4 erbracht wurden aus dem Schema. Testleiterin 1 wurde in Richtung „pro

Einzel“ manipuliert. Die Testung zeigte, dass in dieser Bedingung die Kinder ein etwas höheres Leistungsniveau zeigten. Genau so verhielt es sich bei Testleiterin 3. Die Testleiterinnen 2 und 4 waren in Richtung „pro Gruppe“ eingestellt. Bei Testleiterin 2 zeigte sich auch der erwartete Effekt, dass die SchülerInnen in der Gruppentestung besser abschneiden würden. Jedoch bei Testleiterin 4 waren die Ergebnisse gegenteilig zur Erwartung. Hier war das Leistungsniveau in der Einzeltestung höher. Erwartet worden wäre eine bessere Leistung in der Gruppentestung.

Auch wenn sich zwar eine tendenzielle Richtung abzeichnet, sind die Unterschiede der beiden Vorgabemodi prinzipiell zu gering (siehe nicht signifikante *Haupthypothese*), um dessen zu große Beachtung zu schenken.

Erklärungen für die oben beschriebenen unterschiedlichen Testleistungen je nach dem von welchem Testleiter/welcher Testleiterin getestet wurde, könnten Effekte wie Sympathie, Routiniertheit, Auftreten, Motivationsgeschick oder Ähnliches sein. Dies würde weitere Vertiefung in das Thema bedürfen, sowie weiterführende verifizierende Untersuchungen.

Im subjektiven Befinden unterscheiden sich zwei Vorgabemodi so gut wie kaum voneinander. Unterschiede ließen sich finden in

- 1) der subjektiven Wichtigkeit des Testes (in der Gruppe wird der Test als mehr egal empfunden)
- 2) der Testleiter beschäftigt sich subjektiv besonders mit Testpersonen in der Einzeltestung
- 3) Die Testpersonen arbeiten gerne alleine weil, es keine Störeinflüsse gibt und das Arbeiten in Gruppen regt subjektiv nicht zu besserer Leistung an.

Durch diese äußerst geringen Unterschiede ist nicht davon aus zu gehen, dass das subjektive Befinden sich bedeutend auf das Niveau bei psychologischen Leistungstests auswirkt.

4. Diskussion und Ausblick

Da die vorliegende Untersuchung im Zuge einer Diplomarbeit durchgeführt wurde, waren die zeitlichen Ressourcen äußerst beschränkt. So ist kritisch anzumerken, dass lediglich ein einziges Verfahren angewendet wurde. Die Verwendung nur eines sprachfreien Verfahrens war jedoch anfänglich nicht geplant. Um korrekte Aussagen treffen zu können, wären mindestens zwei Verfahren notwendig gewesen, wovon eines sprachfrei und das andere ein Test mit sprachlicher Komponente sein sollte. Die Vorgabe einer ganzen Intelligenztestbatterie, wie sie beispielsweise der WIT-2 darstellt, wäre die ideale Ausgangsbasis. Da eine solche Intelligenztestbatterie jedoch meist weit mehr als 90 Minuten in Anspruch nehmen würde, war die Durchführung von Seiten der Schule nicht möglich. Bei einer nur 25-minütigen Testung bedeutete der Zeitaufwand für die Schule knappe 6 Stunden. Diese Zeit war das oberste Limit, welches zur Verfügung gestellt wurde. Die Koordination mit den Schulen stellte sich trotz des aus psychologisch-diagnostischer Forschungssicht relativ geringen Zeitaufwandes als sehr schwierig und langwierig heraus. Da es jedoch nicht selbstverständlich ist, für die Testung an Wiener Schulen überhaupt das Einverständnis zu bekommen, wurde die Testung auf das Minimum reduziert. Auch die Aquirierung von Testleitern/Testleiterinnen hätte sich als nahe zu unmöglich gestaltet, wäre es notwendig gewesen 2 Tage oder mehr den selben/die selbe TestleiterIn zu benötigen, ohne eine Aufwandsentschädigung anbieten zu können. Dies ist, wie schon angedeutet, im Zuge einer Diplomarbeit jedoch schwer möglich.

Da die vorliegende Diplomarbeit eine Wiederholungsstudie von Reindl (1998) darstellt und auch diesmal die Ergebnisse in die selbe Richtung gehen, nämlich dass es bezüglich des Niveaus von Testleistungen keinen Unterschied gibt, abhängig davon ob der Proband/die Probandin mittels Einzel- oder Gruppenvorgabe getestet wird, kann gesagt werden, dass weitere Untersuchungen in diese Richtung nicht als notwendig erachtet werden. Die Annahme der meisten Testautoren, dass keine getrennte Normierung für die Einzel- und Gruppenvorgabe von Nöten sind, da sich die Testleistungen nicht unterscheiden, kann bestätigt werden.

Neben der Hauptfragestellung, sind noch die Ergebnisse der Nebenhypothesen erwähnenswert. Der oft publizierte Testleitererwartungseffekt konnte in dieser Studie, trotz gezielter Manipulation der Testleiterinnen, in diesem Kontext nicht gefunden werden.

Dass der Testleiter/die Testleiterin jedoch prinzipiell Auswirkungen auf das Testergebnis hat, konnte auch hier, wie schon so oft, nachgewiesen werden. Hier konnte ein aussagekräftiger

Unterschied der Testleistungen aufgedeckt werden – die Ergebnisse der Testpersonen wurden hierfür je Testleiter ausgewertet und nicht nach Einzel- und Gruppentestung geteilt. Für die weitere Forschung im Bereich der psychologischen Diagnostik, sowie auch in der Anwendung psychologisch-diagnostischer Verfahren, bedeutet dies die Notwendigkeit der vermehrten Beachtung der TestleiterInnen Auswahl. Jedoch müsste vorab geklärt werden, welche Eigenschaften ein Testleiter/eine Testleiterin mitbringen muss, um bessere oder schlechtere Testergebnisse zu provozieren. Da dies in der vorliegenden Diplomarbeit jedoch nur ein „Nebenfund“ ist, sollen die Ergebnisse und die kurz angedeutete Konsequenz nicht näher erörtert werden, dennoch zum Weiterdenken anregen.

5. Zusammenfassung

Diese Diplomarbeit beschäftigt sich mit dem Thema, ob das Leistungsniveau von SchülerInnen eindeutig unterschiedlich ist, in Abhängigkeit davon ob ein Leistungstest in der Gruppe oder Einzeln vorgegeben wird. Die Studie stellt eine Wiederholung von der Diplomarbeit von Reindl (1998) dar und dient zur Absicherung der damaligen Ergebnisse.

Nebenfragestellungen beschäftigten sich mit dem Einfluss des Testleiters, Testleitererwartungseffekten und dem Befinden der Kinder während der Testung. Diese Faktoren sollen Erklärungsansätze für etwaige Unterschiede im Leistungsniveau darstellen.

Die Studie wurde an 3 Wiener Schulen durchgeführt. Um Testleitererwartungseffekte aufdecken zu können, wurden 4 Testleiterinnen eingesetzt, welche in ihrer Einstellung manipuliert wurden. Insgesamt wurden 111 SchülerInnen mit dem Untertest „Abwicklungen“ des Wilde Intelligenztests 2 getestet. Dieser Papier-Bleistift-Test wurde von jeder Testleiterin je einen Teil der ihr zugeteilten Kinder im Gruppen- und im Einzelsetting vorgegeben. Neben dem Leistungstest wurde am Ende der Testung jeder Testperson ein Befindensfragebogen vorgegeben, um einen Teil der Nebenfragestellungen beantworten zu können.

Die Datenauswertung und statistische Berechnungen wurden mittels *PASW Statistics 18* und *GPower 3.1* durchgeführt. Die Ergebnisse zeigen, dass keine Unterschiede in der Leistung zu finden sind, in Abhängigkeit davon ob Testpersonen in einer Gruppe oder Einzeln getestet werden. Auch ein Testleitererwartungseffekt konnte nicht nachgewiesen werden. Auch die Auswertung des Befindensfragebogens ergab keine relevanten Unterschiede im subjektiven Befinden der Testpersonen während der Testung. Ein signifikantes Ergebnis allerdings brachte die Auswertung des Testleitereinflusses. Die Ergebnisse der Testpersonen (nicht nach Einzel- oder Gruppenvorgabe getrennt – es wurden alle von der Testleiterin getesteten Personen miteinbezogen) verglichen nach Testleiterinnenzugehörigkeit unterscheiden sich bedeutsam. Daraus ist abzuleiten, dass jeder Testleiter in anderer Art und Weise die Testpersonen zu Leistung anspornt. Eine unterschiedliche Sympathie gegenüber den Testleiterinnen konnte anhand des Befindensfragebogens nicht gefunden werden. Um das Provozieren unterschiedlicher Leistungen zu unterbinden oder zu minimieren ist der Einsatz testleiterunabhängiger Verfahren angebracht. PC-Verfahren beispielsweise halten den Einfluss des Testleiters so gering wie möglich, um genau diesem Effekt des Testleitereinflusses aus dem Weg zu gehen. Aufgrund dieser Ergebnisse ist zu sagen, dass es keiner weiteren Untersuchung bezüglich der Niveauunterschiede von Einzel- und Gruppenverfahren bedarf.

Abstract

Viele psychologisch-diagnostische Testverfahren können laut Manual sowohl als Einzel- sowie auch als Gruppentest angewandt werden. Ob diese zwei Gruppen auch vergleichbare Leistungen erbringen, soll die vorliegende Diplomarbeit klären. Reindl (1998) ging vor mehr als 10 Jahren ebenfalls dieser Frage nach. Die damaligen Ergebnisse sprechen für eine Vergleichbarkeit der Leistungen. In dieser Studie wurden die beiden Gruppen hinsichtlich ihres Niveaus bei einem Leistungstest verglichen, welcher die Dimension *logisch-schlussfolgerndes Denken* erfasst. 111 SchülerInnen zwischen 15 und 19 Jahren flossen in die Berechnung mit ein. Wie schon damals zeigten sich auch heute keine unterschiedlichen Leistungen in der Gruppentestung und in der Einzeltestung. Durch den zweifachen Beleg der Ergebnisse, kann nun davon ausgegangen werden, dass keine weiteren Untersuchungen zu diesem Thema notwendig sind und die Vergleichbarkeit und gemeinsame Normierung von Einzel- und Gruppenvorgabe bei gängigen Leistungstests zulässig sind.

Most of the manuals for psychological tests say that they may be used as an individual or a group test. The focus of this diploma thesis is to find out, if those two groups perform in a comparable manner. More than 10 years ago Mrs. Reindl (1998) discussed the same question and her results showed a comparability of the performances. In this study, the two groups are compared regarding their performance levels in a psychological test, which analyses the intelligence dimension of logical problem-solving. 111 students aged between 15 and 19 participated in the study and their data was assessed. The results are still the same today as they were in the former one, showing no differences between the individual and the group tests. Because of the second time of evidence of the results, more evaluations are not considered necessary. The comparability is proven and the common standardization of individual and group tests regarding current psychological tests is acceptable.

Literatur

Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention*. Berlin: Springer.

Angele, G. (2011). *SPSS Statistics 19: Eine Einführung*. Verfügbar unter <http://www.uni-bamberg.de/fileadmin/uni/service/rechenzentrum/serversysteme/dateien/spss/skript.pdf> [26.7.2011]

Borzt, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer Medizin.

Buchner, A., Erdfelder, E., & Faul, F. (1996). *Teststärkeanalysen*. In E. Erdfelder, R. Mausfeld, T. Meiser & G. Rudinger (Eds.). *Handbuch Quantitative Methoden* (S. 123-136). Weinheim: Psychologie Verlags Union.

Gniech, G. (1976). *Störeffekte in psychologischen Experimenten*. Stuttgart: W. Kohlhammer

Guthke, J. (2003). *Intelligenztest*. In Kubinger K.D. & Jäger R.S. (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 209-215). Weinheim: Beltz.

Heller, K.A. (Hrsg.). (1984). *Leistungsdiagnostik in der Schule*. Bern: Hans Huber

Herkner, W. (2003). *Lehrbuch Sozialpsychologie*. Bern: Hans Huber.

Herle, M. (2003). *Individualtestung*. In Kubinger K.D. & Jäger R.S. (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 205-206). Weinheim: Beltz.

Huber, O. (1987). *Das psychologische Experiment*. Bern: Hans Huber.

Jäger, R.S. (2003). *Pädagogisch-psychologische Diagnostik*. In Kubinger K.D. & Jäger R.S. (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 313-316). Weinheim: Beltz.

Kersting, M., Althoff, K. & Jäger, A.O. (2008). *WIT-2 - Wilde-Intelligenz-Test 2*. Göttingen: Hogrefe.

Kiesel, C. (2009). *Auf welche Persönlichkeitseigenschaften sind Niveauunterschiede bei Einzel- vs. Gruppenvorgabe im Leistungsbereich rückführbar?* Unveröffentlichte Diplomarbeit. Universität Wien.

Kubinger, K.D. & Wurst, E. (2000). *Adaptives Intelligenzdiagnostikum – Version 2.1 (AID 2)*. Göttingen: Beltz.

Kubinger, K.D. (2003). *Gütekriterien*. In Kubinger K.D. & Jäger R.S. (Hrsg.). *Schlüsselbegriffe der psychologischen Diagnostik* (S. 195-204). Weinheim: Beltz.

Kubinger, K.D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.

Langfeldt, H.P. (1984). *Die klassische Testtheorie als Grundlage normorientierter (standardisierter) Schulleistungstests*. In Heller, K.A. (Hrsg.). *Leistungsdiagnostik in der Schule* (S. 65-97). Bern: Hans Huber.

Moosbrugger, H. & Ochlschlägel, J. (1996). *Frankfurter Aufmerksamkeits-Inventar (FAIR)*. Bern: Hans Huber.

Rasch, D., Kubinger, K.D. & Moder, K. (2011). The two-sample *t* test: pre-testing ist assumptions does not pay off. *Statistical Papers*, 2011(1), 219-231.

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

Schneider, K., Wegge, J., & Konradt, U. (1993). *Motivation und Leistung*. In J. Beckmann, H. Strang & E. Hahn (Eds.), *Aufmerksamkeit und Energetisierung* (S. 101–131). Göttingen: Hogrefe.

Steger, R. (1979). *Schichtspezifische Intelligenzleistungen in Gruppen- und Einzeltestungen mit unterschiedlichem Testleiterverhalten*. Unveröffentlichte Diplomarbeit. Universität Wien.

Wagener, D. (2003). *Leistungstest*. In Kubinger K.D. & Jäger R.S. (Hrsg.). Schlüsselbegriffe der psychologischen Diagnostik (S. 264-267). Weinheim: Beltz.

<http://www.testzentrale.de/programm/d2.html>, zuletzt aufgerufen am 8.11.2011

Anhang

Raffaella Neustifter

a0325146@unet.univie.ac.at

Ihr Zeichen,	Unser Zeichen/GZ	Koordinatorin	TEL 525 25	Datum
Ihre Nachricht	240.128/0025-kanz2/2010	Engleder, Claudia MMag. Claudia.engleder@ssr-wien.gv.at	DW 77221 FAX 99-77221	15.09.2010

Sehr geehrte Frau Neustifter!

Der Stadtschulrat für Wien erteilt Ihnen die Bewilligung, im Rahmen Ihrer Diplomarbeit eine Untersuchung zum Thema „Einzel- vs. Gruppenvorgabe: Niveauunterschiede im Leistungsbereich“ vorbehaltlich der Zustimmung der Direktion an den von Ihnen in Ihrem Ansuchen genannten Wiener Schulen

pGRgORg 21 De La Salle Schule Strebersdorf
GRG 21 Bertha v. Suttner – Schulschiff
GRg21/F26 Franklinstraße 26
durchzuführen.

Für die Untersuchung sind die vorgelegten Unterlagen verbindlich. Bei der Mitarbeit von Lehrer/innen muss der Nachweis der Freiwilligkeit vorliegen.

Alle Untersuchungsergebnisse unterliegen der Anonymität und dürfen nur für die wissenschaftliche Arbeit Verwendung finden.

Mit freundlichen Grüßen
für die Amtsführende Präsidentin:

Mag. Margit Auer eh.
Abteilungsleiterin

Raffaella Neustifter
Wagramerstr. 98/6/17
1220 Wien
Tel.: 0660 5694849
a0325146@unet.univie.ac.at

Elternbrief

Liebe Eltern,

im Rahmen meiner Diplomarbeit in der Studienrichtung Psychologie, führe ich in den nächsten Wochen, mit Einverständnis des Stadtschulrates für Wien, der Direktion und der LehrerInnen, in der von Ihrem Kind besuchten Klasse eine Testung durch.

Meine Arbeit beschäftigt sich mit der Frage, ob Leistungsunterschiede zu finden sind, je nachdem ob der Test in einer Gruppe (z.B. Schulklasse) oder einzeln vorgegeben wird. Da es in diesem Bereich noch keine ausreichende Forschung gibt, können die Ergebnisse welche hier gewonnen werden, zur Verbesserung von psychologischen Tests beitragen. Es können wesentliche Beiträge zur Schullaufbahnberatung, Hochbegabendiagnostik, Abhilfe bei Lese- und Rechtschreibproblemen und vielem mehr gegeben werden.

Die einmalige Testung wird im Rahmen des Unterrichts durchgeführt und nimmt somit keine Freizeit in Anspruch.

Die Testergebnisse werden garantiert **ANONYM** behandelt. Die Daten werden ausschließlich im Rahmen meiner Diplomarbeit verwendet und nicht weitergegeben.

Die Einzelleistungen der Kinder sind für mich nicht von Interesse und es ist somit unmöglich zurück zu verfolgen wer welchen Test ausgefüllt hat!

Des Weiteren wird ein publizierter und fundierter Leistungstest verwendet, um Sorgen bezüglich des Tests an sich aus dem Weg zu räumen.

In diesem Sinne ersuche ich sie als Erziehungsberechtigten/Erziehungsberechtigte, Ihr Einverständnis für die Teilnahme Ihres Kindes an der Testung zu geben. Mit jeder Zustimmung ist mir sehr geholfen!

Bei Fragen stehe ich gerne telefonisch oder per mail zur Verfügung.

Vielen Dank für Ihre Hilfe!
Mit freundlichen Grüßen
Raffaella Neustifter

-----Bitte geben Sie diesen Abschnitt ihrem Kind in die Schule mit-----

Mein Sohn/ meine Tochter....., geboren am,
darf an der Testung im Rahmen der Diplomarbeit von Raffaella Neustifter teilnehmen.

.....

Datum, Unterschrift
Raffaella Neustifter
Wagramerstr. 98/6/17
1220 Wien
Tel.: 0660 5694849
a0325146@unet.univie.ac.at

Informationsschreiben an die LehrerInnen

Sehr geehrte Lehrerschaft,

ich arbeite zur Zeit an meiner Diplomarbeit im Fach Psychologie an der Universität Wien bei Univ.-Prof. Dr. Klaus Kubinger im Fachgebiet Diagnostik. Thema der Arbeit ist die Frage, ob die Ergebnisse bei Leitungstests davon abhängig sind, ob eine Testperson einzeln oder in der Gruppe getestet wurde. Anlass für diese Fragestellung ist die möglicherweise unberechtigte Annahme vieler Testautoren, dieser und welcher Leistungstest sei gleichermaßen geeignet ihn einer Einzelperson vor zu geben, genau so wie ihn einer Gruppe vor zu geben. Diesbezüglich gibt es nämlich noch keine ausreichenden Untersuchungen.

Die Ergebnisse dieser Studie können dazu beitragen, psychologische Leistungstests noch genauer zu machen.

Es werden von mir speziell eingeschulte Testleiter die Testung übernehmen.

Bei einem Teil der Kinder werden die Schüler im Klassenverband gemeinsam getestet, dem anderen Teil wird der Test jedem Kind einzeln und separat von den Mitschülern vorgegeben.

Vom Stadtschulrat Wien und der Direktorin/des Direktors habe ich die Einverständnis die Testung an ihrer Schule vor zu nehmen.

Eine einzelne Testung ist in ca. 20 Minuten durchführbar. Die Testunterlagen werden natürlich von mir zur Verfügung gestellt und werden von den Testleitern mitgebracht. Getestet wird mit dem WIT-2 (Wilde Intelligenztest in der 2., revidierten Auflage), sowie einem nachfolgenden Befindensfragebogen.

Die Ergebnisse werden natürlich ANONYM behandelt und es ist nicht nachvollziehbar, wer welchen Test ausgefüllt hat, da es mir nicht um die Leistung eines einzelnen Kindes geht, sondern lediglich um den Unterschied von Gruppen- und Einzeltestung.

Das Einverständnis der Eltern wird natürlich eingeholt und ist mittels mit zu bringenden Abschnitts mit Unterschrift zu bestätigen.

Für weitere Fragen stehe ich gerne telefonisch oder per mail zur Verfügung!

Vielen Dank für Ihre Mithilfe!
Mit freundlichen Grüßen

Raffaella Neustifter

Einzeltestung

Mein Geschlecht: (bitte ankreuzen):

männlich ☐

weiblich ☐

Mein Alter: _____Jahre

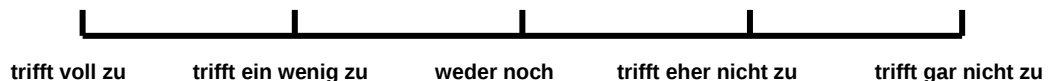
Wie ist es dir während DIESER Testung gegangen?

Bitte beantworte die Fragen **EHRlich** und kreuze so an, dass es auf **DICH** am ehesten zutrifft. Es gibt kein Richtig und Falsch! 🤔

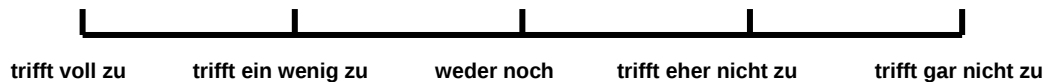
Bsp.: Heute fühle ich mich gut



1. Es war mir unangenehm, dass ich mit dem Testleiter alleine in einem Raum war.



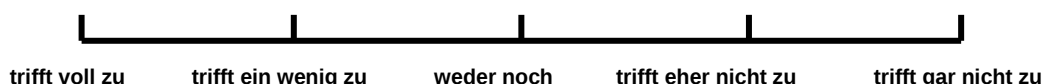
2. Ich kann mich wenn ich alleine bin nicht so gut konzentrieren. In einer Gruppe würde mir die Testung leichter fallen.



3. Der Test war mir egal.



4. Alleine habe ich mich wohlgeföhlt, weil so meine Mitschüler nicht gesehen haben, was ich alles ausgefüllt habe und wie schnell ich war.



5. *Ich mag Aktivitäten mit meinen Klassenkollegen und –kolleginnen.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

6. *Ich konnte schummeln.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

7. *Es gab Störeinflüsse (Lärm usw.), die mich abgelenkt haben.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

8. *Ich habe mich im Allgemeinen während der Testung wohl gefühlt.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

9. *Der Testleiter war mir sympathisch.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

10. *Ich hatte das Gefühl, dass sich der Testleiter mit mir besonders beschäftigt hat.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

11. *Ich arbeite gerne alleine, weil mich so niemand stören kann.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

VIELEN DANK FÜR DEINE MITARBEIT!!

Testung in der Gruppe

Mein Geschlecht: (bitte ankreuzen):

männlich ☐

weiblich ☐

Mein Alter: _____Jahre

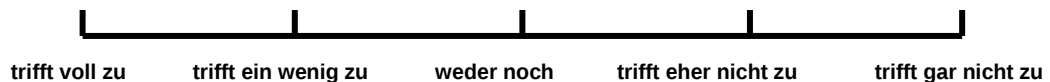
Wie ist es dir während DIESER Testung gegangen?

Bitte beantworte die Fragen EHRlich und kreuze so an, dass es auf DICH am ehesten zutrifft. Es gibt kein Richtig und Falsch! 🍌

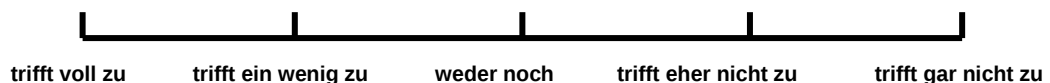
Bsp.: Heute fühle ich mich gut



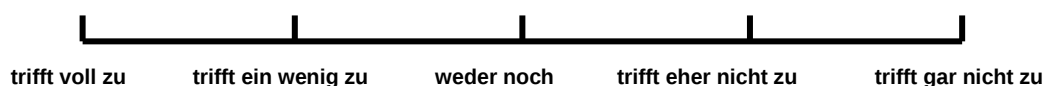
1. Ich habe mich von anderen Mitschülern gestört gefühlt.



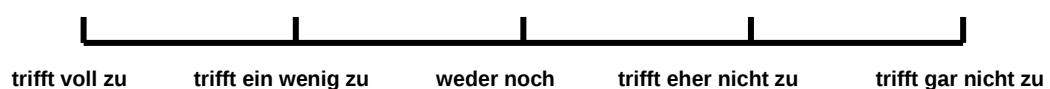
2. Ich kann mich in der Gruppe nicht so gut konzentrieren. Wenn ich alleine gewesen wäre, wäre mir die Testung leichter gefallen.



3. Der Test war mir egal.



4. In der Gruppe habe ich mich wohlgefühlt, weil man nicht so auffällt.



5. *Ich mag Aktivitäten mit meinen Klassenkollegen und –kolleginnen.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

6. *Ich habe bei jemand anderem abgeschrieben.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

7. *Es gab Störeinflüsse (Lärm, Gespräche der Mitschüler usw.), die mich abgelenkt haben.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

8. *Ich habe mich im Allgemeinen während der Testung wohl gefühlt.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

9. *Ich war froh, dass ich mit dem Testleiter nicht alleine sein musste.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

10. *Ich hatte das Gefühl, dass sich der Testleiter mit mir besonders beschäftigt hat.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

11. *Das Arbeiten in der Gruppe regt mich zu besseren Leistungen an.*

trifft voll zu trifft ein wenig zu weder noch trifft eher nicht zu trifft gar nicht zu

VIELEN DANK FÜR DEINE MITARBEIT!!

Zusammenfassende und interessante Infos über die Untersuchung

Ziel der Untersuchung:

Es soll gezeigt werden, dass Kinder die in der Gruppe getestet werden **bessere Leistungen** erzielen als Kinder unter der Einzeltestbedingung

Grund für die Annahme:

Es gibt **verschiedene Quellen** die das Ergebnis schon im Vorfeld dieser Untersuchung angedeutet haben:

- Frühere Untersuchungen belegen die Hypothese (neueste aus dem Jahr 2000)
- Prof. Kubinger schließt sich dieser Meinung an und geht davon aus, dass sich diese Ergebnisse auch jetzt noch replizieren lassen
- Diplomarbeit von A. Schrott aus dem Jahr 1998 fiel ebenso in dieser Richtung aus

In diesem Sinne wünsche ich mir, dass auch meine Ergebnisse so ausfallen!

Bei Fragen dazu, kannst du mich gerne jederzeit kontaktieren.

Vielen Dank für deine Unterstützung

Raffaela N.

Zusammenfassende und interessante Infos über die Untersuchung

Ziel der Untersuchung:

Es soll gezeigt werden, dass Kinder die **alleine getestet** werden **bessere Leistungen** erzielen als Kinder in der Gruppentestung.

Grund für die Annahme:

Es gibt **verschiedene Quellen** die das Ergebnis schon im Vorfeld dieser Untersuchung angedeutet haben:

- Frühere Untersuchungen belegen die Hypothese (neueste aus dem Jahr 2000)
- Prof. Kubinger schließt sich dieser Meinung an und geht davon aus, dass sich diese Ergebnisse auch jetzt noch replizieren lassen
- Diplomarbeit von A. Schrott aus dem Jahr 1998 fiel ebenso in dieser Richtung aus

In diesem Sinne wünsche ich mir, dass auch meine Ergebnisse so ausfallen!

Bei Fragen dazu, kannst du mich gerne jederzeit kontaktieren.

Vielen Dank für deine Unterstützung

Raffaela N.

Kreuztabellen - Fragebogenauswertung

Vorgabemodus * E1 G9 Kreuztabelle

Anzahl		E1 G9			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	22	4	7	33
	Gruppentestung	33	28	17	78
Gesamt		55	32	24	111

Vorgabemodus * E2 G2 Kreuztabelle

Anzahl		E2 G2			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	19	5	9	33
	Gruppentestung	49	15	14	78
Gesamt		68	20	23	111

Vorgabemodus * E3 G3 Kreuztabelle

Anzahl		E3 G3			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	15	8	10	33
	Gruppentestung	26	27	25	78
Gesamt		41	35	35	111

Vorgabemodus * E4 G4 Kreuztabelle

Anzahl		E4 G4			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	19	9	5	33
	Gruppentestung	26	25	27	78
Gesamt		45	34	32	111

Vorgabemodus * E5 G5 Kreuztabelle

Anzahl		E5 G5			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	4	27	2	33
	Gruppentestung	6	68	4	78

Vorgabemodus * E5 G5 Kreuztabelle

Anzahl

		E5 G5			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	4	27	2	33
	Gruppentestung	6	68	4	78
Gesamt		10	95	6	111

Vorgabemodus * E6 G6 Kreuztabelle

Anzahl

		E6 G6		Gesamt
		stimme nicht zu	weder noch	
Vorgabemodus	Einzeltestung	32	1	33
	Gruppentestung	75	3	78
Gesamt		107	4	111

Vorgabemodus * E7 G7 Kreuztabelle

Anzahl

		E7 G7			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	26	5	2	33
	Gruppentestung	62	12	4	78
Gesamt		88	17	6	111

Vorgabemodus * E8 G8 Kreuztabelle

Anzahl

		E8 G8			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	5	17	11	33
	Gruppentestung	10	51	17	78
Gesamt		15	68	28	111

Vorgabemodus * E10 G10 Kreuztabelle

Anzahl

		E10 G10			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	5	13	15	33
	Gruppentestung	66	4	8	78
Gesamt		71	17	23	111

Vorgabemodus * E11 G11 Kreuztabelle

Anzahl

		E11 G11			Gesamt
		stimme nicht zu	stimme zu	weder noch	
Vorgabemodus	Einzeltestung	7	18	8	33
	Gruppentestung	24	22	32	78
Gesamt		31	40	40	111

Welch-Test

ONEWAY deskriptive Statistiken

twert

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
Einzeltestung	33	51,06	9,906	1,724	47,55	54,57	32	70
Gruppentestung	78	49,62	8,739	,989	47,65	51,59	32	64
Gesamt	111	50,05	9,080	,862	48,34	51,75	32	70

Robuste Testverfahren zur Prüfung auf Gleichheit der Mittelwerte

twert

	Statistik ^a	df1	df2	Sig.
Welch-Test	,528	1	54,106	,470

a. Asymptotisch F-verteilt

Kolmogorov-Smirnov-Test bei zwei Stichproben

Häufigkeiten

Vorgabemodus		N
twert	Einzeltestung	33
	Gruppentestung	78
	Gesamt	111

Statistik für Test^a

		twert
Extremste Differenzen	Absolut	,121
	Positiv	,121
	Negativ	-,027
Kolmogorov-Smirnov-Z		,584
Asymptotische Signifikanz (2-seitig)		,885

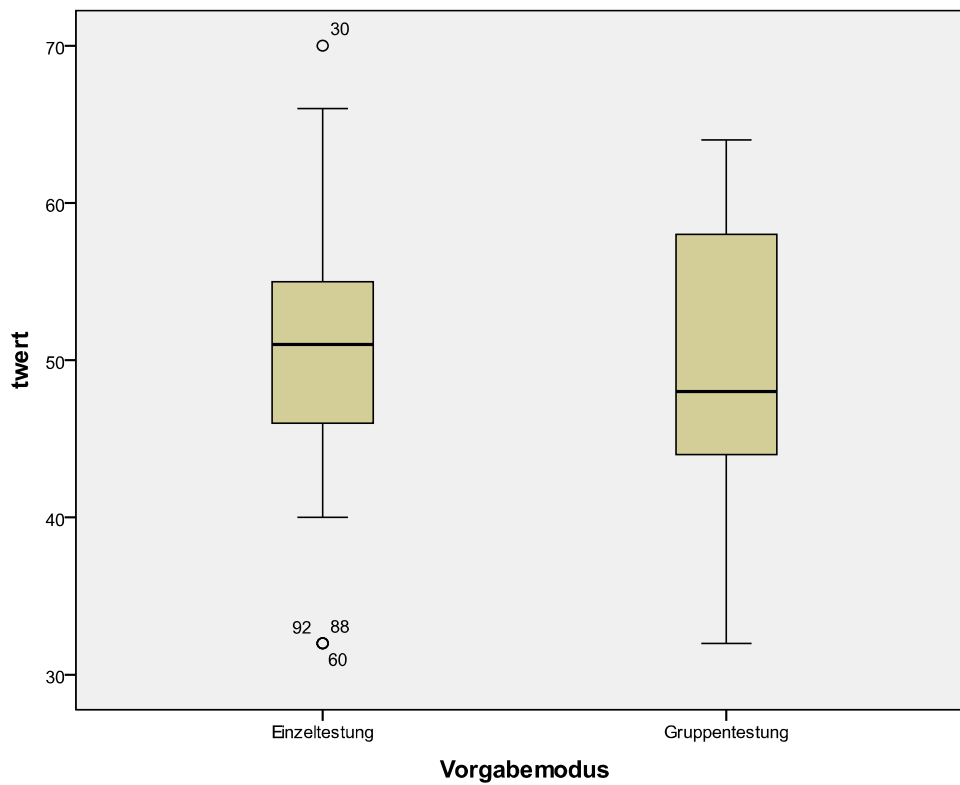
a. Gruppenvariable: Vorgabemodus

```
EXAMINE VARIABLES=twert BY modus
/PLOT=BOXPLOT
/STATISTICS=NONE
/NOTOTAL.
```

Explorative Datenanalyse

Verarbeitete Fälle

Vorgabemodus		Fälle					
		Gültig		Fehlend		Gesamt	
		N	Prozent	N	Prozent	N	Prozent
twert	Einzeltestung	33	100,0%	0	,0%	33	100,0%
	Gruppentestung	78	100,0%	0	,0%	78	100,0%



Univariate Auswertung

Test der Homogenität der Varianzen

twert

Levene-Statistik	df1	df2	Signifikanz
,800	3	107	,497

Post-Hoc-Tests

Mehrfachvergleiche

twert
Scheffé-Prozedur

(I) Testleiter	(J) Testleiter	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
Testleiter 1	Testleiter 2	2,672	2,439	,753	-4,26	9,60
	Testleiter 3	1,196	2,201	,961	-5,06	7,45
	Testleiter 4	7,068*	2,289	,027	,57	13,57
Testleiter 2	Testleiter 1	-2,672	2,439	,753	-9,60	4,26
	Testleiter 3	-1,476	2,500	,950	-8,58	5,63
	Testleiter 4	4,396	2,578	,410	-2,93	11,72
Testleiter 3	Testleiter 1	-1,196	2,201	,961	-7,45	5,06
	Testleiter 2	1,476	2,500	,950	-5,63	8,58
	Testleiter 4	5,872	2,355	,108	-,82	12,56
Testleiter 4	Testleiter 1	-7,068*	2,289	,027	-13,57	-,57
	Testleiter 2	-4,396	2,578	,410	-11,72	2,93
	Testleiter 3	-5,872	2,355	,108	-12,56	,82

*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

twert

Scheffé-Prozedur^{a, b}

Testleiter	N	Untergruppe für Alpha = 0.05.	
		1	2
Testleiter 4	26	45,46	
Testleiter 2	21	49,86	49,86
Testleiter 3	30	51,33	51,33
Testleiter 1	34		52,53
Signifikanz		,118	,743

Die Mittelwerte für die in homogenen Untergruppen befindlichen Gruppen werden angezeigt.

a. Verwendet ein harmonisches Mittel für Stichprobengröße = 26,877.

b. Die Gruppengrößen sind nicht identisch. Es wird das harmonische Mittel der Gruppengrößen verwendet. Fehlerniveaus des Typs I sind nicht garantiert.

Kolmogorov-Smirnov-Anpassungstest

		Standardisierte Residuen für tWert
N		111
Parameter der Normalverteilung ^{a, b}	Mittelwert	,0000
	Standardabweichung	,98627
Extremste Differenzen	Absolut	,068
	Positiv	,068
	Negativ	-,049
Kolmogorov-Smirnov-Z		,719
Asymptotische Signifikanz (2-seitig)		,679

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Deskriptive Statistiken

Abhängige Variable: tWert

Testleiter	Vorgabemodus	Mittelwert	Standardabweichung	N
Testleiter 1	Einzeltestung	54,22	9,121	9
	Gruppentestung	51,92	8,981	25
	Gesamt	52,53	8,938	34
Testleiter 2	Einzeltestung	48,22	5,426	9
	Gruppentestung	51,08	7,902	12
	Gesamt	49,86	6,945	21
Testleiter 3	Einzeltestung	53,00	11,747	8
	Gruppentestung	50,73	7,542	22
	Gesamt	51,33	8,691	30
Testleiter 4	Einzeltestung	48,43	13,113	7
	Gruppentestung	44,37	8,674	19
	Gesamt	45,46	9,941	26
Gesamt	Einzeltestung	51,06	9,906	33
	Gruppentestung	49,62	8,739	78
	Gesamt	50,05	9,080	111

Levene-Test auf Gleichheit der Fehlervarianzen^a

Abhängige Variable: tWert

F	df1	df2	Sig.
1,152	7	103	,337

Prüft die Nullhypothese, daß die Fehlervarianz der abhängigen Variablen über Gruppen hinweg gleich ist.

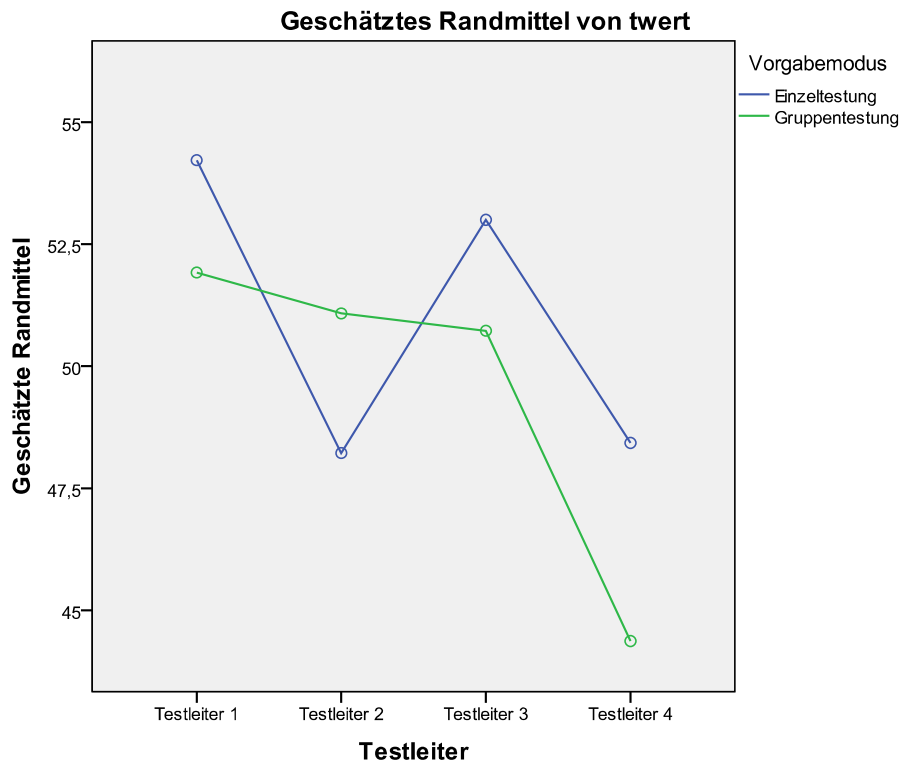
a. Design: Konstanter Term + tl + modus + tl * modus

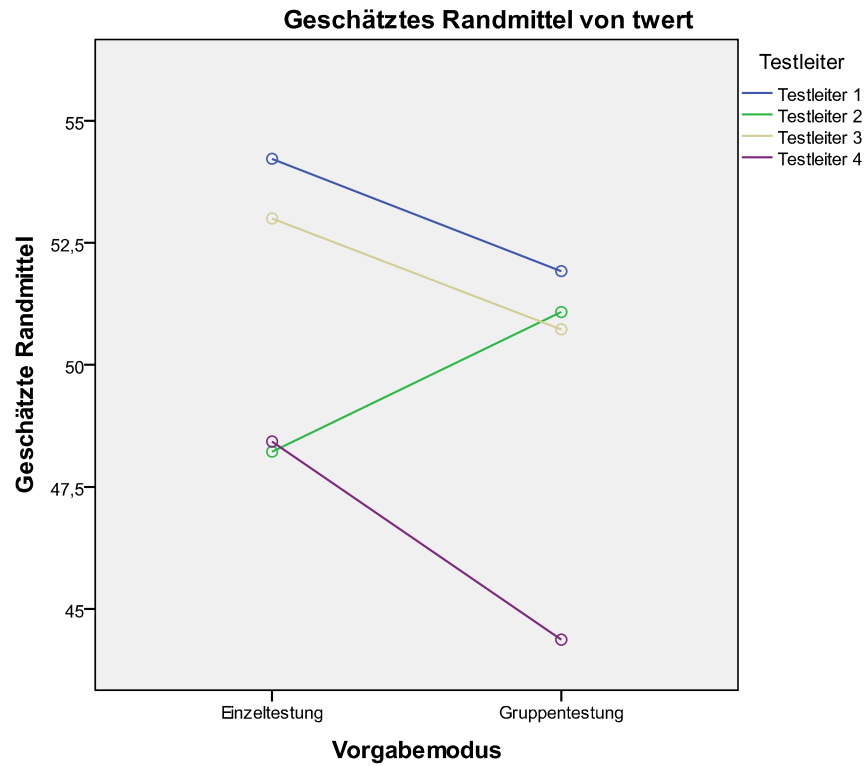
Tests der Zwischensubjekteffekte

Abhängige Variable:twert

Quelle	Quadratsumme vom Typ I	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	998,408 ^a	7	142,630	1,820	,091
Konstanter Term	278000,225	1	278000,225	3548,045	,000
tl	806,605	3	268,868	3,431	,020
modus	52,711	1	52,711	,673	,414
tl * modus	139,092	3	46,364	,592	,622
Fehler	8070,367	103	78,353		
Gesamt	287069,000	111			
Korrigierte Gesamtvariation	9068,775	110			

a. R-Quadrat = ,110 (korrigiertes R-Quadrat = ,050)





Kolmogorov-Smirnov-Anpassungstest

		Standardisierte Residuen für twert tl und modus
N		111
Parameter der Normalverteilung ^{a, b}	Mittelwert	,0000
	Standardabweichung	,96766
Extremste Differenzen	Absolut	,085
	Positiv	,064
	Negativ	-,085
Kolmogorov-Smirnov-Z		,894
Asymptotische Signifikanz (2-seitig)		,401

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Persönliche Daten

Raffaella Neustifter
a0325146@unet.univie.ac.at
09.10.1984, Wien

Staatsbürgerschaft: Österreich

Ausbildung

1990-1994 Volksschule „Schulbrüder Strebersdorf“, Wien

1994-1995 Gymnasium „Schulbrüder Strebersdorf“, Wien

1995-2003 Gymnasium „Schulschiff“, Wien

2003-2004 1 Semester Studium der technischen Physik an der
TU Wien

2004-dato Studium der Psychologie an der Uni Wien (2007
Abschluss des 1. Abschnittes; 2012 voraussichtlicher
Abschluss des Studiums)

Berufliche Tätigkeiten

2001 sowie in den Folgejahren 2002 und 2003
Bank Austria
Ferialjob als Filialmitarbeiterin

2003 **H&M** (www.hm.com)
Ferialjob im Verkauf (Filiale 116)

2004-dato **H&M** (www.hm.com)
geringfügig Beschäftigt im Verkauf (Filiale 116)

2007-dato **INFAR** – Verkehrspsychologisches Institut
(www.infar.at)
Mitarbeit bei Projekten, Aushilfe im Büro,
administrative Tätigkeit, verkehrspsychologische
Diagnostik, Statistische Auswertungen, Verfassen
von Gutachten
2007 Absolvierung des Pflichtpraktikums

2009

1-monatiges freiwilliges Praktikum auf der
akutpsychiatrischen Station des SMZO (Einzel- und
Gruppengespräche)
(*Zeugnis beiliegend*)

Besondere Kenntnisse

Fremdsprachen

Englisch – fließend in Wort und Schrift
Französisch – Maturaniveau

fachbezogen

SPSS, sämtliche diagnostische
Verfahren (Leistungstests, AC,
Fragebögen, Anamnesegespräche),
verfassen von Gutachten,
Wahlfachspezialisierung: Psychologische
Diagnostik und Wirtschaftspsychologie