



**universität
wien**

**Vergleich einer revidierten Form des Wiener Matrizen-Tests
(WMT) in Berlin und Österreich**

Diplomarbeit

**eingereicht von
Natascha Marakovits**

**zur Erlangung des
Magistrgrades der Naturwissenschaften (Mag. rer. nat.)**

Wien, November 2008

**Studienkennzahl: 298
Studienrichtung: Psychologie
Betreuerin: Dr. Karin Waldherr**

In Liebe und Dankbarkeit
für
meine Eltern

Inhaltsverzeichnis

Vorwort	5
1. Das Problem der Kulturfreiheit von Intelligenztests	7
2. Die Entstehungsgeschichte des Wiener Matrizen-Tests	9
2.1 Der Matrizen test von Raven	9
2.2 Vorläufer des WMT: Formanns averbaler Matrizen test (FAM)	9
2.3 Die erste Version des WMT	15
2.4 Die 2., revidierte Fassung des WMT (WMT-2)	18
3. Bisherige Untersuchungen zum WMT	19
3.1 Dimensionalitätsanalysen des WMT in Verbindung mit anderen Rasch-skalierten Tests	19
3.2 Bisherige Untersuchungsergebnisse zum interkulturellen Vergleich mittels WMT	21
3.3 Weitere Untersuchungen	23
3.4 Zusammenfassung	25
4. Ziele und Hypothesen der Untersuchung	26
5. Planung der Durchführung der Untersuchung	27
6. Unterschiede der Schulsysteme Deutschland – Österreich	29
6.1 Schulformen	29
6.2 Notensystem	31
6.3 Punktesystem	32
7. Datenerhebung	33
8. Stichprobe	34
8.1 Geschlecht	36
8.2 Muttersprache	36

8.3 Alter	37
8.4 Schultyp	37
8.5 Mathematiknote	38
8.6 Darstellende Geometrie und Geometrisch Zeichnen	39
8.7 Bearbeitungsdauer	40
9. Rasch-Modellprüfung	41
9.1 Teilungskriterien	41
9.2 Zusammenfassung der Ergebnisse der Modelltests Berlin	45
9.3 Die Durchführung der Alpha-Adjustierung	45
9.4 Grafische Modellkontrolle Berlin	48
9.5 Ergebnisse der Modelltests Berlin – Österreich	54
9.6 Grafische Modellkontrolle Berlin – Österreich	55
9.7 Zusammenfassung der Modelltests	58
9.8 Dimensionalitätsprüfung nach Martin – Löff	60
10. Mittelwertsvergleiche mit der österreichischen Normierungsstichprobe	62
10.1 Ergebnisse	64
11. Normierung	65
12. Testgütekriterien	66
12.1 Objektivität	66
12.2 Reliabilität	66
12.3 Validität	67
13. Zusammenfassung	70
14. Literaturverzeichnis	73
15. Anhang	76
15.1 Schulenverzeichnis	77
15.2 Antwortblatt	78

15.3 Normtabellen	79
15.4 Ergebnisse der Rasch-Modelltests	82
15.5 Ergebnisse der Modelltests nach Martin-Löf	96
15.6 Ergebnisse der Varianzanalyse	98
Lebenslauf	99

Vorwort

Der Wiener Matrizen-Test (WMT, Formann & Piswanger, 1979) ist ein Test zur sprachfreien Erfassung der Intelligenz. Aufgrund der sprachfreien Testaufgaben gilt er als kulturunabhängig („culture fair“) und wird nun seit beinahe 30 Jahren in verschiedenen Bereichen der Psychologie (Schulpsychologie, Betriebspsychologie) eingesetzt.

Aufgrund intensiver Forschungen zum Rasch-Modell und sozialer sowie bildungspolitischer Veränderungen in den letzten Jahren, erschien es sinnvoll, eine Neunormierung des WMT vorzunehmen, sowie die Gültigkeit des probabilistischen Testmodells von Rasch (1960) mittels neuester Software zu überprüfen.

Das Ergebnis ist der WMT-2 mit 18 Items, sowie Normen einer österreichischen Stichprobe mit 2494 Personen im Alter von 14 bis 64 Jahren (Formann, Waldherr & Piswanger, in Druck).

Wie bereits erwähnt handelt es sich beim WMT um einen kulturfreien Test. Beginnen möchte ich meine Arbeit daher mit dem allgemeinen Problem der Kulturfreiheit von Intelligenztests. Anschließend möchte ich einen Überblick über die Entstehungsgeschichte und die bisherigen Untersuchungen zum WMT geben.

Aufgrund der Eliminierung von sechs Items des WMT für den WMT-2, erschien es mir sinnvoll zu klären, ob beim WMT-2 Rasch-Homogenität bei einer ausländischen Stichprobe gegeben ist.

Da sich bisherige Untersuchungen zum WMT auf Kulturen beschränkten, die im Vergleich zu Österreich sehr unterschiedlich sind, habe ich mich entschieden, für meine Untersuchung eine Stichprobe zu wählen, die Österreich sehr ähnlich ist. Bei der Stichprobe dieser Arbeit handelt es sich daher um Berliner SchülerInnen aus Gesamt- und Berufsschulen. Ursprünglich sollte die Stichprobe auch Gymnasien beinhalten, dies war jedoch leider aufgrund fehlender Kooperation von Seiten der Schulleiter nicht möglich.

Die Rasch-Homogenität des WMT-2 sollte dabei anhand verschiedener Modelltests mit dem Programm LPCM-Win (Fischer & Ponocny-Seliger, 1998) überprüft werden.

In weiterer Folge sollte eine Überprüfung der Eindimensionalität der Items mittels Modelltests nach Martin – Löff erfolgen.

Ein weiteres Ziel meiner Arbeit war es, die Testleistungen der Berliner Stichprobe mit den Testleistungen der österreichischen Normstichprobe zu vergleichen und anschließend eine Normierung der Berliner Stichprobe vorzunehmen. Eine Normierung erschien sinnvoll, da der WMT-2 ebenso wie der WMT international eingesetzt werden soll und daher eine Erweiterung der Vergleichsstichproben nützlich ist.

1. Das Problem der Kulturfreiheit von Intelligenztests

Der WMT hat im Gegensatz zu vielen anderen Tests den Vorteil, dass er aus grafischen und symbolischen Elementen besteht und somit sprachfrei gestaltet ist. Daher besteht ein wesentliches Merkmal darin, dass er kulturgerecht („culture fair“) bzw. kulturfrei („culture free“) ist.

Somit kann davon ausgegangen werden, dass es hinsichtlich der Ergebnisse auch in unterschiedlichen Kulturen keine Unterschiede gibt. Es zeigt sich nämlich, dass bei nicht-sprachfreien Intelligenztests die Testpersonen aus jener Kultur bessere Ergebnisse erzielen, in welcher der Test entwickelt wurde (Piswanger, 1975). Das ist darauf zurückzuführen, dass es bei nicht-sprachfreien Tests unausweichlich ist, dass sie kulturspezifisches Wissen, Reaktionen und Verhaltensweisen messen. Somit kann nicht davon ausgegangen, dass derselbe Test in unterschiedlichen Kulturen zu denselben Ergebnissen führt. Catell spricht in diesem Zusammenhang von "crystallized intelligence" oder vom „General Crystallized Ability Faktor“ (Catell, 1971). Laut Catell sind dies jene kognitiven Fähigkeiten, in denen sich angehäuften Wissen aus bisherigen Lernprozessen kristallisiert und verfestigt hat. Diese Intelligenz ist gewissermaßen das Endprodukt dessen, was flüssige Intelligenz und Bildung und Ausbildung gemeinsam hervorgebracht haben. Dieser Faktor beinhaltet in hohem Maße kulturspezifische Elemente.

Um zu verhindern, dass bei interkulturellen Untersuchungen die Ergebnisse durch kulturspezifische Elemente im Test beeinflusst werden, sollten sprachfreie Tests angewendet werden. Diese versuchen durch die Verwendung von grafischen oder symbolischen Elementen die sogenannte „fluid general intelligence“ (Catell, 1971) zu messen. Catell beschreibt diesen „General Fluid Ability Faktor“ als allgemeine, weitgehend angeborene Leistungsfähigkeit. Sie spiegelt die Fähigkeit wieder, sich neuen Problemen und Situationen anzupassen, ohne dass dazu frühere Lernerfahrungen notwendig sind. Fluidität zeigt sich in Verhaltensweisen, die ohne Zögern das Unwichtige vom Wichtigen trennen, viele Zusammenhänge zwischen Informationen "auf einen Schlag erfassen" und ordnen. Somit ist „fluide Intelligenz“ vor allem durch „Schlussfolgerndes Denken“ gekennzeichnet. Eben diese Fähigkeiten und Verhaltensweisen lassen sich relativ kulturfrei („culture-fair“) erfassen.

Bei sprachfreien Tests wird also versucht, solche Darstellungen zu verwenden, die allen Kulturen möglichst gleich gut vertraut sind. Das „Schlussfolgernde Denken“ soll anhand symbolischer Elemente gemessen werden, zwischen denen bestimmte Zusammenhänge erkannt und fehlende Symbole ergänzt werden sollen.

Zur Aufgabenlösung sind somit keine sprachlichen Formulierungen nötig, sodass soziale Faktoren wie kultureller Background, Schulbildung und die persönliche und sprachliche Entwicklung nicht von Bedeutung sind.

Bereits in den 20er Jahren verwendeten Spearman's Schüler Feingold, Sarason und Catell kulturunabhängige Tests und konnten in ihren Arbeiten später auch zeigen, dass sie sehr gut mit allgemeinen Intelligenzschätzungen und dem allgemeinen Intelligenzfaktor korrelieren und dass die Testkennwerte somit nicht von kulturellen Faktoren beeinflusst werden. Catell war jedoch der Erste, der versuchte, komplexe Beziehungen zwischen „Grundelementen“ herzustellen, die alle Menschen kennen und die nicht in verbaler Form, sondern in Bildern ausgedrückt werden.

Die kulturfreien Tests haben sich seither in vielen Ländern wie China, Afrika und Indien als brauchbar erwiesen.

Kulturfreie Tests messen somit gemäß Cattells Theorie den Faktor der „flüssigen“ allgemeinen Begabung, der sich zwar von der „kristallierten“ allgemeinen Begabung unterscheidet, jedoch mit ihr korreliert (Catell, 1971).

2. Die Entstehungsgeschichte des Wiener Matrizen-Tests (WMT)

In diesem Abschnitt sollen die Konstruktionsprinzipien beschrieben werden, auf dessen Grundlage der WMT entwickelt wurde. Weiters sollen die Methoden, mit denen er überprüft wurde erklärt werden und letztendlich soll auf die Verbreitung als Test eingegangen werden.

2.1 Der Matrizentest von Raven

Der Matrizentest von Raven, der die Grundlage für den WMT bildete, ist einer der bekanntesten nonverbalen Intelligenztests. Der Test hat zwei Ausgaben: die Form von 1938 in der Revision von 1956 (Standard Progressive Matrices) und die Form von 1947 in der Revision von 1956 (farbige Sets A und B der Standard Progressive Matrices = coloured Progressive Matrices).

Die Standard Form enthält 60 Items („matrices“), geordnet in 5 Gruppen (A, B, C, D, E) mit je 12 prinzipiell gleichartigen Aufgaben und kann für Erwachsene im Alter bis 65 Jahren und für Kinder von 6-11 Jahren verwendet werden. Jedes Item besteht aus geometrischen Formen, mit einer fehlenden Figur. Die Versuchsperson muss die bestehenden Zusammenhänge herausfinden und danach die fehlende Figur aus den Lösungsvorschlägen auswählen. Die Schwierigkeit der Aufgaben nimmt dabei von Gruppe A nach Gruppe E zu.

Die Coloured Progressive Matrices können für Kinder von 5-11 Jahren und für Erwachsene mit niedriger geistiger Leistungsfähigkeit verwendet werden (Raven, 1960).

2.2 Vorläufer des WMT: Formanns averbaler Matrizentest (FAM)

In seiner Dissertation von 1973 versuchte Formann auf Grundlage der Matrizentestaufgaben von Raven (1938) einen neuen Matrizentest zu konstruieren. Es sollten dabei bestimmte formale Regeln berücksichtigt werden, wobei er sich mit folgenden testtheoretischen Fragen auseinandersetzte:

1. Itemuniversum

Wie und wodurch kann in Bezug auf einen Matrizentest ein Itemuniversum definiert werden? Nach welchem Konstruktionsprinzip soll vorgegangen werden?

2. Dimensionalität

Misst ein Test, der nach festgelegten Regeln konstruiert wurde, tatsächlich eindimensional, wie es aufgrund der einheitlich gebildeten Items zu erwarten wäre, d. h. erfasst er genau eine latente Eigenschaft?

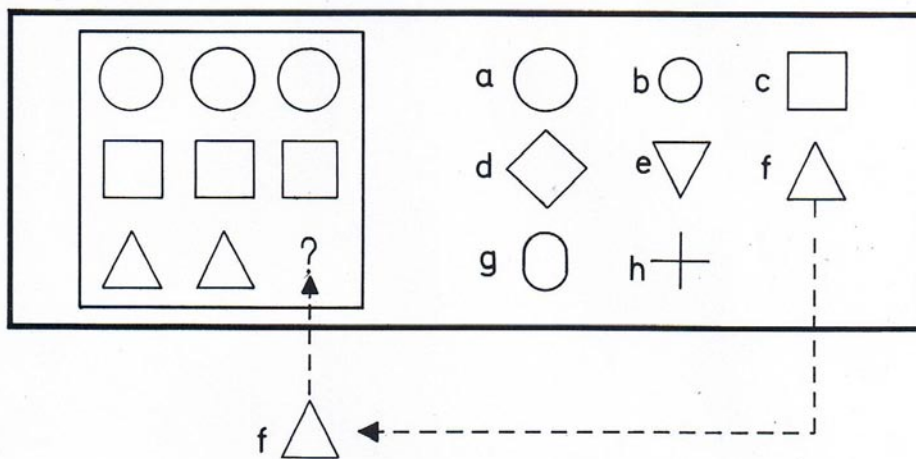
3. Lösungsverhalten

Hier stellt Formann die Frage „ob für ein konkretes Item die gemäß den postulierten Konstruktionsprinzipien geltenden Gesetzmäßigkeiten dieselben sind, wie jene, welche die Personen bei ihrer Lösung beachten,...“ (Formann, 1973, S. 2).

Das heißt, ist der Lösungsweg nur mit den Aufbauregeln des Items verknüpft oder gibt es noch andere Einflüsse, die nicht erfasst werden und in welchem Zusammenhang stehen die beachteten „Einzeloperationen“ zueinander?

Aufgrund dieser Fragen wurden Aufgaben entwickelt, die folgendermaßen aussehen: Die Aufgaben sind zweigeteilt: Links befindet sich ein unvollständiges Neun-Figureschema. Das Fragezeichen soll durch eine der rechts abgebildeten Antwortmöglichkeiten a bis h sinnvoll ergänzt werden.

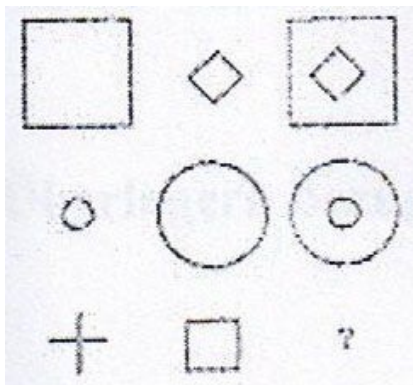
Abb. 1: Einführungsbeispiel aus dem WMT (Formann & Piswanger, 1979)



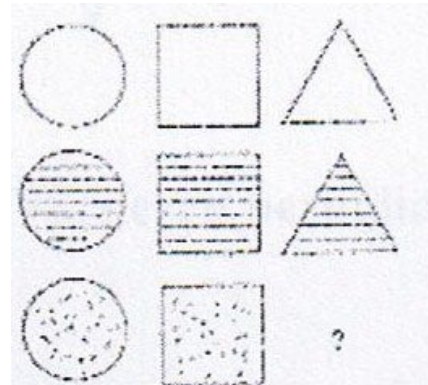
Die Figuren sind aus bestimmten Elementen aufgebaut, die Komponenten genannt werden. Jede Komponente einer Figur kann aufgrund der Materialeigenschaften Form, Muster, Anzahl oder räumliche Anordnung beschrieben werden. Die Beschreibungsmerkmale der Materialeigenschaften werden Repräsentanten genannt. Die Form wird durch eine geometrische Figur bestimmt, zum Beispiel: Dreiecke, Vierecke, Kreise, usw.. Weiters unterscheiden sich die Figuren hinsichtlich ihres Musters. Dieses kann weiß, gestreift usw. sein. Die Komponente Anzahl wird durch die Zahl der verschiedenen Formen repräsentiert und die Komponente räumliche Anordnung unterscheidet sich z. B. durch die Richtung, in die ein Pfeil zeigt, oder oben – unten oder links – rechts Anordnungen. Beispiele dazu sind in Abbildung 2 dargestellt.

Abbildung 2: Beispiele für Materialeigenschaften (Formann, 1973, S. 14ff)

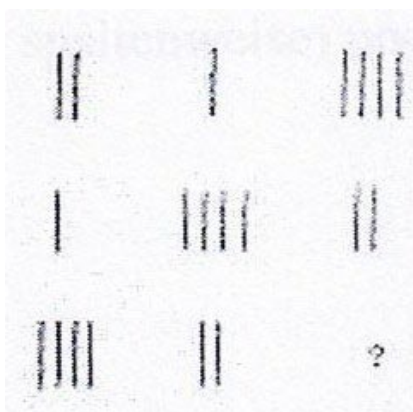
unterschiedliche Formen



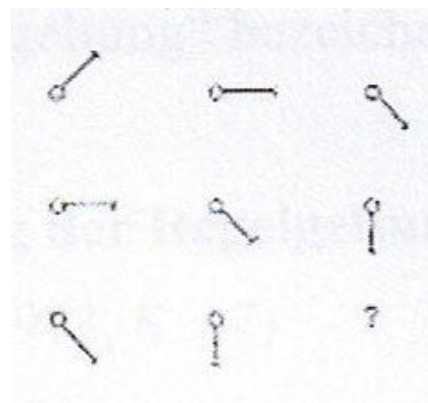
unterschiedliche Muster und Formen



unterschiedliche Anzahl



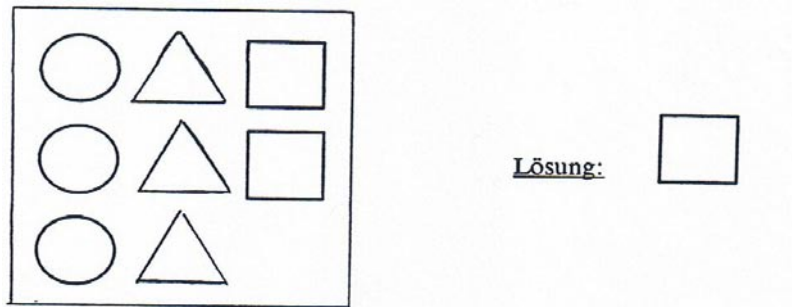
unterschiedliche räumliche Anordnung



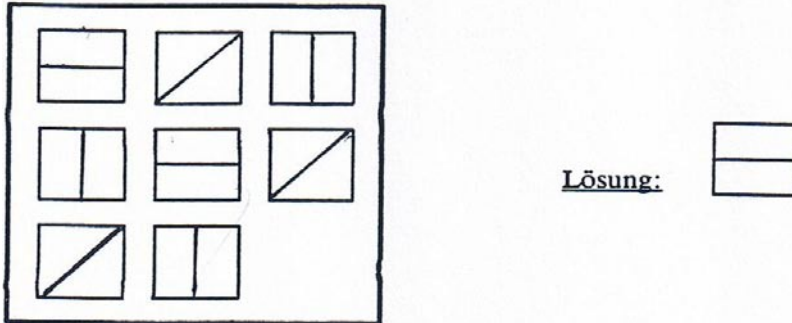
Um auf die fehlende Figur schließen zu können, muss der logische Zusammenhang, der zwischen den Figuren besteht, herausgefunden werden. Nach Formann werden die Regeln Fortsetzen, Variieren und Überlagern unterschieden. Diese sind in Abbildung 3 dargestellt.

Abb. 3: Regeln (Mohammadzadeh Koucheri, 1993, S.11):

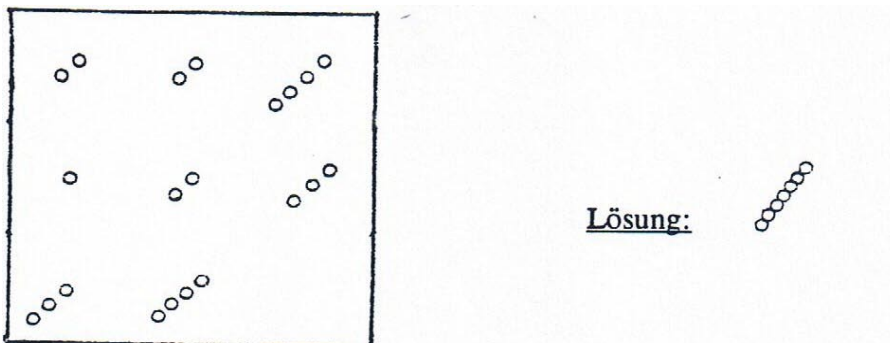
Fortsetzen: z.B. Form fortsetzen



Variieren: z.B. räumliche Anordnung variieren



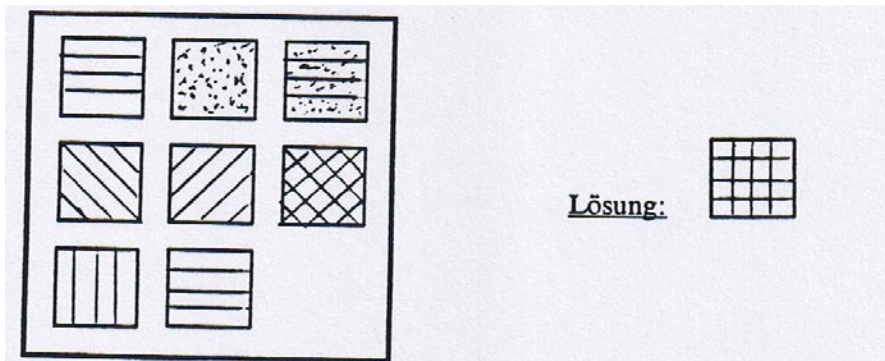
Überlagern: z.B. Anzahl überlagern



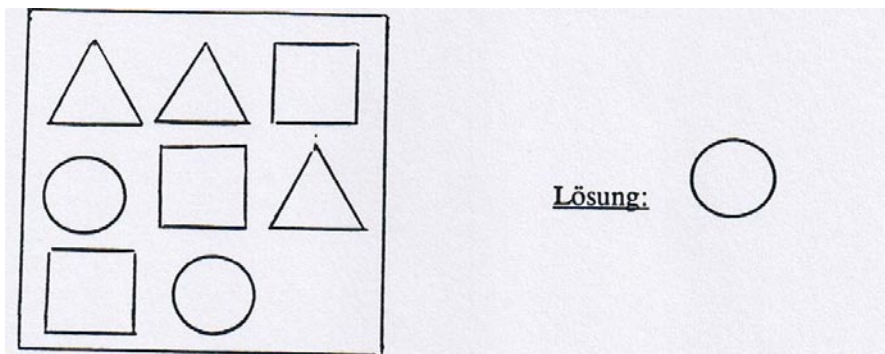
Weiters werden nach Formann folgende Richtungen der Regelgeltung unterschieden: waagrecht, senkrecht oder waag- und senkrecht. Hier wird die Regel entweder nur auf die Zeilen, nur auf die Spalten oder auf beides angewandt. In Abbildung 4 sind die unterschiedlichen Richtungen der Regelgeltung dargestellt.

Abb. 4: Unterschiedliche Richtungen der Regelgeltung (Formann, 1973, S. 17)

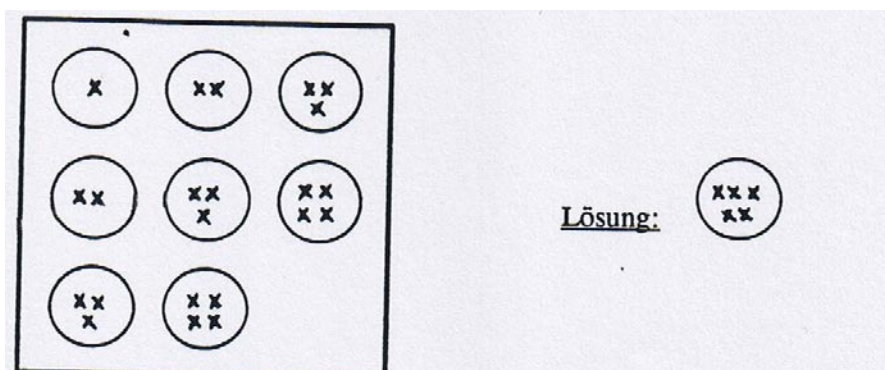
Waagrechte Regelgeltung: z.B. Muster waagrecht überlagern



Senkrechte Regelgeltung: z.B. Form senkrecht variieren



Waag- und senkrechte Regelgeltung: z.B. Anzahl waagrecht und senkrecht



Bei der Lösung haben die Testpersonen somit drei wesentliche Punkte zu beachten:

1. Welche Regeln gelten bezüglich
2. welcher relevanten Materialeigenschaften in
3. welcher Richtung des Schemas.

Schlussendlich zeigte sich, dass mit den Faktoren:

F1 = Anzahl der Komponenten je Figur

F2 = relevante Materialeigenschaft

F3 = geltende Regel je relevanter Materialeigenschaft je Figureschema

F4 = Regelrichtung

Aufgaben mit unterschiedlicher Schwierigkeit konstruiert werden konnten. F1 bis F4 mit der Menge aller Repräsentanten der Materialeigenschaften (relevante und irrelevante) definieren das Itemuniversum. Aus diesem Itemuniversum wurden von Formann (1973) 42 Items ausgewählt.

Diese 42 Items gab er 348 SchülerInnen im Alter von 15-19 Jahren aus Allgemeinbildenden Höheren Schulen vor.

Um die Rasch-Homogenität für die 42 Items zu überprüfen, wurden die Daten graphischen Modelltests unterzogen. Der Rohscore (hoch/niedrig) wurde als internes Teilungskriterium herangezogen und das Alter und die Mathematiknote fungierten als externe Teilungskriterien.

Ein Item erwies sich als zu leicht und wurde daher eliminiert. Bei Modelltest 1 mit dem Teilungskriterium Rohscore, erwiesen sich sechs Items als signifikant. Bei Modelltest 2 mit dem Teilungskriterium Alter gab es ein signifikantes Item und bei Modelltest 3 mit dem Teilungskriterium Mathematiknote erwiesen sich drei Items als signifikant. Insgesamt mussten also 10 Items ausgeschlossen werden. Dies wurde mit der Spezifität des Rasch-Modells und den verbleibenden Freiheitsgraden bei der Itemkonstruktion erklärt.

Nach Elimination dieser 10 Items war das Ergebnis ein homogener Test mit 32 Items, der genau eine latente Eigenschaft erfasst, nämlich „Schlussfolgerndes Denken“.

Mittels dem Linearen Logistischen Testmodell wurden unterschiedliche Hypothesen zum Lösungsverhalten der Testpersonen untersucht. Laut Formann (1973) reicht die Aufgabenstruktur in der hier verstandenen Art nicht aus, um die

Aufgabenschwierigkeit mit hinreichender Genauigkeit zu erklären (Formann, 1973, S. 86).

2.3 Die erste Version des WMT

Die 42 Items des FAM bildeten somit die Grundlage für einen neuen Matrizentest. Die notwendige Normierung machte Piswanger (1975) im Rahmen seiner Dissertation.

Hierfür stellte er einen interkulturellen Vergleich zwischen Österreich und Westafrika (Togo, Nigeria) an.

„Die Datenerhebung in Österreich erfolgte,..., nach folgenden wesentlichen Gesichtspunkten:

1. Es sollte für die Modelltests und für die klassische Eichung des selektierten, homogenen Tests eine genügend große Gesamtstichprobe von österreichischen, 14-19 jährigen, männlichen und weiblichen Schülern aus Berufsschulen, Fachschulen und höheren Schulen zur Verfügung stehen (etwa 2500 Pbn.)
2. Die Stichprobe sollte Schüler aus ländlichen und städtischen Gebieten etwa in gleichem Maße erfassen.
3. Es sollten je für die Modelltests gewähltem Teilungskriterium genügend große Gruppen zustande kommen.
4. Die Eichgruppen für die klassische Eichung sollten stark genügend besetzt sein.
5. Nicht zuletzt sollte die österreichische Stichprobe in etwa altersmäßig und in möglichst vielen anderen Kriterien der togolesischen und auch der nigerianischen Stichprobe entsprechen“ (Piswanger, 1975, S. 65)

Insgesamt wurden 2485 österreichische SchülerInnen getestet. In Togo waren es 159 Antwortbogen, die verwendet werden konnten und in Nigeria waren es 200 Antwortbogen.

Bei der Eindimensionalitätsprüfung der Daten der österreichischen Stichprobe wurden Modelltests für folgende acht Teilungskriterien gerechnet:

- Hoher/niedriger Rohscore (Trennscore 21)

- Geschlecht
- Schultyp (mit/ohne Matura)
- Beruf des Vaters (Selbständige, höhere oder mittlere Angestellte, Beamte, Facharbeiter vs. Arbeiter, Hilfsarbeiter, Rentner, Landwirte)
- Alter
- Selbsteinschätzung vorher (Trennscore 21)
- Selbsteinschätzung nachher
- Mathematiknote

Nach signifikanten Modelltests mussten 10 Items ausgeschlossen werden.

Mit den Daten der togolesischen Stichprobe wurden drei Modelltests gerechnet:

- Hoher / niedriger Rohscore (Trennscore 16)
- Geschlecht
- Stadt / Land

Aufgrund signifikanter Ergebnisse mussten vier weitere Items ausgeschlossen werden. Um zu überprüfen, ob der Test in Österreich und in Togo eindimensional misst, wurden auch die kombinierten Daten Österreich/Togo diversen Modelltests unterzogen. Danach mussten nochmals vier Items eliminiert werden.

Ebenso wurden Modelltests der nigerianischen Daten mit folgenden Teilungskriterien gerechnet:

- Hoher / niedriger Rohscore (Trennscore 18)
- Geschlecht
- Slum / Nobelgegend

Dies führte, ebenso wie die Kombination der österreichischen und nigerianischen Daten sowie der nigerianischen und togolesischen Daten, zu keiner weiteren Itemeliminierung.

Die endgültige Fassung des WMT umfasste nach Eliminierung aller Items, die in einem der durchgeführten Modelltests signifikant waren, noch 24 Items. Diese erfassten in allen drei Stichproben dasselbe latente Merkmal.

Im Anschluss daran konnten für Österreich Normtabellen erstellt werden und die konvergente Validität wurde mit Hilfe von Ravens Progressive Matrices (1938) und dem Intelligenz-Struktur-Test (Amthauer, 1973) berechnet. Beide lieferten hohe Korrelationskoeffizienten, sodass der WMT ein „Indikator-Merkmal für die intellektuelle Leistungsfähigkeit“ (Piswanger, 1975, S. 159) ist und daher als allgemeiner Intelligenztest gilt.

Im Anschluss daran, erfolgten weitere Modelltests. Hier wurden die Itemparameterschätzungen des 42-Item-Sets mit denen der Validitätsstichprobe und mit der nigerianischen Stichprobe verglichen. Teilungskriterien hierbei waren:

- Itemparameterschätzungen unter und über Rohscore 12
- Schultyp
- Geschlecht
- Alter

Das Ergebnis war „die Gewissheit, einen homogenen, eindimensional messenden Matrizentest zur praktischen Anwendung zur Verfügung stellen zu können“ (Piswanger, 1975, S. 165). Im Anschluss daran, erhielt der ursprünglich genannte FAM endgültig den Namen WMT.

Somit konnte 1979 das Testmanual „WMT. Wiener Matrizen-Test. Ein Rasch-skaliertes sprachfreier Intelligenztest“ von Formann und Piswanger veröffentlicht werden.

Laut diesem Testmanual ist der WMT ein Test zur Erfassung der sprachfreien Intelligenz, genauer, der „individuellen Fähigkeit (des) schlussfolgernden Denkens im Umgang mit abstrakten Symbolen“ (Formann & Piswanger, 1979, S. 7).

Er kann als Einzel- oder Gruppentest ohne strenge Zeitbegrenzung (ca. 25 Min.) durchgeführt werden. Nach drei Übungsbeispielen folgen 24 Aufgaben.

Im Testmanual der Version mit 24 Aufgaben sind Normtabellen für die Gesamtstichprobe und 8 nach Schultyp und/oder Alter differenzierter Teilgruppen enthalten. Die Normtabellen sind für beide Geschlechter gültig. Die Gesamtstichprobe kann hinsichtlich Schultyp, Alter und Geschlecht als repräsentativ angesehen werden. (Formann & Piswanger, 1979)

2.4 Die 2., revidierte Fassung des WMT (WMT-2)

Aufgrund der Tatsache, dass die Veröffentlichung des WMT nun beinahe 30 Jahre zurückliegt, erschien es notwendig, eine Neunormierung des WMT zu machen. Weiters war es interessant, die Eindimensionalitätshypothese der Items mittels nun zur Verfügung stehender Modelltests zu überprüfen. Denn um die Gültigkeit des probabilistischen Testmodells von Rasch (1960) zu überprüfen, wurden in der Erstversion des WMT die Modellkontrollen ausschließlich mittels grafischer Methode überprüft.

Doch nicht nur das Vorhandensein neuer Software zur Überprüfung der Modellkontrollen war ausschlaggebend für eine Neunormierung. Auch die sozialen und bildungspolitischen Veränderungen der letzten 30 Jahre waren ein Anstoß, um eine Neunormierung zu veranlassen. Die Entstehung der revidierten Fassung hatte somit mehrere Gründe. (Formann, Waldherr & Piswanger, in Druck)

Nach umfangreichen Modelltests mussten sechs Items aufgrund des Verdachts mangelnder Modellkonformität ausgeschlossen werden. Somit bilden die revidierte Fassung des WMT 18 Items, die geeignet sind, um „einen Eindruck vom allgemeinen intellektuellen Leistungsniveau einer Person zu erhalten“ (Formann, Waldherr & Piswanger, in Druck, S. 3).

In der 2., revidierten Fassung wurde eine österreichische Stichprobe von 2494 Personen erfasst. Dabei wurde die Normstichprobe um Erwachsene im Alter zwischen 19 und 64 Jahren erweitert.

Mit der revidierten Fassung des WMT steht somit ein sehr ökonomischer, reliabler und valider Kurztest zum sprachfreien Screening der intellektuellen Leistungsfähigkeit für den Altersbereich von 14-64 Jahren zur Verfügung. (Formann, Waldherr, Piswanger, in Druck, S. 3)

3. Bisherige Untersuchungen zum WMT

In diesem Abschnitt möchte ich zum einen auf die Untersuchungsergebnisse bei diversen Dimensionalitätsanalysen des WMT in Verbindung mit anderen Rasch-skalierten Tests eingehen und zum anderen die bisherigen Untersuchungsergebnisse diverser Kulturvergleiche mittels WMT darstellen.

3.1 Dimensionalitätsanalysen des WMT in Verbindung mit anderen Rasch-skalierten Tests

Nepita (1991) wollte im Rahmen ihrer Diplomarbeit herausfinden, ob der „Wiener Matrizen-Test“ (WMT), der „Mannheimer Test zur Erfassung des physikalisch-technischen Problemlösens“ (MTP, Conrad et al., 1980), das „Adaptives Intelligenz Diagnostikum“ (AID, Kubinger & Wurst, 1985) und der „Dreidimensionale Würfeltest“ (3DW, Gittler, 1990) jeweils eine eigene Dimension der Intelligenz messen. Dabei verwendete sie aus Zeitgründen beim 3DW nur die Kurzversion mit 10 Items. Beim AID wurden die Subtests 5, 7 und 8 nicht vorgegeben.

Die Stichprobe umfasste 30 AHS- und 87 HauptschülerInnen, somit insgesamt 117 Wiener SchülerInnen im Alter zwischen 13 und 15 Jahren.

Anhand der Ergebnisse der Likelihood-Quotiententests konnte festgestellt werden, dass „die acht zur Diskussion stehenden Subtests des AID andere Dimensionen erfassen als der WMT, der 3DW und der MTP“ (Nepita, 1991, S. 129). Die Kombination WMT/3DW erwies sich als Rasch-homogen. Nach Durchführung einer Faktorenanalyse konnte die Hypothese, dass der WMT, der 3DW und der MTP eine Dimension der Intelligenz erfassen, bestätigt werden, da alle drei Tests ihre höchsten Ladungen in einem einzigen Faktor ausweisen (Nepita, 1991, S. 122).

Kritisch dabei ist zu sehen, dass es sich um eine sehr kleine Stichprobengröße handelte (N=117) und davon 60 Testpersonen erst 13 Jahre alt waren, wobei gültige Normen erst ab einem Alter von 14 Jahren vorhanden sind. Weiters ist kritisch anzumerken, dass hier die Durchführung einer Faktorenanalyse und des weiteren deren Ergebnisse aufgrund der geringen Stichprobengröße in Frage gestellt werden müssen.

Gittler erscheint die positive Korrelation zwischen WMT und 3DW von $r=.35$ ($p<.001$) als „plausibel, da beide Verfahren komplexe kognitive Leistungen erfordern, und das Lösen von Testaufgaben somit auch gemeinsame Teilaspekte (intellektuelle Basisfertigkeiten, Motivation, Aufmerksamkeit etc.) beinhaltet“ (Gittler, 1999, S. 73).

Gittler wollte mit seinen Dimensionalitätsanalysen von WMT und 3DW (1999) herausfinden, ob beide Tests jeweils verschiedene latente Merkmale erfassen.

Seine Publikation unterteilt sich in zwei Studien, die eine Abgrenzung von Raumvorstellung (3DW) und schlussfolgerndem Denken (WMT) darstellen sollen. Studie I basiert dabei auf einer Reanalyse der Originaldaten von Nepita (1991) und Studie II bezieht sich auf Kapitel 6.1 der Habilitationsschrift von Gittler (1992), wobei die Daten von Stephenson (1987) stammen.

Im ersten Schritt erfolgten für jeden Test Modellprüfungen nach den Teilungskriterien Rohscore, Reaktion auf das erste Item, Geschlecht, Alter bzw. Jahrgang als Altersäquivalent und Schultyp bzw. Ausbildungsrichtung. Die Ergebnisse zeigten, dass Eindimensionalität des 3DW in beiden Studien gegeben war, jedoch waren beim WMT in Studie II 3 von 5 Modelltests signifikant. Nach Elimination von insgesamt 13 Aufgaben anhand verschiedener Ausschlusskriterien, wurden die homogenen 3DW Aufgaben mit den verbliebenen 11 homogenen WMT Aufgaben zu einem Test zusammengefügt.

Anhand der Teilungskriterien Leistung im jeweils anderen Test, Leistungsniveau, Geschlecht, Alter bzw. Jahrgang, Schultyp (Studie I) bzw. Ausbildungsrichtung (Studie II) und Reaktion auf das erste Item je Test sollte überprüft werden, ob beide Tests ein und dasselbe latente Merkmal erfassen.

Das Ergebnis waren signifikante Modelltests in beiden Tests, sodass sich folgendes sagen lässt: „Zusammenfassend kann festgestellt werden, dass die beiden Tests – laut Rasch-Modell - nachweislich nicht dieselbe latente Dimension erfassen, sondern jeweils etwas Spezifisches, was positive Aussagen für die Testgütekriterien (Konstukt-) Validität und Nützlichkeit beider Verfahren impliziert“ (Gittler, 1999, S. 80).

Auch bei dieser Studie sind einige Kritikpunkte anzumerken: Die Ausschlusskriterien, die zur Eliminierung von insgesamt 13 WMT Aufgaben führten, sind genauer zu betrachten, da der WMT laut Testautoren Rasch-Homogenität aufweist. Von den 13

Aufgaben wurden 5 ausgeschlossen, weil sie schlecht kopiert waren. Stephenson bemerkt dazu, dass dies „nicht in einer dem Test an und für sich immanenten Inhomogenität zu suchen“ sei (Stephenson, 1987, S. 86). Weitere 5 Aufgaben wurden ausgeschlossen, weil die Bearbeitungszeit zu knapp bemessen wurde, wobei im Testmanual eine Bearbeitungszeit von ca. 25 Minuten angegeben ist und somit eine Verkürzung zu einer Reduzierung der bearbeiteten Aufgabenanzahl und in Folge dessen zu fehlender Homogenität führen kann. Da jedoch auch die Aufgaben 6 und 7 ausgeschlossen werden mussten, erscheint es fraglich, ob der WMT tatsächlich eindimensional misst.

3.2 Bisherige Untersuchungen zum interkulturellen Vergleich mittels WMT

Edijeh (1980) nahm für seine Untersuchung die Studie von Piswanger (1975) als Anlass, daran anzuknüpfen und einen Kulturvergleich Österreich – Iran zu machen. Dabei gab er die ursprüngliche Form des WMT mit den 42 Aufgaben 303 iranischen MittelschülerInnen vor. Es zeigte sich, dass der WMT im Iran auf 28 Aufgaben reduziert werden muss, um Rasch-Homogenität zu erzielen und um weitere 6 Aufgaben im Kulturvergleich Österreich – Iran. Das Ergebnis waren somit 22 Rasch-homogene Items.

Mohammadzadeh Koucheri (1993) erweiterte den WMT um die Komponente „Vorgabeart“ und konstruierte so einen Test mit insgesamt 32 Aufgaben, den VMT (Variierter Matrizen-Test). Die Bearbeitungszeit betrug mit Instruktion 50 Minuten. Mittels diesem Test wollte sie untersuchen, „ob die Bearbeitung und Lösung einer Aufgabe aus dem Matrizentest von der kulturspezifischen Schriftrichtung abhängig ist“ (Mohammadzadeh Koucheri, 1993, S. 32).

Diesen modifizierten Test, der an die Schriftrichtung angepasst war, gab sie 756 IranerInnen und 289 ÖsterreicherInnen vor und prüfte anhand der Modelltests Rohscore, Geschlecht, Alter, Schultyp, Klasse, Beruf des Vaters, Links- vs. Rechtshänder und österreichische vs. iranische SchülerInnen auf Eindimensionalität. Von den 32 Aufgaben mussten 17 Aufgaben aufgrund signifikanter Ergebnisse der Modelltests eliminiert werden. Einen Einfluss der Schriftrichtung zeigte sich besonders bei der iranischen Stichprobe, da eine diskordante Vorgabeart die

Aufgabenschwierigkeit signifikant erhöhte. „Eine Aufgabe ist diskordant, wenn die Schriftrichtung entgegengesetzt der optimalen Bearbeitungsrichtung verläuft“ (Mohammadzadeh Koucheri, 1993, S. 17). Bei der österreichischen Stichprobe zeigten sich hingegen keine Unterschiede. Die Bearbeitung gegen die kulturspezifische Schriftrichtung zeigte keinen Einfluss auf die Aufgabenschwierigkeit.

Einen Kulturvergleich Österreich-Chile machte Broer (1996) in seiner Untersuchung mittels WMT und 3DW. Der WMT wurde dabei um die Antwortmöglichkeiten „Ich weiß die Lösung nicht“ und „Keine Figur vervollständigt das Muster“ erweitert. Die Stichprobe bestand aus chilenischen und österreichischen StudentInnen. Nach verschiedenen Modelltests (Score, 3DW-Leistung, D2 – ein Aufmerksamkeits- und Belastungstest, Bearbeitungszeit, Alter und Bildung der Eltern) konnte Eindimensionalität des WMT in beiden Stichproben nachgewiesen werden. Die Modellanpassung bei der chilenischen Stichprobe war im Vergleich zu anderen interkulturellen Untersuchungen sehr gut. Hier wird vermutet, dass dies auf die „Nähe“ der unterschiedlichen Kulturen zurückzuführen ist oder auch mit den zusätzlichen Antwortmöglichkeiten in Verbindung stehen könnte.

Der Modelltest Österreich-Chile brachte ein signifikantes Ergebnis, womit keine Rasch-Homogenität gegeben war, im Gegensatz zum 3DW, der ein nicht signifikantes Ergebnis brachte und somit Rasch-Homogenität angenommen werden konnte. Daraus schloss Broer, dass der WMT für diesen Kulturvergleich nicht geeignet ist.

Im Rahmen der Neunormierung wurde jeweils eine Stichprobe in Deutschland und den Vereinigten Arabischen Emiraten erhoben. Beide Stichproben waren sehr klein: in Deutschland betrug die Stichprobe 66 Personen, die sich aus SchülerInnen einer Abiturklasse eines Gymnasiums und jungen Erwachsenen aus Nürnberg und Karlsruhe zusammensetzte.

Die Stichprobe der Vereinigten Arabischen Emirate repräsentiert mit 48 Personen die Gesellschaftsstruktur der Einwohner mit höherem Bildungsstand: 87% der Personen haben einen Matura- oder Universitätsabschluss.

Anhand grafischer Modelltests konnte gezeigt werden, dass eine gute Übereinstimmung der geschätzten Itemparameter mit den Itemparameterschätzwerten aus der österreichischen Normstichprobe vorliegt.

3.3 Weitere Untersuchungen

Hahne (1999) untersuchte, ob mögliche Lerneffekte bei Leistungstests eine Rolle spielen. Dazu gab sie den WMT insgesamt 403 AHS-SchülerInnen in unterschiedlicher Reihenfolge vor und untersuchte, ob sich bei den unterschiedlichen Itemreihenfolgen Lerneffekte zeigten. Mit der zusätzlichen Antwort „Ich weiß es nicht“ sollte Raten vermieden werden. Zur Bearbeitung hatten die SchülerInnen ca. eine Stunde Zeit.

Laut Hahne lassen die Ergebnisse „den Schluss zu, dass der Wiener Matrizen-Test (WMT) nicht durch Lerneffekte belastet ist“ (Hahne, 1999, S. 72). Als Kritik ist hierbei wiederum zu bemerken, dass es sich bei der Untersuchung um eine sehr kleine Stichprobe handelte.

Dies widerspricht der allgemeinen Annahme, dass Lerneffekte bei Matrizentests zu beobachten sind. Hier sei die Untersuchung von Stephenson (1987) zu erwähnen. Stephenson ging der Frage nach, ob die Berufsausbildung einen Einfluss auf die Leistungen im 3DW und in einem speziell konstruierten Perspektivaufgaben-Test hat. Dazu gab er den WMT, den 3DW und den Perspektivaufgaben-Test insgesamt 395 BerufsschülerInnen verschiedener Ausbildungsrichtungen im ersten und dritten Lehrjahr vor. Der WMT wurde dabei nur zur Parallelisierung der Testpersonen als Intelligenztest verwendet. In der Stichprobe ohne Parallelisierung konnten Unterschiede beobachtet werden: die Technischen Zeichner hatten in beiden Tests die besten Ergebnisse. Am schlechtesten waren die Setzer und Drucker (die graphischen und kaufmännischen BerufsschülerInnen). Eine signifikante Verbesserung der Ergebnisse in Zusammenhang mit der Dauer der Berufsausbildung konnte jedoch nur bei den Technischen Zeichnern beobachtet werden.

Nach der Durchführung des Modelltests anhand des Kriteriums Rohscore wies der WMT ein signifikantes Ergebnis auf. Grund dafür waren schlechte Kopien von 4 Items, woraufhin diese ausgeschlossen wurden. Nach Eliminierung der Items erwies sich der WMT als Rasch-homogen. Auch beim Perspektivaufgaben-Test mussten 3

Items eliminiert werden, um Rasch-Homogenität zu erreichen. Beim 3DW zeigte sich kein signifikantes Ergebnis.

Vogt (2007) erstellte im Rahmen ihrer Diplomarbeit eine Neunormierung des WMT. Hierfür wurden 1386 SchülerInnen zwischen 14 und 27 Jahren aus Wien, Tirol, Kärnten, Oberösterreich und Salzburg getestet, wovon letztendlich 1296 Testhefte für die Neunormierung herangezogen werden konnten. Das Ergebnis waren neun Normtabellen für unterschiedliche Altersstufen und Schultypen für die Gesamtstichprobe.

Weiters überprüfte Vogt (2007) anhand der 1296 SchülerInnen, ob Rasch-Homogenität gegeben war. Nach Durchführung von 11 Modelltests anhand der externen Kriterien Alter, Muttersprache, Geschlecht, Mathematiknote, Schultyp, Angabe eines Kodenamens, Darstellende Geometrie, Geometrisches Zeichnen und Testbearbeitungsdauer und der internen Kriterien Rohscore und Reaktion auf das erste Item, wiesen fünf Modelltests nicht signifikante und sechs Modelltests signifikante Ergebnisse auf. Auch nach Durchführung der Alpha-Adjustierung nach Bonferroni blieben drei Modelltests signifikant.

Weiters untersuchte sie mittels Modelltests nach Martin – Löf, ob Homogenität verschiedener Itemteilmengen gegeben ist. Dazu verwendete sie folgende Kriterien: gerade/ungerade Itemnummer, Richtung der Regelgeltung (waag- und senkrecht), geltende Regel Variieren oder Fortsetzen oder Überlagern und erste vs. zweite Testhälfte.

Es zeigten sich signifikante Ergebnisse bei den Modelltests Überlagern und beim Vergleich der ersten Testhälfte mit der zweiten Testhälfte. Die Ergebnisse sprechen daher dafür, dass bei der Regel Überlagern unterschiedliche Dimensionen der Intelligenz angesprochen werden. Ebenso zeigten die Ergebnisse, dass die erste Testhälfte signifikant leichter zu lösen ist, als die zweite Testhälfte.

Vogt (2007) hat sich trotz dieser Ergebnisse gegen eine Eliminierung von Items entschieden, da „einerseits die Anzahl von 24 Items nicht weiter dezimiert werden sollte und andererseits der WMT in seiner vorliegenden Fassung durchaus als gut eingeführter, praktikabler und ökonomischer Intelligenztest betrachtet werden kann“ (Vogt, 2007, S. 79).

3.4 Zusammenfassung

Zusammenfassend kann gesagt werden, dass der WMT bereits in verschiedenen Kulturen vorgegeben wurde. In den meisten Untersuchungen mussten Aufgaben ausgeschlossen werden, um Rasch-Homogenität zu erreichen. Das heißt, es konnte nach Ausschluss von Aufgaben gewährleistet werden, dass in verschiedenen Kulturen das gleiche latente Merkmal erfasst wurde (vgl. Piswanger, 1975; Edijeh, 1980; Mohammadzadeh-Koucheri, 1993; Broer, 1996).

Im Zuge dieser Untersuchungen konnte somit gezeigt werden, dass der WMT in verschiedenen Kulturen ein und dasselbe Merkmal misst und dass die Leistungen gleichaltriger Testpersonen aus verschiedenen Kulturkreisen miteinander verglichen werden können.

Hierbei ist zu beachten, dass der WMT eine Intelligenzdimension misst, die bei allen Personen zu beobachten ist. Das heißt jedoch nicht, dass das Leistungsniveau der Personen unterschiedlicher Kulturen gleich sein muss.

Vogt (2007) erstellte im Rahmen ihrer Diplomarbeit neue Normen für eine österreichische Stichprobe und überprüfte den Datensatz auf Rasch-Homogenität. Obwohl sich in den Modelltests signifikante Ergebnisse zeigten, wurde keine Eliminierung von Items veranlasst, sodass für diese Daten Rasch-Homogenität nicht gegeben ist.

In Bezug auf die Rasch-Homogenität des WMT liefern all diese Untersuchungen unterschiedliche Ergebnisse. Gemeinsam ist allen Untersuchungen, dass Items ausgeschlossen werden mussten oder sollten (siehe Vogt, 2007), um Rasch-Homogenität zu erreichen. Jedoch konnte nach Eliminierung von Items gezeigt werden, dass der WMT eine Dimension der Intelligenz erfasst, nämlich „Schlussfolgerndes Denken“.

An diese Ergebnisse möchte ich anknüpfen und untersuchen, ob die „Nähe“ von Kulturen einen Einfluss auf die Ergebnisse hat und ob sich zwei Stichproben, die sich kulturell kaum voneinander unterscheiden, in ihren Ergebnissen voneinander abweichen.

4. Ziele und Hypothesen der Untersuchung

In Anlehnung an bisherige Untersuchungen erschien es mir interessant zu untersuchen, ob sich in den Testleistungen des WMT-2 eine Berliner Stichprobe von einer österreichischen Stichprobe unterscheidet bzw. ob Rasch-Homogenität für diese Daten gegeben ist.

Die Hauptfragestellung dieser Arbeit beschäftigt sich daher mit der Frage, inwieweit für den Berliner Datensatz die Eindimensionalitätshypothese angenommen werden kann. Überprüft werden soll dies mittels graphischer und parametrischer Modelltests. Weiters möchte ich untersuchen, ob verschiedene Itemteilmengen dieselbe latente Dimension der Intelligenz erfassen.

Aufgrund der Ergebnisse aus dem Testheft des WMT-2 liegt die Vermutung nahe, dass die Eindimensionalitätshypothese angenommen werden kann.

Im Rahmen dieser Arbeit soll auch eine Normierung der Berliner Stichprobe vorgenommen werden. Im Zuge dessen soll ebenso untersucht werden, ob es Unterschiede in den Testleistungen der Berliner Stichprobe und der österreichischen Stichprobe gibt. Hier vermute ich aufgrund der uneinheitlichen Schulsysteme der beiden Länder, dass sich signifikante Unterschiede zeigen werden.

Weiters erschien es mir sinnvoll die Hauptgütekriterien Objektivität, Reliabilität und Validität zu überprüfen. Um die Validität des WMT-2 zu ermitteln, sollte zusätzlich zum WMT-2 der Zahlen-Verbindungs-Test (ZVT, Oswald & Roth, 1987) den SchülerInnen zur Bearbeitung vorgegeben werden.

5. Planung der Durchführung der Untersuchung

Die Stichprobe sollte Berliner Schüler und Schülerinnen aus den mit Österreich vergleichbaren Schultypen erfassen. Geplant war daher eine Testdurchführung in Berliner Berufsschulen, in Gesamtschulen und in Gymnasien. Der Altersbereich sollte SchülerInnen im Alter von 14-20 Jahren erfassen.

Die Datenerhebung sollte nach den folgenden Merkmalen erfolgen:

1. Die Gesamtstichprobe sollte eine Größe aufweisen, anhand derer ein aussagekräftiger Vergleich der Normen gemacht werden kann. Es wurde daher im Vorfeld eine Größe von ca. 1000 SchülerInnen festgelegt.
2. Die Stichprobe sollte so gewählt werden, dass unter Berücksichtigung der Sozialstruktur der Stadt Berlin die Stichprobe als repräsentativ für Berlin angesehen werden kann. Die Schulen wurden daher unter Berücksichtigung der Sozialstruktur so ausgewählt, dass Repräsentativität gegeben ist. Auf diesen Punkt wird noch genauer eingegangen.
3. Hinsichtlich der Teilungsgruppen der Modelltests sollte eine ausreichend große Stichprobe zustande kommen.

Die Auswahl der Schulen fand, wie bereits kurz erwähnt, unter Berücksichtigung der Sozialstruktur Berlins statt. Dazu verwendete ich den „Sozialstrukturatlas Berlin 2003“ (Senatsverwaltung für Gesundheit, Soziales und Verbraucherschutz (2004). Erhoben wurden dabei verschiedene Variablen wie Schulabschluss, Einkommen, Ausländeranteil und Sozialhilfeempfänger. Demnach lässt sich Berlin in drei Zonen unterteilen: Stadtbezirke mit schlechter, mittelmäßiger und guter Sozialstruktur. Anhand dieser Unterteilung sollten die Schulen ausgewählt werden.

Nachdem die wesentlichen Kriterien der Stichprobe festgelegt waren, nahm ich Kontakt mit der Senatsverwaltung für Bildung, Wissenschaft und Forschung in Berlin auf, um mir die notwendige Genehmigung für die Testdurchführung an den Schulen zu holen. Danach erfolgte die Kontaktaufnahme mit den in Frage kommenden Schulen. Insgesamt sollte ich an mindestens neun Schulen testen, um Repräsentativität zu gewährleisten.

Um die Genehmigung der Senatsverwaltung zu bekommen, war es notwendig, bereits unterschriebene Einverständniserklärungen von den Schulen zu haben.

Daher war bereits im Vorfeld der Untersuchung ein Aufenthalt in Berlin notwendig. Somit habe ich im November 2007 die für die Testdurchführung vorgesehenen Schulen persönlich kontaktiert, um mir die notwendigen Einverständniserklärungen zu besorgen. Am Ende des Aufenthalts konnte ich die Einverständniserklärungen von insgesamt neun Schulen in der Senatsverwaltung für Bildung, Wissenschaft und Forschung abgeben und in Folge dessen, erhielt ich die Genehmigung zur Durchführung der Testungen an den Schulen.

6. Unterschiede der Schulsysteme Deutschland – Österreich

Bevor ich auf die genaue Datenerhebung eingehe, möchte ich noch kurz die bestehenden Unterschiede des deutschen, im Vergleich zum österreichischen Schulsystem beschreiben.

6.1 Schulformen

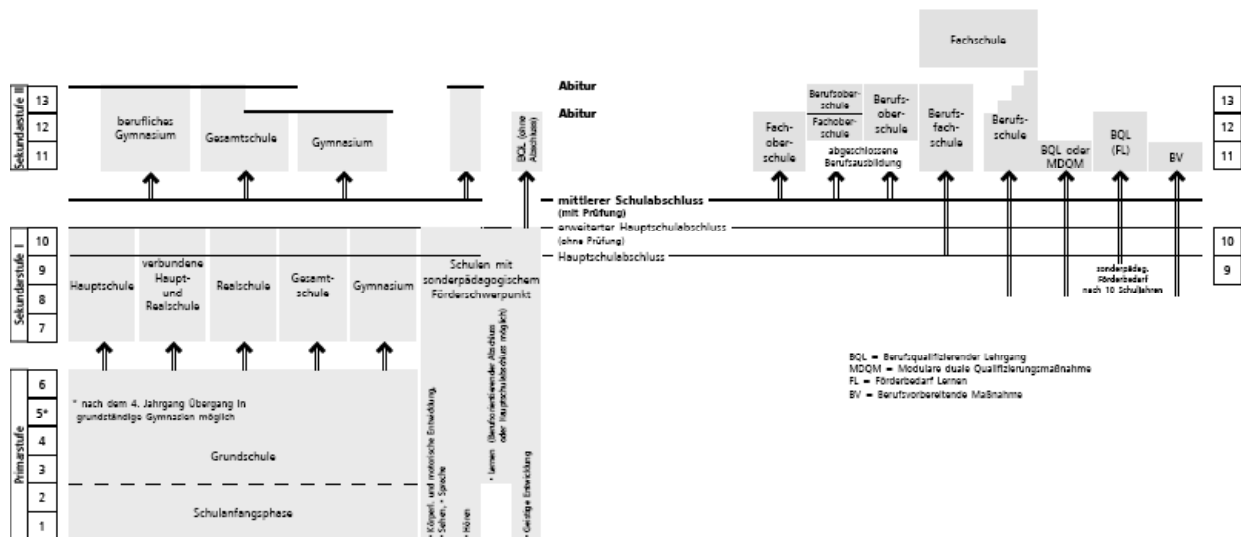
Bereits in den Vorbereitungen für meine Datenerhebung musste ich mich ausführlich mit dem deutschen, insbesondere dem Berliner Schulsystem auseinandersetzen. Das Berliner Schulsystem unterteilt sich in sehr viele verschiedene Schulformen, die in der nachfolgenden Abbildung dargestellt sind:

Abb. 5: Aufbau des Berliner Schulsystems (Senatsverwaltung für Bildung, Wissenschaft und Forschung, 2006)

Aufbau des Berliner Schulsystems

Allgemein bildende Schulen

Berufliche Schulen



Anhand dieser Abbildung lässt sich erkennen, dass die Grundschule, anders als in Österreich, sechs Schulstufen beinhaltet. Danach fällt je nach Fähigkeiten der SchülerInnen die Entscheidung für die weitere Schullaufbahn. Mögliche Schulformen sind dabei, wie auch in Österreich die Hauptschule und das Gymnasium. Anders als in Österreich stehen weitere Schulformen zur Verfügung, nämlich die Verbundene

Haupt- und Realschule, die Realschule und die Gesamtschule (Schulversuche in Österreich lasse ich hierbei außer Acht).

Um Repräsentativität meiner Stichprobe zu gewährleisten, habe ich anhand der österreichischen Normstichprobe des WMT-2 vergleichbare Schulformen ausgewählt. Die Schulformen Berufsschule, Gesamtschule und Gymnasium kamen dabei in Frage, die ich nun etwas genauer beschreiben möchte.

6.1.1 Berufsschule

Sie vermitteln den SchülerInnen, die in einem Berufsausbildungsverhältnis stehen, vor allem die für den gewählten Beruf erforderlichen fachtheoretischen Kenntnisse und erweitern die Allgemeinbildung in Anknüpfung an die beruflich erworbenen Einsichten und Erfahrungen.

(Schulgesetz für Berlin. Berlin 2004, S. 38 ff.)

6.1.2 Berufsoberschule

Die Berufsoberschulen vermitteln in einem zweijährigen Vollzeitausbildungsgang eine allgemeine und fachtheoretische Bildung. Sie führt zur fachgebundenen Hochschulreife und beim Nachweis der notwendigen Kenntnisse einer zweiten Fremdsprache zur allgemeinen Hochschulreife. (Schulgesetz für Berlin. Berlin 2004, S. 38 ff.)

6.1.3 Gesamtschule

Gesamtschulen fassen in den Klassenstufen 7 bis 10 das Unterrichtsangebot der Hauptschule, der Realschule und des Gymnasiums in einer Schule zusammen. Ziel der Gesamtschule ist es, in diesen Klassenstufen alle SchülerInnen unabhängig von ihren Lernvoraussetzungen gemeinsam zu unterrichten und zu betreuen und sie gleichzeitig durch gezielte Maßnahmen individuell zu fördern.

Der erfolgreiche Besuch dieser Schulart führt in Jahrgangsstufe 9 zum Hauptschulabschluss und in Jahrgangsstufe 10 zum erweiterten Hauptschulabschluss oder zum mittleren Schulabschluss, welcher zum Übergang in

die zweijährige Form der gymnasialen Oberstufe berechtigt. (Schulgesetz für Berlin. Berlin 2004, S. 38 ff.)

6.1.4 Gymnasien

An dieser Stelle muss ich erwähnen, dass trotz dem Bemühen meinerseits, sich leider kein Schulleiter eines Gymnasiums bereit erklärt hat, an der Untersuchung teilzunehmen.

Somit besteht die Gesamtstichprobe aus SchülerInnen von Berufs- und Gesamtschulen.

6.2 Notensystem

Im Unterschied zum österreichischen, ist das deutsche Notensystem in sechs Stufen gegliedert:

- 1 (*sehr gut*) – wenn die Leistung den Anforderungen in besonderem Maße entspricht.
- 2 (*gut*) – wenn die Leistung den Anforderungen voll entspricht.
- 3 (*befriedigend*) – wenn die Leistung im Allgemeinen den Anforderungen entspricht.
- 4 (*ausreichend*) – wenn die Leistung zwar Mängel aufweist, aber im Ganzen den Anforderungen noch entspricht.
- 5 (*mangelhaft*) – wenn die Leistung den Anforderungen nicht entspricht, jedoch erkennen lässt, dass die notwendigen Grundkenntnisse vorhanden sind und die Mängel in absehbarer Zeit behoben werden können.
- 6 (*ungenügend*) – wenn die Leistung den Anforderungen nicht entspricht und selbst die Grundkenntnisse so lückenhaft sind, dass die Mängel in absehbarer Zeit nicht behoben werden können. (Online unter URL: <http://de.wikipedia.org/wiki/Schulnote>)

6.3 Punktesystem

An der gymnasialen Oberstufe und in Berufsoberschulen gilt in Deutschland ein aufwärts zählendes Punktesystem, das im Abitur eingesetzt wird und der leichteren Vergleichbarkeit und Berechnung der Endnote dient:

15–13 Punkte entsprechen der Schulnote 1 (*sehr gut*)

12–10 Punkte entsprechen der Schulnote 2 (*gut*)

9–7 Punkte entsprechen der Schulnote 3 (*befriedigend*)

6–5 Punkte entsprechen der Schulnote 4 (*ausreichend*)

4 Punkte entsprechen der Schulnote 4– (*schwach ausreichend*)

3–1 Punkte entsprechen der Schulnote 5 (*mangelhaft*)

0 Punkte entsprechen der Schulnote 6 (*ungenügend*)

Eine voll ausreichende Leistung ist erst ab 5 Punkten erbracht. (Online unter URL: <http://de.wikipedia.org/wiki/Schulnote>)

7. Datenerhebung

Die Datenerhebung erstreckte sich vom Zeitraum Februar bis April 2008. Die Testvorgabe erfolgte klassenweise. Insgesamt wurden an jeder Schule ca. 120 SchülerInnen getestet, sodass sich die Testungen meist über 5 oder mehr Schulstunden erstreckten.

Am Beginn der Testung stellte ich mich kurz vor und erklärte den Grund der Testung. Ebenso erwähnte ich die Freiwilligkeit der Teilnahme an der Testung, die Anonymität und die Rückmeldung der Testergebnisse anhand der Codenamen. Danach teilte ich die Antwortblätter aus und ging die einzelnen Punkte mit den SchülerInnen gemeinsam durch. Dann bekamen sie die Testhefte und ich erklärte die Aufgabenstellung anhand der Übungsbeispiele. Weiters fügte ich hinzu, dass die SchülerInnen, wenn sie eine Aufgabe nicht wissen, Raten vermeiden und die betreffende Antwort besser auslassen sollen. Erst dann begann die Testbearbeitung. Im Anschluss an den WMT wurde der Zahlen-Verbindungs-Test (ZVT, Oswald & Roth, 1987) ausgeteilt. Nachdem ich den SchülerInnen erklärt hatte, worum es bei dem Test geht, bearbeiteten sie die beiden Übungsbeispiele auf der ersten Seite. Danach erfolgte die Testbearbeitung nach den Vorgaberegeln des Testmanuals. Am Ende der Testung hatten die SchülerInnen noch die Möglichkeit Fragen zu stellen.

Aufgrund mangelnder Motivation und Kooperation von Seiten der Schüler gestaltete sich die Vorgabe des ZVT als sehr schwierig. Ein wesentlicher Grund, weshalb nur ein Teil der Gesamtstichprobe den ZVT bearbeiteten lag auch daran, dass die Schulleiter die Testungen in die verkürzten Unterrichtsstunden legten, die anstelle von 45 Minuten nur 30 Minuten dauerten.

8. Stichprobe

Insgesamt wurden 728 SchülerInnen getestet, wobei 68 aufgrund mangelnder Motivation, die Testhefte ernsthaft zu bearbeiten, ausgeschlossen werden mussten. Diese 68 SchülerInnen mussten ausgeschlossen werden, da der Schulleiter die Testungen in der Freistunde der SchülerInnen ansetzte. Die Bereitschaft für eine ernsthafte Bearbeitung war, trotz Bemühen meinerseits, nicht gegeben. Aufgrund dessen habe ich mich entschieden, die Testung vorzeitig abubrechen (geplant war eine Testung von 120 SchülerInnen).

Ein weiteres Problem bei der Datenerhebung stellte die Terminfindung in den einzelnen Schulen dar. Leider waren von den anfangs zustimmenden neun Schulen nur vier Schulen bereit, an der Untersuchung teilzunehmen. Die anderen fünf Schulen wollten aufgrund zeitlicher Probleme doch nicht an der Testung teilnehmen. Somit musste ich Ersatz für diese Schulen finden, was sich als ziemliche Herausforderung gestaltete.

Letztendlich gelang es mir, drei weitere Schulen für mein Vorhaben zu gewinnen, sodass ich auf den Datensatz von 728 SchülerInnen kam. Wie bereits erwähnt, mussten davon 68 SchülerInnen ausgeschlossen werden, sodass die Gesamtstichprobe 660 SchülerInnen umfasst.

Tabelle 1 zeigt die Stichprobenzusammensetzung im Vergleich mit der österreichischen Normstichprobe des WMT-2.

Tabelle 1

	Berliner Stichprobe		WMT-2	
	N	%	N	%
Schultypen				
Fach- und Berufsschulen	172	26	661	51.3
AHS/BHS bzw. Gesamtschulen	488	74	628	48.7
Gesamt	660	100	1289	100
Alter				
14-16 Jahre	306	47.8	761	55.14
17-18 Jahre	196	30.6	528	38.26
19-25 Jahre	138	21.6	91	6.6
Gesamt	640	100	1380	100
Geschlecht				
Weiblich	285	43.2	604	43.8
Männlich	375	56.8	776	56.2
Gesamt	660	100	1380	100

Es zeigt sich eine ungleiche Verteilung der Prozentwerte in den beiden Stichproben: So ist der Anteil an Fach- und BerufsschülerInnen in der österreichischen Stichprobe mit 51.3% wesentlich höher als in der Berliner Stichprobe, wo der Anteil an BerufsschülerInnen 26% beträgt. Auch hinsichtlich des Alters gibt es gravierende Unterschiede. Der Anteil an 19-25 jährigen liegt in der Berliner Stichprobe mit 21.6% wesentlich höher, als der Anteil an 19-25 jährigen bei der österreichischen Stichprobe mit 6.6%. In den beiden anderen Altersgruppen ist der Prozentsatz in der Berliner Stichprobe jeweils um einiges niedriger, als in der österreichischen Stichprobe. Lediglich das Geschlecht ist in beiden Stichproben beinahe gleichverteilt.

Im Folgenden möchte ich auf die einzelnen Variablen und deren Ausprägungen in der Gesamtstichprobe näher eingehen.

8.1 Geschlecht

Die Gesamtstichprobe von 660 SchülerInnen unterteilt sich hinsichtlich des Geschlechts folgendermaßen:

Tabelle 2: **Geschlecht der Testperson**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	männlich	375	56,8	56,8	56,8
	weiblich	285	43,2	43,2	100,0
	Gesamt	660	100,0	100,0	

Der Anteil der Burschen betrug 56,8% und war somit erhöht.

8.2 Muttersprache

Da in den bisherigen Untersuchungen auch die Muttersprache erhoben wurde und es ebenso aufgrund der „Kulturfreiheit“ interessant erschien, nahm ich diesen Aspekt in meine Datenerhebung mit hinein. Hinsichtlich der Muttersprache gestaltete sich die Gesamtstichprobe wie folgt:

Tabelle 3: **Muttersprache der Testperson**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	deutsch	504	76,4	76,4	76,4
	andere	156	23,6	23,6	100,0
	Gesamt	660	100,0	100,0	

Wie erwartet war der Anteil der SchülerInnen mit Deutsch als Muttersprache, mit 76,4% überrepräsentiert.

8.3 Alter

Obwohl für den WMT Normen ausschließlich für SchülerInnen von 14 bis 18 Jahren vorliegen, habe ich auch die jüngeren und älteren SchülerInnen in die Berechnungen mit hineingenommen, da sich deren Ergebnisse nicht signifikant von der Normstichprobe unterschieden.

Der Mittelwert der Gesamtstichprobe betrug 17,13 Jahre, bei einer Standardabweichung von 2,53.

Die genaue Verteilung ist in der unten stehenden Tabelle ersichtlich:

Tabelle 4: **Alter der Testperson**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 12	2	,3	,3	,3
13	17	2,6	2,6	2,9
14	39	5,9	5,9	8,8
15	112	17,0	17,0	25,8
16	155	23,5	23,5	49,2
17	121	18,3	18,3	67,6
18	75	11,4	11,4	78,9
19	24	3,6	3,6	82,6
20	36	5,5	5,5	88,0
21	31	4,7	4,7	92,7
22	17	2,6	2,6	95,3
23	11	1,7	1,7	97,0
24	11	1,7	1,7	98,6
25	8	1,2	1,2	99,8
27	1	,2	,2	100,0
Gesamt	660	100,0	100,0	

8.4 Schultyp

Wie bereits im Vorwort erwähnt, konnte ich leider keine Testungen an Gymnasien durchführen, sodass sich die Gesamtstichprobe letztendlich nur aus Berufs- und GesamtschülerInnen zusammensetzt.

Von den 660 SchülerInnen besuchten 172 eine Berufsschule. Aufgrund mangelnder Kooperation von Seiten der Schulleiter konnte ich leider nur an zwei Berufsschulen Testungen durchführen, sodass diese Gruppe mit 26,1% unterrepräsentiert ist. Die restlichen 488 SchülerInnen besuchten eine Gesamtschule. In Tabelle 5 ist die

genaue Verteilung der Gesamtstichprobe hinsichtlich der Variable Schultyp dargestellt.

Tabelle 5: **Besuchter Schultyp der Testperson**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Berufsschule	172	26,1	26,1	26,1
	Gesamtschule	488	73,9	73,9	100,0
	Gesamt	660	100,0	100,0	

8.5 Mathematiknote

Um einen möglichen Zusammenhang zwischen dem Mathematikverständnis und dem Lösen der Aufgaben zu erkennen, sollten die SchülerInnen ihre Mathematiknote im letzten Zeugnis angeben.

Wie bereits in Abschnitt 6.2 erwähnt, wird das Notensystem in Deutschland in 6 Stufen unterteilt. Bei der Berliner Gesamtstichprobe zeigt sich an Anteil von 5,3% mit der Note 1 und ein Anteil von 1,1% mit der Note 6. Der Mittelwert betrug 3,25 mit einer Standardabweichung von 1,1.

Von den 660 SchülerInnen haben 19 keine Angabe gemacht. Das entspricht 2,9% der Gesamtstichprobe und stellt den höchsten Prozentsatz an fehlenden Werten dar.

Tabelle 6 zeigt die genaue Verteilung der Noten in der Gesamtstichprobe.

Tabelle 6: **Letzte Mathematiknote der Testperson**

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1	34	5,2	5,3	5,3
	2	135	20,5	21,1	26,4
	3	202	30,6	31,5	57,9
	4	185	28,0	28,9	86,7
	5	78	11,8	12,2	98,9
	6	7	1,1	1,1	100,0
	Gesamt	641	97,1	100,0	
Fehlend	999	19	2,9		
Gesamt		660	100,0		

8.6 Darstellende Geometrie und Geometrisch Zeichnen

Aufgrund der symbolhaften Aufgabendarstellung des WMT und in Anlehnung an bisherige Untersuchungen, war es auch sinnvoll zu erheben, ob die SchülerInnen Vorkenntnisse in Darstellender Geometrie oder Geometrisch Zeichnen haben. Daher wurden zum Schluss noch die beiden Fragen zur Ausbildung in Darstellender Geometrie bzw. in Geometrisch Zeichnen gestellt.

Tabelle 7: **Ausbildung der Testperson in Darstellender Geometrie**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ja	98	14,8	14,8	14,8
nein	562	85,2	85,2	100,0
Gesamt	660	100,0	100,0	

Tabelle 8: **Ausbildung der Testperson in Geometrisch Zeichnen**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ja	202	30,6	30,6	30,6
nein	458	69,4	69,4	100,0
Gesamt	660	100,0	100,0	

Nach einer genauen Erklärung, was der Unterschied dieser beiden Fragen ist und unter welchen Bedingungen die Fragen mit „Ja“ anzukreuzen sind, haben alle SchülerInnen diese beiden Fragen auch beantwortet, sodass keine missing values zu verzeichnen sind.

Hierbei muss angemerkt werden, dass das Fach „Geometrisch Zeichnen“ in Deutschland mit dem Fach „Technisches Zeichnen“ gleichzusetzen ist.

Die beiden Tabellen zeigen, dass der Anteil jener SchülerInnen, die eine Ausbildung in Geometrisch Zeichnen hatten mit 30.6% höher ist, als der Anteil jener SchülerInnen, die eine Ausbildung in Darstellender Geometrie hatten.

8.7 Bearbeitungsdauer

Da der WMT keine strenge Zeitbeschränkung hat und die durchschnittliche Bearbeitungsdauer 30 Minuten beträgt (bei 24 Aufgaben), ließ ich im Gegensatz zu bisherigen Untersuchungen die Variable „Bearbeitungsdauer“ außer Acht.

In der Regel bekam ich für die Dauer der Testung pro Klasse eine Schulstunde, das entspricht an Berliner Schulen 45 Minuten, zur Verfügung gestellt. Einige Schulleiter überließen mir jedoch nur eine verkürzte Schulstunde für die Testung. Dies waren 30 Minuten anstelle von 45 Minuten, was sich auf die Bearbeitungszeit des WMT-2 nicht auswirkte, sehr wohl jedoch auf den ZVT, da dieser im Anschluss an den WMT-2 nicht mehr vorgegeben werden konnte. Nach der Instruktion hatten die SchülerInnen die gesamte Schulstunde Zeit, die Aufgaben zu bearbeiten. Meist betrug die Bearbeitungsdauer zwischen 20 und 30 Minuten. Nur sehr wenige SchülerInnen waren sehr viel schneller oder um einiges langsamer.

9. Rasch-Modellprüfung

Da sich eine Hauptfragestellung dieser Arbeit auf die Überprüfung der Gültigkeit des Rasch-Modells bezieht, soll diese in diesem Abschnitt anhand parametrischer Modelltests überprüft werden. Dazu wurde das Programm LPCM.WIN 1.0 (Fischer & Ponocny-Seliger, 1998) verwendet.

Die Gesamtstichprobe wurde anhand von bestimmten Kriterien geteilt um die Gültigkeit des Rasch-Modells sowohl grafisch als auch mittels Likelihood – Quotienten Tests (Andersen, 1973) zu überprüfen. Die Teilungskriterien, sowie die Ergebnisse der Modelltests, sollen nun dargestellt und interpretiert werden.

9.1 Teilungskriterien

Damit man die Itemparameter von verschiedenen Teilstichproben miteinander vergleichen kann, war es notwendig, die Gesamtstichprobe anhand verschiedener Kriterien zu teilen. Dabei wurden in dieser Arbeit folgende externe Teilungskriterien verwendet:

- Alter
- Geschlecht
- Muttersprache
- Schultyp
- Mathematiknote
- Darstellende Geometrie
- Geometrisch Zeichnen

Weiters wurden noch zwei interne Teilungskriterien verwendet:

- Rohscore
- Reaktion auf das erste Item

Da bereits im Vorfeld diejenigen SchülerInnen von der Gesamtstichprobe herausgenommen wurden, die den WMT-2 nicht ernsthaft bearbeiteten, konnten alle 660 verbleibenden Testhefte für die Berechnungen verwendet werden.

Der erste Modelltest wurde anhand des Teilungskriteriums Alter durchgeführt. Dabei wurden die Itemparameterschätzungen der jüngeren SchülerInnen (12-16 Jahre) mit denen der älteren SchülerInnen (17-27) verglichen.

Die beiden Gruppen waren beinahe gleich groß mit 49.2% (n = 325) jüngeren und 50.8% (n = 335) älteren SchülerInnen. Der Mittelwert betrug bei den jüngeren 9.36 mit einer Standardabweichung von 3.440 und bei den älteren SchülerInnen 10.14 mit einer Standardabweichung von 3.748. Nach Durchführung eines t-Tests erwiesen sich die geringen Unterschiede als signifikant. Die jüngeren SchülerInnen haben demnach signifikant schlechtere Testleistungen als die älteren SchülerInnen.

Als zweites externes Kriterium wurde das Geschlecht zum Vergleich der Itemparameterschätzungen herangezogen. Der Anteil der männlichen Schüler war wie bereits erwähnt mit 56.8% gegenüber dem Anteil der weiblichen Schülerinnen mit 43.2% erhöht. Der Mittelwert der Burschen lag mit 9.47 (s = 3.539) unter dem der Mädchen mit 10.13 (s = 3.692). Auch hier brachte ein t-Test ein signifikantes Ergebnis, wonach die Burschen signifikant schlechtere Testleistungen erbrachten als die Mädchen.

Im dritten Modelltest wurde die Stichprobe anhand des externen Kriteriums Muttersprache geteilt. Dabei wurden die Itemparameterschätzungen der SchülerInnen mit Deutsch als Muttersprache mit den Itemparameterschätzungen der SchülerInnen verglichen, die eine andere Muttersprache angaben. Die Stichprobe verteilte sich dabei folgendermaßen: 76.4% der SchülerInnen gaben Deutsch als Muttersprache an. Der Mittelwert der SchülerInnen mit deutscher Muttersprache betrug 10.12 (s = 3.554) und war somit trotz der Kulturfreiheit des WMT-2 deutlich höher als jener SchülerInnen mit nicht deutscher Muttersprache. Hier lag der Mittelwert bei 8.48 mit einer Standardabweichung von 3.585. Nach Durchführung eines t-Tests konnte gezeigt werden, dass es sich dabei wiederum um ein signifikantes Ergebnis handelt. SchülerInnen mit Deutsch als Muttersprache haben signifikant bessere Testleistungen als SchülerInnen mit einer anderen Muttersprache.

Ein weiteres externes Teilungskriterium war der besuchte Schultyp. Hier wurde unterschieden zwischen den GesamtschülerInnen und den BerufsschülerInnen. Wie bereits in Abschnitt 8 beschrieben, war der Anteil der GesamtschülerInnen mit 73.9%

(n = 488) wesentlich höher als der Anteil der BerufsschülerInnen mit 26,1% (n = 172). Der Mittelwert der GesamtschülerInnen lag mit 9.58 (s = 3.469) unter dem der BerufsschülerInnen, die einen Mittelwert von 10.26 (s = 3.980) hatten. Der Grund dafür liegt wahrscheinlich in der Altersverteilung der beiden Teilstichproben, da der Altersdurchschnitt der GesamtschülerInnen bei 16 Jahren lag und die BerufsschülerInnen im Durchschnitt 20 Jahre alt waren. Auch hier brachte ein t-Test ein signifikantes Ergebnis, wonach GesamtschülerInnen signifikant schlechtere Testergebnisse hatten, als BerufsschülerInnen.

Der vierte Modelltest wurde anhand des externen Teilungskriteriums Mathematiknote durchgeführt. Dabei wurden die Itemparameterschätzungen der SchülerInnen mit einer zufriedenstellenden Mathematiknote (Note 1-3) mit jenen SchülerInnen verglichen, die eine wenig oder nicht zufriedenstellende Mathematiknote (4-6) angegeben hatten. Die Gruppe derjenigen, die eine zufriedenstellende Mathematiknote angaben, war mit 56.2% um einiges höher, als der Anteil an SchülerInnen mit einer wenig oder nicht zufriedenstellenden Mathematiknote, deren Anteil bei 43.8% lag. Der Mittelwert der SchülerInnen mit einer zufriedenstellenden Mathematiknote war mit 10.29 (s = 3.604) höher als der Mittelwert der SchülerInnen mit einer wenig oder nicht zufriedenstellenden Mathematiknote, der bei 9.13 (s = 3.387) lag. Der t-Test zeigt, dass es sich dabei um ein signifikantes Ergebnis handelt. SchülerInnen mit einer zufriedenstellenden Mathematiknote haben demnach auch bessere Testleistungen im WMT-2.

Beim externen Teilungskriterium Darstellende Geometrie wurden die Itemparameterschätzungen der SchülerInnen mit Vorkenntnissen in Darstellender Geometrie (14.8%) mit jenen SchülerInnen verglichen, die keine Vorkenntnisse in Geometrie angegeben hatten (85.2%). Der Mittelwert der SchülerInnen mit Ausbildung in Geometrie lag mit 10.71 (s = 3.535) höher als jener SchülerInnen, die keine Ausbildung in Geometrie angegeben hatten, da dieser Mittelwert bei 9.59 mit einer Standardabweichung von 3.610 lag. Nach Durchführung eines t-Tests kann gesagt werden, dass SchülerInnen mit Ausbildung in Geometrie signifikant bessere Testleistungen haben, als SchülerInnen ohne Ausbildung in Geometrie.

Als letztes externes Kriterium wurde die Stichprobe anhand der Vorkenntnisse in Geometrisch Zeichnen geteilt. Dabei waren jene SchülerInnen, die keine Vorkenntnisse hatten mit 30.6% wesentlich weniger, als die SchülerInnen, die angegeben hatte, Vorkenntnisse in Geometrisch Zeichnen zu haben (69.4%). Der Mittelwert der SchülerInnen ohne Vorkenntnisse war mit 9.72 ($s = 3.609$) nur geringfügig unter dem Mittelwert jener SchülerInnen, die Vorkenntnisse in Geometrisch Zeichnen hatten. Der Mittelwert betrug hier nämlich 9.85 mit einer Standardabweichung von 3.647. Hier brachte der t-Test ein nicht signifikantes Ergebnis, wonach sich SchülerInnen mit Vorkenntnissen in Geometrisch Zeichnen in ihren Testleistungen nicht von SchülerInnen ohne Vorkenntnisse in Geometrisch Zeichnen unterscheiden.

Bei der Durchführung des Modelltests anhand des internen Kriteriums Rohscore lag der Trennscore bei 10 gelösten Aufgaben, da der Mittelwert über die Gesamtstichprobe bei 9.77 Aufgaben lag. Die Gruppe der SchülerInnen, die mehr als 10 Aufgaben richtig löste, war mit 42.1% ($n = 278$) kleiner, als die Gruppe, die bis zu 10 Aufgaben richtig löste. Deren Anteil lag bei 57.9% ($n = 382$). Der Mittelwert der Gruppe, die mehr als 10 Aufgaben löste, betrug 13.22 ($s = 1.833$) und der Mittelwert der anderen Gruppe lag bei 7.24 ($s = 2.263$).

Beim letzten Modelltest wurde die Reaktion auf das erste Item als internes Teilungskriterium herangezogen. Hier folgte ein Vergleich der Itemparameterschätzungen der SchülerInnen, die das erste Item gelöst hatten, mit den SchülerInnen, die das erste Item nicht gelöst hatten. Fehlende Werte wurden als falsche Antworten gewertet, sodass alle 660 Testhefte verwendet werden konnten. Insgesamt konnten 81.1% der SchülerInnen das erste Item lösen und nur 18.9% konnten es nicht lösen. Der Mittelwert derjenigen, die es nicht lösen konnte, betrug 6.92 ($s = 3.171$) und der Mittelwert der anderen Gruppe lag bei 10.42 ($s = 3.390$). Auch hier erwies sich der t-Test als signifikant. Demnach haben SchülerInnen, die das erste Item nicht lösen konnten, signifikant schlechtere Testleistungen als SchülerInnen, die das erste Item lösen konnten.

Dass Item 1 zu den leichtesten Items des WMT gezählt wird, spielt in diesem Zusammenhang keine Rolle. Auch andere Items hätten hier als Teilungskriterium verwendet werden können.

Laut Testmanual des WMT-2 (Formann, Waldherr & Piswanger, in Druck) ist Item 2 das leichteste und Item 17 das schwierigste. Auch in dieser Arbeit stellt sich Item 2 als das leichteste heraus, da 88.5% dieses Item lösen konnten. Anders verhält es sich mit dem schwierigsten Item. Hier zeigte sich, dass Item 18 für die Stichprobe in dieser Arbeit am schwierigsten war. Nur 18.8% konnten dieses Item richtig lösen. Dabei ist jedoch anzumerken, dass diejenigen SchülerInnen, die mit den Aufgaben bereits fertig waren, langsamere SchülerInnen oft drängten, fertig zu werden. So sollte dieses Ergebnis nur mit Vorsicht interpretiert werden, da nicht ausgeschlossen werden kann, dass es an mangelnder Fairness von Seiten der SchülerInnen lag, dass leider nicht alle SchülerInnen die Aufgaben in Ruhe bearbeiten konnten.

9.2 Ergebnisse der Modelltests

Die kritischen Chi-Quadrat Werte betragen laut Tabelle „Verteilungsfunktion der Chi-Quadrat-Verteilung“ in Bortz & Döring (2005, S. 749) bei Annahme einer Irrtumswahrscheinlichkeit $\alpha = 0.01$ und 17 Freiheitsgraden (df) 33.41 und bei 16 Freiheitsgraden 32.

Da es aufgrund der Durchführung von neun Modelltests zu einer Zunahme des Alpha-Fehler-Risikos kommt, wurde zusätzlich zur grafischen Modellkontrolle auch eine Alpha-Adjustierung vorgenommen.

9.3 Die Durchführung der Alpha-Adjustierung

Bei jedem statistischen Hypothesentest besteht das Risiko des Alpha-Fehlers, das heißt, die Nullhypothese wird irrtümlicherweise verworfen, obwohl sie gültig ist. Daher spricht man auch von Irrtumswahrscheinlichkeit. Bei einem Signifikanzniveau von $\alpha = 5\%$ wird die Nullhypothese mit einer Wahrscheinlichkeit von 5% verworfen, auch

wenn sie zutrifft. Die Wahrscheinlichkeit, dass sie beibehalten wird, beträgt demnach $1-\alpha$ also $1-0.05 = 95\%$.

Der Alpha-Fehler steigt nun mit zunehmender Zahl an durchgeführten Signifikanztests. Dies wird als Alpha-Kumulierung bezeichnet. Eine Alpha-Fehler-Kumulierung tritt somit auf, wenn eine Hypothese über mehrere Signifikanztests geprüft wird.

Um dies zu verdeutlichen, soll nun veranschaulicht werden, wie sich die Wahrscheinlichkeit die Nullhypothese beizubehalten, bei der Durchführung von neun Modelltests, verändert:

Bei neun unabhängigen Modelltests beträgt die Wahrscheinlichkeit, in allen neun Tests die Nullhypothese beizubehalten, $(1-\alpha)$ potenziert mit der Anzahl an Tests, also in diesem Fall mit 9. Als Ergebnis erhalten wir bei $\alpha = 0.01$ eine Wahrscheinlichkeit von $(1-0.01)^9 = 0.9135$.

Die Wahrscheinlichkeit die Nullhypothese beizubehalten beträgt demnach nur noch 91%. Anhand der Gegenwahrscheinlichkeit kann gesagt werden, mit welcher Wahrscheinlichkeit die Nullhypothese trotz Gültigkeit verworfen wird, nämlich mit einer Wahrscheinlichkeit von $1-(1-0.01)^9 = 0.0865\%$.

Die Wahrscheinlichkeit, die Nullhypothese, das heißt, die Gültigkeit des Rasch-Modells irrtümlicherweise zu verwerfen hat sich somit beinahe um das Neunfache erhöht.

Um dem Entgegenzuwirken wird eine Alpha-Adjustierung durchgeführt. Bei der Alpha-Adjustierung wird das Signifikanzniveau der einzelnen Tests so angepasst, dass für alle Tests zusammengekommen das angestrebte Signifikanzniveau eingehalten wird. Nachteil dabei ist der Verlust an Teststärke (Power).

Die einfachste und am häufigsten durchgeführte Alpha-Adjustierung ist die Korrektur nach Bonferroni. Dabei wird das ursprüngliche Signifikanzniveau durch die Anzahl der Modelltests dividiert: $\alpha^* = \alpha / (\text{Anzahl der Tests})$.

In der hier vorliegenden Arbeit beträgt das adjustierte Alpha demnach: $\alpha^* = 0.01/9$, also 0.001. Die kritischen Chi-Quadrat Werte verändern sich ebenfalls. Diese werden wieder in der Tabelle von Bortz & Döring (2005) abgelesen und betragen bei $df = 17$ 40.79 und 39.52 bei $df = 16$.

In Tabelle 9 sind die Ergebnisse ausführlich dargestellt.

Tabelle 9: **Ergebnisse der Modelltests**

Signifikante Werte sind mit einem * gekennzeichnet.

Modelltest	Stichprobe N1 / N2	χ^2 - Wert	df
Geschlecht (männlich / weiblich)	375 / 285	22.01	17
Alter (12-16 / 17-27)	325 / 335	22.71	17
Muttersprache (deutsch / andere)	504 / 156	30.15	17
Schultyp (Berufsschule / Gesamtschule)	172 / 488	17.94	17
Mathematiknote (1-3 / 4-6)	371 / 270	29.44	17
Darstellende Geometrie (ja / nein)	98 / 562	25.27	17
Geometrisch Zeichnen (ja / nein)	202 / 458	20.85	17
Rohscore (Trennscore 10)	382 / 278	*91.62	17
Erstes Item (gelöst / nicht gelöst)	535 / 125	22.33	16

$\alpha = 1\%$, $df = 17$, kritischer Wert $\chi^2_{(.01/df=17)} = 33.41$

$\alpha = 1\%$, $df = 16$, kritischer Wert $\chi^2_{(.01/df=17)} = 31.99$

Nach Alpha-Adjustierung:

$\alpha = 0.1\%$, $df = 17$, kritischer Wert $\chi^2_{(.001/df=17)} = 40.79$

$\alpha = 0.1\%$, $df = 16$, kritischer Wert $\chi^2_{(.001/df=16)} = 39.25$

Somit zeigt sich in den Ergebnissen, dass von den neun Modelltests nur beim Teilungskriterium Rohscore ein signifikantes Ergebnis vorliegt, es demnach signifikante Unterschiede in den Itemparameterschätzungen der Teilstichproben gibt.

Wie im nächsten Abschnitt genau dargestellt, zeigt die grafische Modellkontrolle, dass es sich dabei jedoch nur um sehr geringe Abweichungen handelt, somit die Gültigkeit des Rasch-Modells nicht in Frage gestellt werden muss. Dies entspricht auch den Ergebnissen der Modelltests für den WMT-2 mit der österreichischen Stichprobe (Formann, Waldherr & Piswanger, in Druck). Auch hier wies nach Durchführung der Alpha-Adjustierung das Teilungskriterium Median einen signifikanten Wert auf, wobei die grafische Modellkontrolle keine relevanten Abweichungen erkennen ließ.

Müller-Phillip & Tarnai (1989, S. 239) meinten zur Annahme der Gültigkeit eines Modells: „Geht man davon aus, dass eine völlige Übereinstimmung von Modell und Daten aufgrund der idealisierten Abbildung der Realität durch das Modell grundsätzlich unmöglich ist, dann muss sich die Frage nach der Gültigkeit eines Modells danach richten, welche Modellabweichungen tolerierbar, d.h. wegen ihrer praktischen Relevanz vernachlässigbar sind. Dieses allgemeine Problem statistischer Tests wird beim Rasch-Modell besonders deutlich,...“.

9.4 Grafische Modellkontrolle

Eine weitere Möglichkeit, um die Gültigkeit des Rasch-Modells zu überprüfen, ist die Grafische Modellkontrolle (Rasch, 1960). Dazu werden die mittels LPCM-Win Itemparameter der beiden Gruppen in einem Diagramm dargestellt. Weichen die Daten nur minimal von der imaginären 45° Geraden ab, passen sie gut zum Rasch-Modell und sind somit Rasch-homogen. Betrachtet man die Ergebnisse der Modelltests, so sollte es nur bei der Grafischen Modellkontrolle „Rohscore“ Abweichungen der Items von der 45° Geraden geben.

Es lässt sich erkennen, dass es nur geringe Abweichungen der Items im Modelltest „Rohscore“ von der 45° Geraden gibt. Auch die anderen Modelltests weisen eine gute Anpassung auf, lediglich beim Modelltest „Darstellende Geometrie“ gibt es

etwas größere Abweichungen. Da jedoch der χ^2 -Wert deutlich unter dem kritischen Wert liegt, können die Abweichungen der Grafischen Modellkontrolle vernachlässigt werden.

In Folge sind die grafischen Modellkontrollen für alle neun Modelltests dargestellt.

Abb. 6: Grafische Modellkontrolle Geschlecht

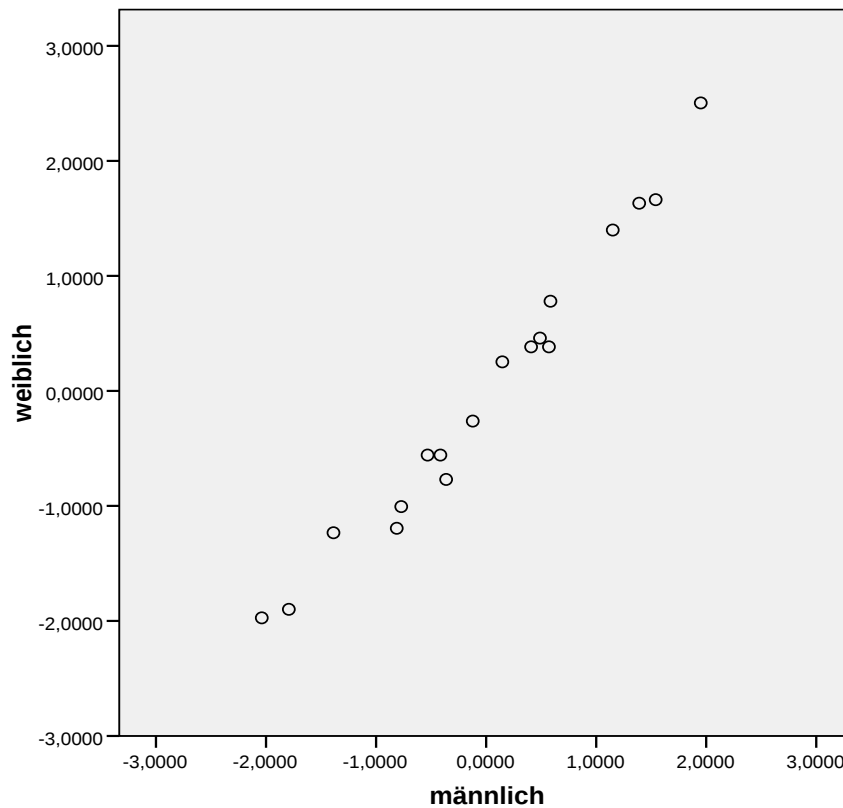


Abb. 7: Grafische Modellkontrolle Alter

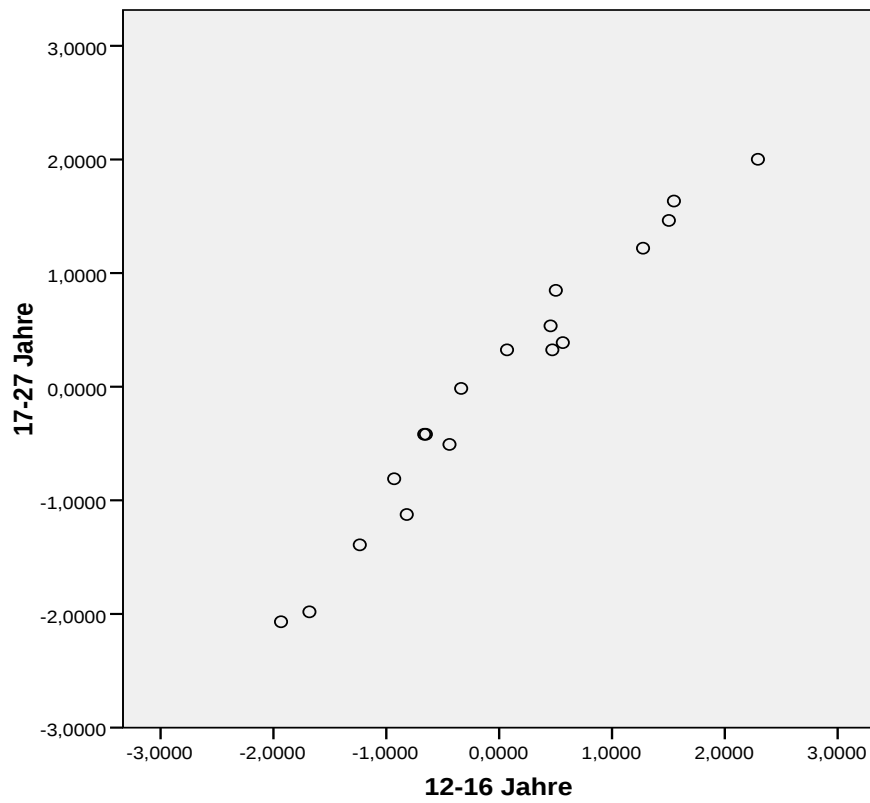


Abb. 8: Grafische Modellkontrolle Muttersprache

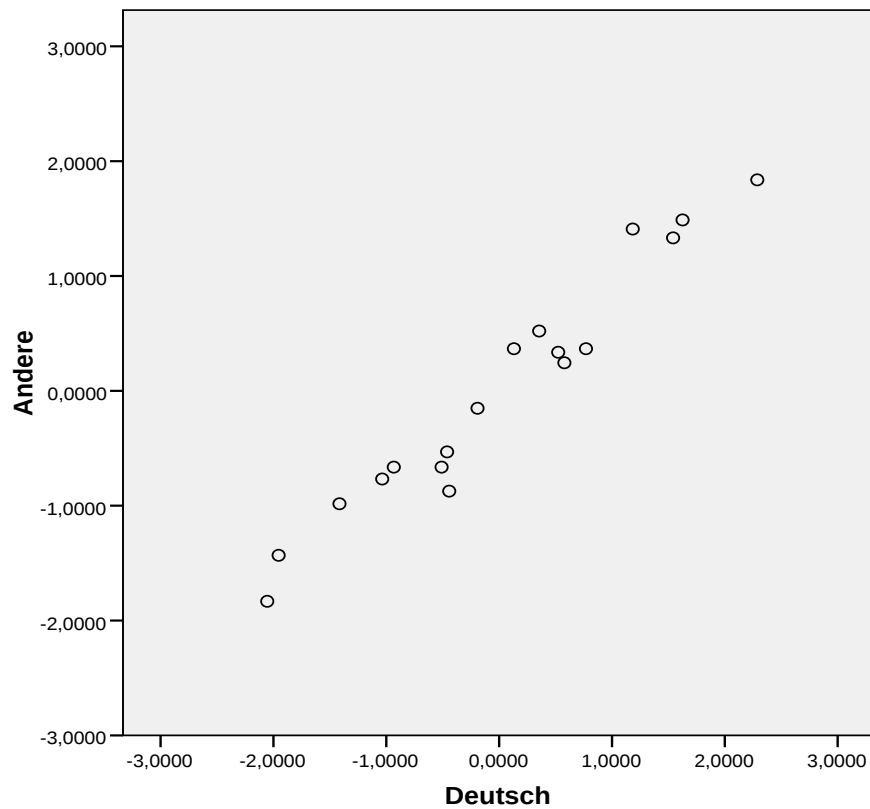


Abb. 9: Grafische Modellkontrolle Schultyp

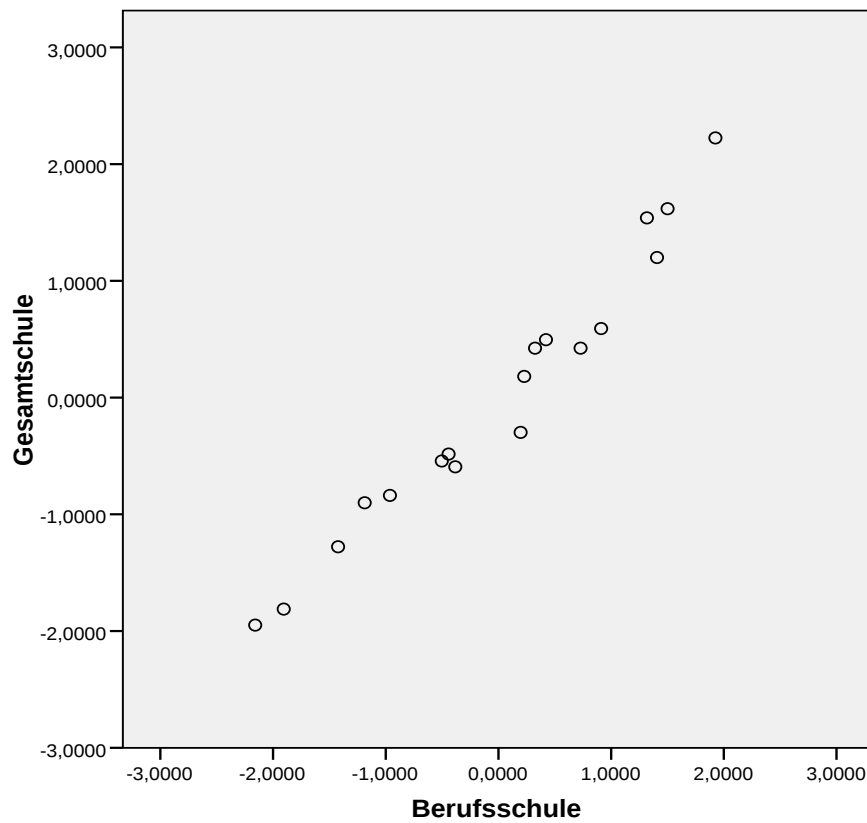


Abb. 10: Grafische Modellkontrolle Mathematiknote

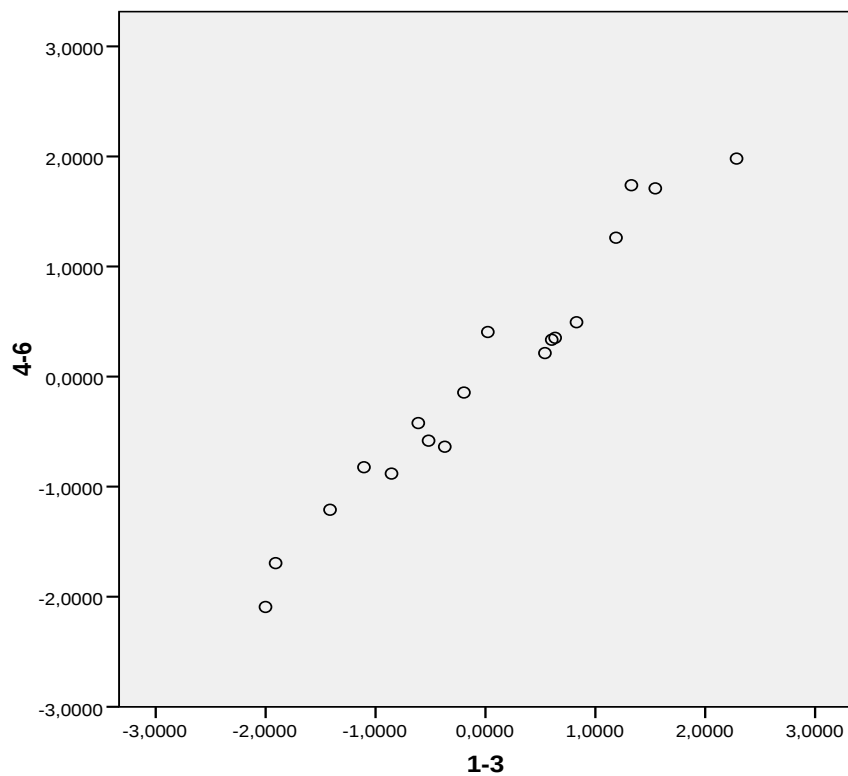


Abb. 11: Grafische Modellkontrolle Darstellende Geometrie

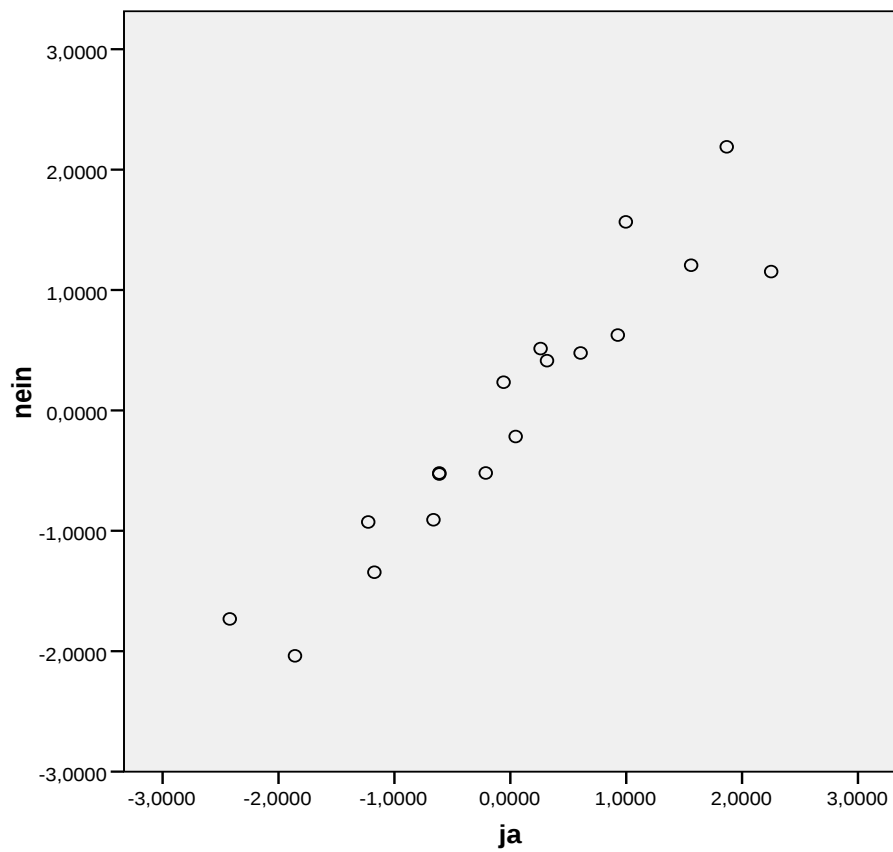


Abb. 12: Grafische Modellkontrolle Geometrisch Zeichnen

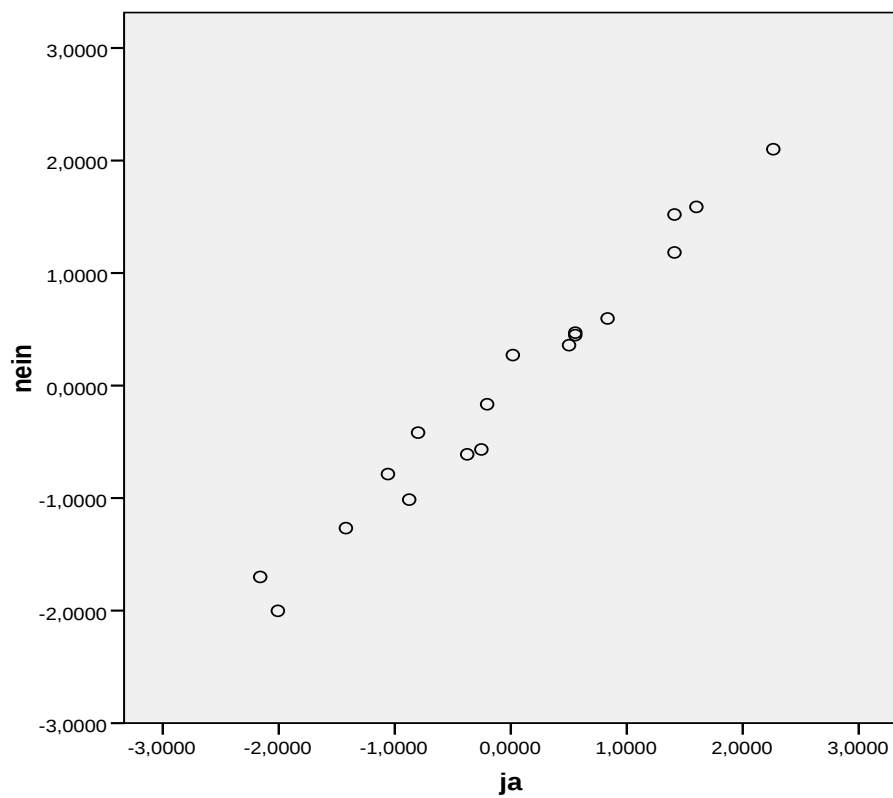


Abb. 13: Grafische Modellkontrolle Rohscore

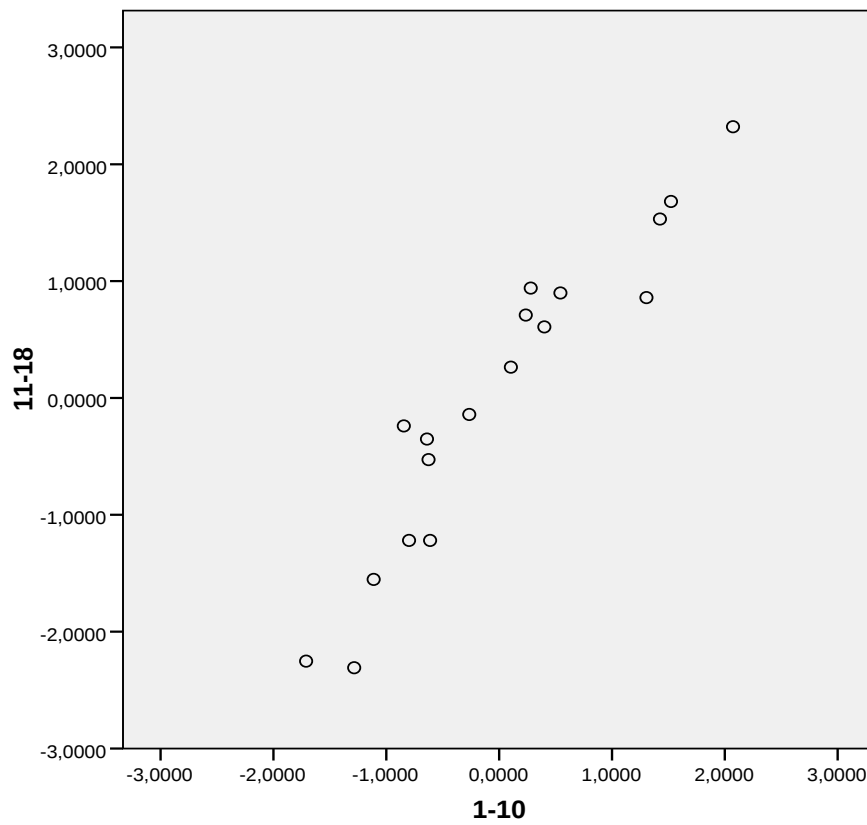
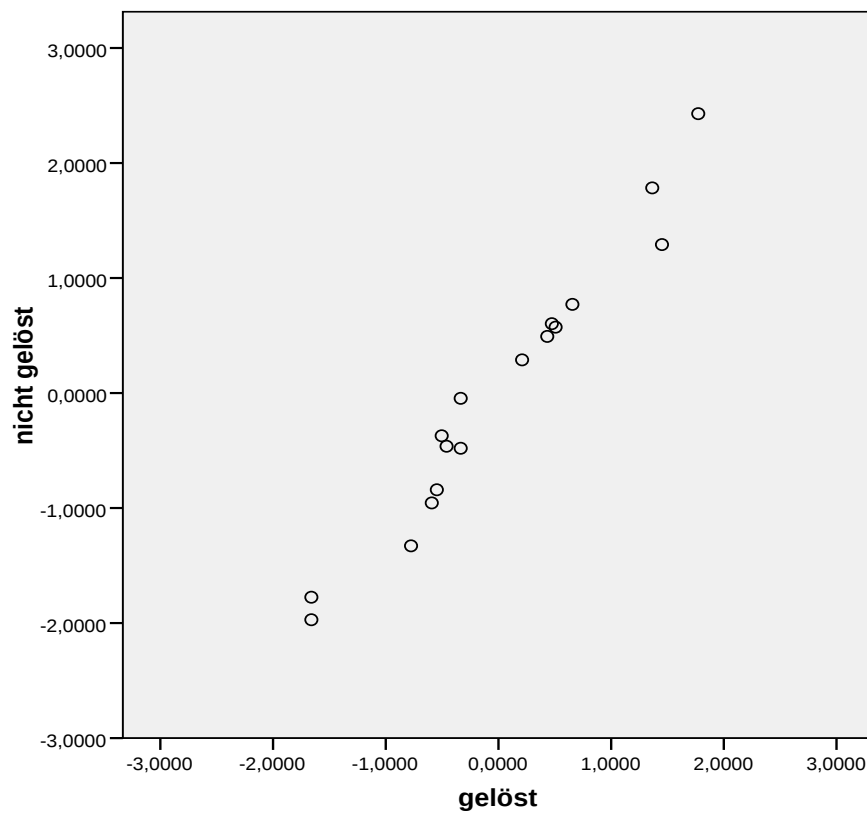


Abb. 14: Grafische Modellkontrolle Item 1



9.5 Modelltests Berlin – Österreich

Um zu überprüfen, ob der WMT-2 für die Berliner Stichprobe dieselbe latente Dimension erfasst wie für die österreichische Stichprobe, sollten nun die Itemparameterschätzungen der beiden Stichproben miteinander verglichen werden. Dazu wurde der Datensatz aus Berlin mit der österreichischen Normierungsstichprobe verglichen. Die Übereinstimmung der beiden Länder wurde in verschiedenen Teilstichproben überprüft:

1. Alter
 - 14-16 Jahre
 - 17-18 Jahre
 - 19-25 Jahre
2. Schultyp
 - Gesamtschule / AHS, BHS
 - Berufsschule

Teilungskriterium war dabei immer die Nationalität, also Berlin und Österreich. Dabei wurden diejenigen Personen aus der Stichprobe herausgenommen, die nicht dem jeweiligen Kriterium des Modelltests entsprachen. Stichprobe N1 und N2 stehen somit für die jeweilige Anzahl an Berliner und österreichischen Testpersonen. Auch hier wurde, um das Alpha-Risiko zu reduzieren, eine Alpha-Adjustierung durchgeführt. Jedoch wiesen sich auch nach Durchführung der Alpha-Adjustierung zwei Modelltests signifikante χ^2 Werte auf.

Tabelle 10 zeigt eine Übersicht der Ergebnisse für die jeweiligen Modelltests.

Tabelle 10

Modelltests Alter	Stichprobe N1 / N2	χ^2 - Wert	df
14-16 Jahre	325 / 722	40.45	17
17-18 Jahre	198 / 528	*49.74	17
19-15 Jahre	136 / 91	26.03	17
Modelltests Schultyp	Stichprobe N1 / N2	χ^2 - Wert	df
Berufsschule	42 / 661	20.11	17
Gesamtschule, AHS / BHS	481 / 628	*77.64	17

$\alpha = 1\%$, $df = 17$, kritischer Wert $\chi^2_{(.01/df=17)} = 33.41$

Nach Alpha-Adjustierung:

$\alpha = 0.1\%$, $df = 17$, kritischer Wert $\chi^2_{(.001/df=17)} = 40.79$

9.6 Grafische Modellkontrolle

Auch hier wurde eine grafische Modellkontrolle durchgeführt, um mögliche Abweichungen der Items zu veranschaulichen.

Abb. 15: Grafische Modellkontrolle Alter 14-16 Jahre

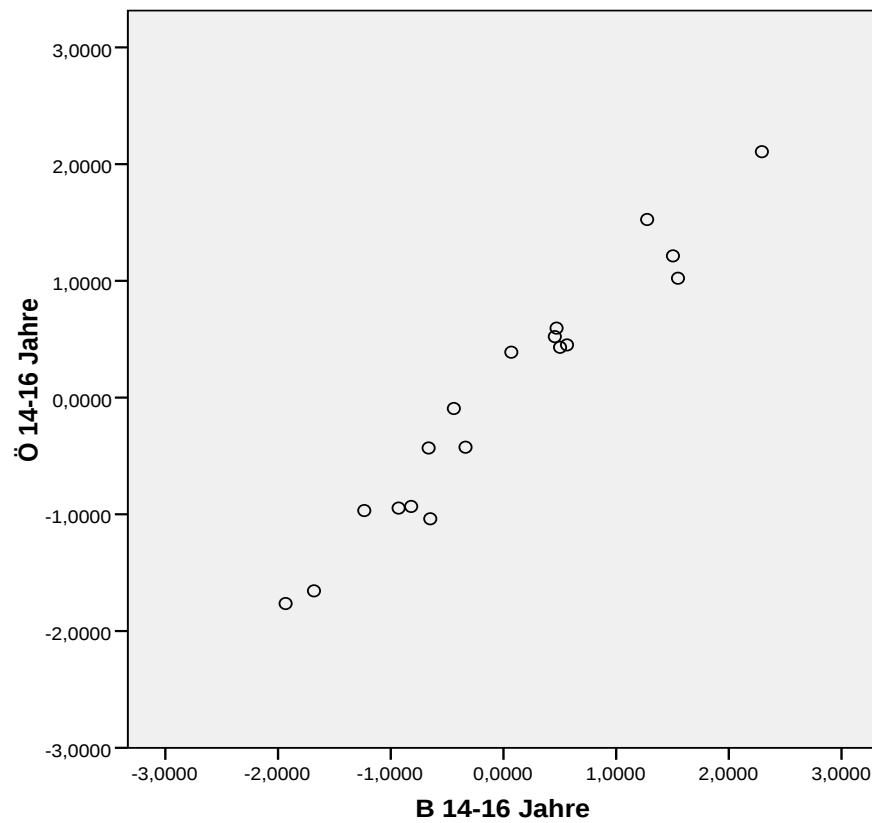


Abb. 16: Grafische Modellkontrolle Alter 17-18 Jahre

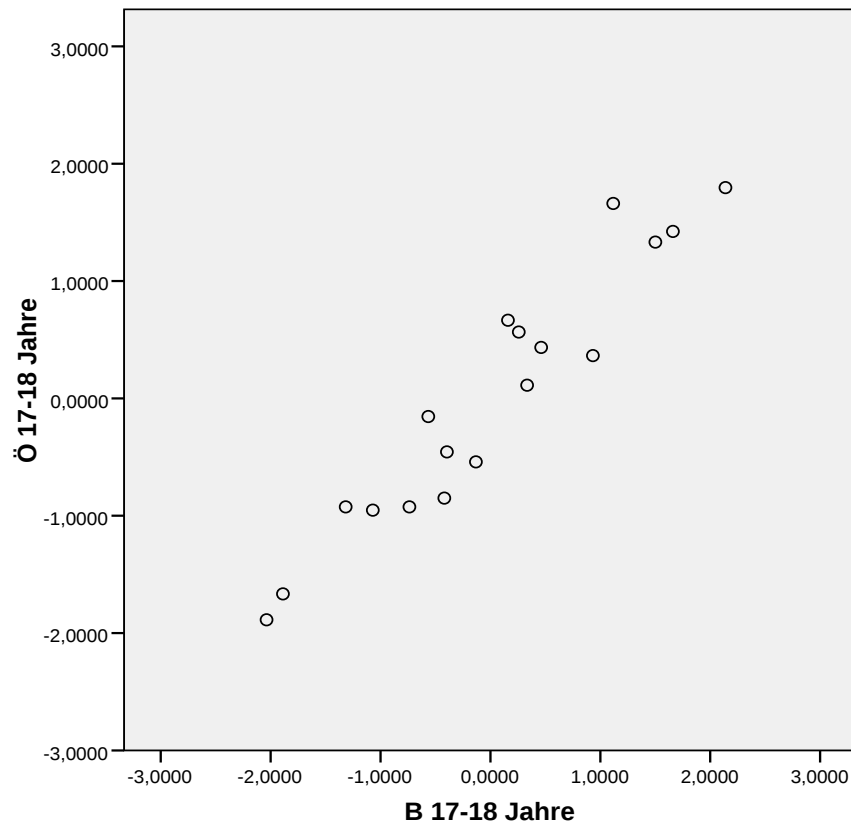


Abb. 17: Grafische Modellkontrolle Alter 19-25 Jahre

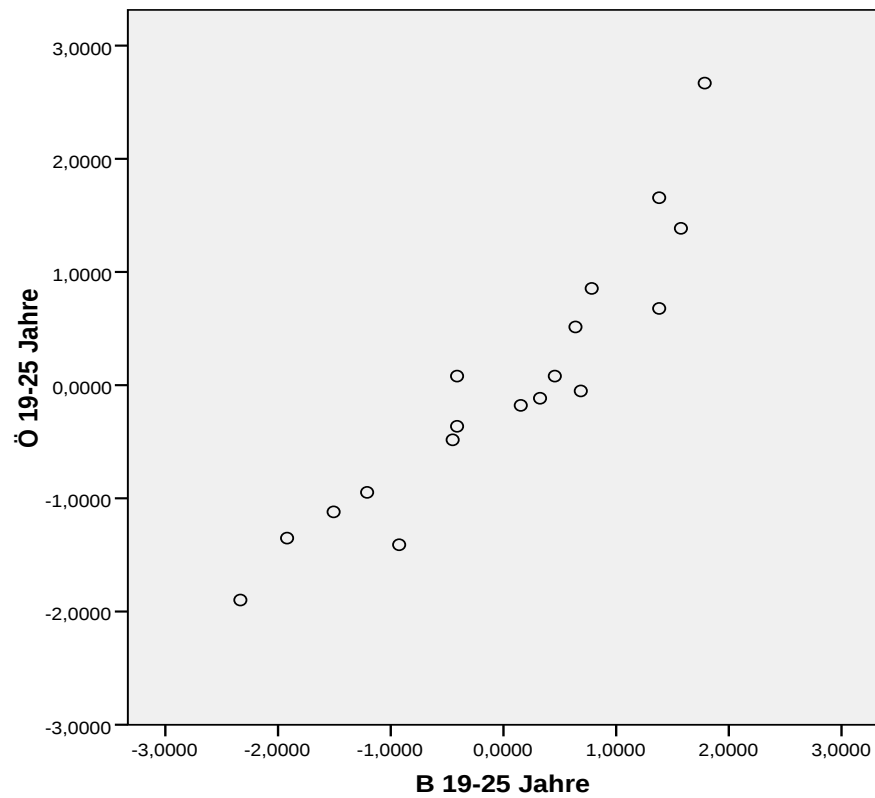


Abb. 18: Grafische Modellkontrolle Schultyp Berufsschule

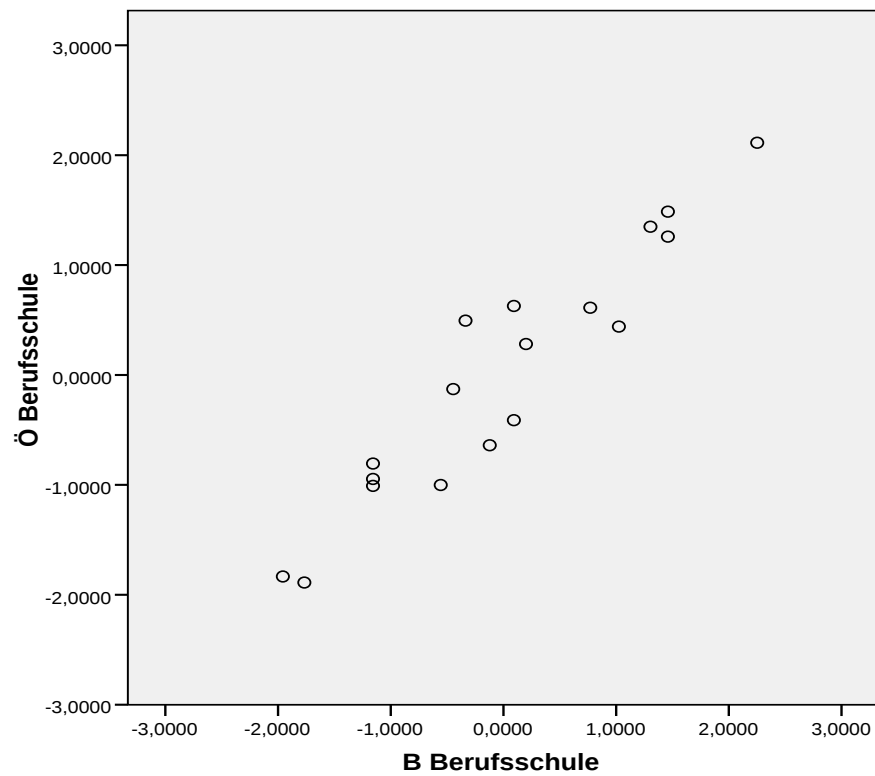
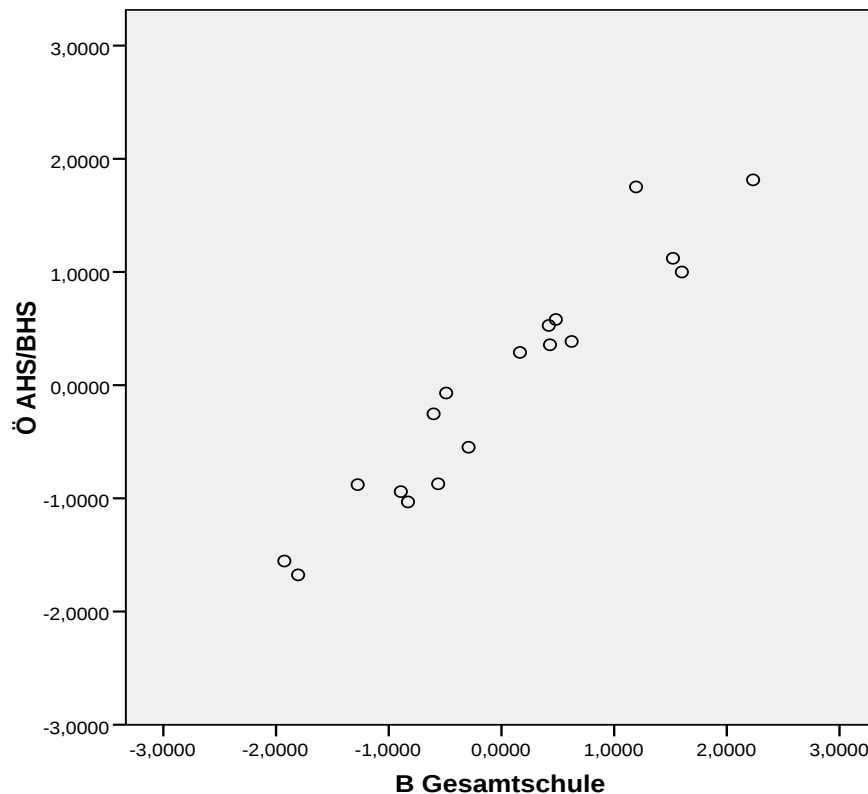


Abb. 19: Grafische Modellkontrolle Schultyp Gesamtschule



Die Grafiken zeigen in allen Modelltests Abweichungen der Items von der 45° Geraden. Diese sind jedoch eher gering, sodass von der Gültigkeit des Rasch-Modells ausgegangen werden kann, zumal auch die Likelihoodquotienten-Tests nicht signifikant sind.

9.7 Zusammenfassung der Modelltests

Insgesamt wurden in dieser Arbeit neun Modelltests mit dem Berliner Datensatz durchgeführt, von denen acht für die Gültigkeit des Rasch-Modells sprechen. Der Modelltest nach dem internen Kriterium Rohscore stellte sich auch nach Durchführung der Alpha-Adjustierung als signifikant heraus. Bei der grafischen Modellkontrolle zeigten sich hingegen nur geringfügige Abweichungen. Auch die Ergebnisse der Modelltests des WMT-2 mit der österreichischen Stichprobe decken sich mit den hier gefundenen Ergebnissen. Das Problem, dass es eine vollständige Übereinstimmung zwischen Modell und Daten nicht geben kann, beschreiben Müller-Phillip & Tarnai (1989).

Die Schlussfolgerung ist daher, dass die Nullhypothese und somit die Gültigkeit des Rasch-Modells, angenommen werden kann.

Von den fünf Modelltests, die zum Vergleich Berlin/Österreich durchgeführt wurden, weisen zwei ein signifikantes Ergebnis auf. Dabei handelt es sich um die Modelltests die in den Teilstichproben Alter 17-18 Jahre und Schultyp Gesamtschule durchgeführt wurden. Wie bereits erwähnt, zeigt die grafische Modellkontrolle jedoch tolerierbare Abweichungen, sodass die Gültigkeit des Rasch-Modells nicht in Frage gestellt werden muss.

9.8 Dimensionalitätsprüfung nach Martin – Löff

Die Dimensionalitätsprüfung nach Martin – Löff (1973) soll Aufschluss darüber geben, ob zwei hypothetische Itemteilmengen gleich sind. Es soll die Homogenität der beiden aufgrund von a priori Hypothesen gebildeten Itemteilmengen überprüft werden.

Dazu müssen Hypothesen vorliegen, dass zwei aufgrund inhaltlicher Überlegungen getrennte Itemteilmengen unterschiedliche Fähigkeitsdimensionen ansprechen.

Da die Items im WMT-2 anhand der formalen Aufbauregeln der Items in jeweils zwei Itemteilmengen getrennt werden können, konnte eine Dimensionalitätsprüfung nach Martin – Löff erfolgen.

Bei der Bestimmung der Itemteilmengen orientierte ich mich an die bei der Dimensionalitätsprüfung nach Martin – Löff im WMT-2 festgelegten Hypothesen:

1. Richtung der Regelgeltung

Dabei wurden jene Items verglichen, die eine waag- und senkrechte Regelgeltung hatten, mit den Items, die nur waag- oder senkrecht aufgebaut waren. Es sollte somit untersucht werden, ob die Richtung der Regelgeltung unterschiedliche Fähigkeitsdimensionen anspricht.

2. Geltende Aufbauregel Überlagern

Hier wurden die Items dahingehend getrennt, ob in ihrem Aufbau die Regel Überlagern vorkommt oder nicht. Es sollte also wiederum untersucht werden, ob sich die beiden Itemteilmengen hinsichtlich der angesprochenen Fähigkeitsdimension unterscheiden.

3. Geltende Aufbauregel Variieren

Ebenso wie bei der Aufbauregel Überlagern erfolgte die Itemtrennung nach der Aufbauregel Variieren nach ihrem Vorkommen. Demzufolge unterteilten sich die Items in jene, bei denen Variieren vorkommt und jene, bei denen Variieren nicht

vorkommt. Auch hier sollten mögliche unterschiedliche Fähigkeitsdimensionen erfasst werden.

4. Schwierigkeit der Items

Entsprechend dieser Hypothese wurden die Items geteilt in leichtere (Items 1-9) und schwerere (Items 10-18). Somit sollte auch hier überprüft werden, ob Eindimensionalität dieser beiden Itemteilmengen vorliegt.

In der unten stehenden Tabelle sind die Ergebnisse der Modelltests aller Hypothesen angeführt.

Tabelle 11: Dimensionalitätsprüfung nach Martin – Löf, kritischer Wert $\chi^2_{(.01/df=76-80)} = 124.84$

Itemsubgruppen	χ^2 - Wert	df
Richtung der Regelgeltung: waag- und senkrecht vs. nur waag- oder senkrecht	75.30	79
Aufbauregel Überlagern: kommt nicht vor vs. kommt mind. 1x vor	119.81	80
Aufbauregel Variieren: kommt nicht vor vs. kommt mind. 1x vor	60.74	76
Schwierigkeit der Items: leichtere vs. schwerere	74.72	80

Wie man anhand der Tabelle erkennen kann, weisen alle Modelltests ein nicht signifikantes Ergebnis auf. Aufgrund der Dimensionalitätsprüfung nach Martin – Löf kann daher die Eindimensionalitätshypothese der Items beibehalten werden. Es werden keine unterschiedlichen Fähigkeitsdimensionen bei den einzelnen Itemteilmengen angesprochen.

10. Mittelwertsvergleiche mit der österreichischen Normierungsstichprobe

In diesem Abschnitt soll überprüft werden, inwieweit die für diese Arbeit erhobenen Daten mit denen der Normierungsstichprobe des WMT-2 übereinstimmen. Die Daten der österreichischen Stichprobe stammen daher von der Normierungsstichprobe des WMT-2. Um zu überprüfen, ob sich die beiden Stichproben hinsichtlich der Testleistungen signifikant voneinander unterscheiden, wurden Mittelwertsvergleiche folgender Teilstichproben durchgeführt:

- 1) - Österreichische BerufsschülerInnen 14-16 Jahre
- Berliner BerufsschülerInnen 14-16 Jahre
- 2) - Österreichische BerufsschülerInnen 17-18 Jahre
- Berliner BerufsschülerInnen 17-18 Jahre
- 3) - Österreichische AHS/BHS SchülerInnen 14-16 Jahre
- Berliner GesamtschülerInnen 14-16 Jahre
- 4) - Österreichische AHS/BHS SchülerInnen 17-18 Jahre
- Berliner GesamtschülerInnen 17-18 Jahre
- 5) - Österreichische Erwachsene 19-25 Jahre
- Berliner BerufsschülerInnen 19-25 Jahre

Zuerst erfolgte die Prüfung der Rohscores der einzelnen Teilstichproben auf Normalverteilung mittels Komogorov-Smirnov-Test. Dieser brachte signifikante Abweichungen von der Normalverteilung bei den Teilstichproben.

In weiterer Folge sollten mittels Varianzanalyse mögliche Unterschiede in den Testleistungen der beiden Stichproben Berlin/Österreich erhoben werden. Da die Homogenität der Varianzen mittels Levene-Test angenommen werden kann, können die Ergebnisse der Varianzanalyse trotz Verletzung der Normalverteilung unter Beachtung der unten angeführten Unterschiede interpretiert werden.

Um einen Überblick über die einzelnen Teilstichproben und deren Mittelwerte zu haben, werden diese in Tabelle 3 dargestellt. Zusätzlich sind neben den Mittelwerten die Standardabweichungen angeführt.

Tabelle 12:

	N (Ö)	N (B)	Mittel= wert (Ö)	Mittel= wert (Berlin)	Standard= abweichung (Ö)	Standard= abweichung (B)
BerufsschülerInnen bis 16 Jahre	432	4	9.23	11.00	3.54	3.37
BerufsschülerInnen 17-18 Jahre	230	38	10.14	8.76	3.36	3.01
AHS/BHS- bzw. GesamtschülerInnen bis 16 Jahre	329	321	11.11	9.34	3.33	3.44
AHS/BHS- bzw. GesamtschülerInnen 17-18 Jahre	299	160	12.50	10.03	3.44	3.50
Alle BerufsschülerInnen	661	42	9.55	8.98	3.51	3.08
Alle AHS/BHS- bzw. GesamtschülerInnen	628	481	11.77	9.58	3.44	3.47
Erwachsene ohne Matura 19-25 Jahre	91	137	12.13	10.66	3.81	4.12
Gesamtstichprobe	1380	660	10.73	9.76	3.68	3.62

Wie anhand der Tabelle zu erkennen ist, weisen die österreichischen Teilstichproben allesamt höhere Mittelwerte auf. Die Gruppe der Berliner BerufsschülerInnen unter 16 Jahre besteht leider nur aus vier SchülerInnen, sodass eine Interpretation des Ergebnisses nicht zulässig ist.

Bevor die Ergebnisse der Varianzanalyse dargestellt werden, sollen die für die Interpretation wichtigen Unterschiede kurz beschrieben werden:

Bei der Interpretation der Varianzanalyse ist zu beachten, dass sich die Berliner Stichprobe in wesentlichen Punkten von der österreichischen Stichprobe unterscheidet. Nicht nur, dass die Berliner Stichprobe wesentlich kleiner ist, auch die Verteilung hinsichtlich der Teilstichproben ist sehr unterschiedlich. So beträgt der Anteil an Berliner Erwachsenen im Alter von 19-25 Jahren 21.6% der Berliner Gesamtstichprobe, wohingegen die österreichischen Erwachsenen derselben Altersgruppe nur 6.6% der österreichischen Gesamtstichprobe ausmachen. Eine ebenso ungleiche Verteilung ist in den beiden anderen Altersgruppen zu beobachten. Die genaue Verteilung der einzelnen Teilstichproben ist in Tabelle 1, Seite 23 ersichtlich.

Die eben genannten Unterschiede der Stichproben und auch die bereits in Abschnitt 6 genannten Unterschiede im Schulsystem von Österreich und Berlin lassen nur eine eingeschränkte Interpretation der Ergebnisse der Varianzanalyse zu. Ebenso wichtig zu erwähnen ist, dass ein Vergleich Berlin/Wien aufgrund fehlender Daten leider nicht möglich war und ein Vergleich Berlin/Österreich somit eigentlich unzulässig ist. Dennoch sollen die Ergebnisse kurz dargestellt werden, wobei noch zu erwähnen ist, dass Vergleiche nur in denselben Altersgruppen und denselben Schultypen erfolgten.

10.1 Ergebnisse

Hinsichtlich der verschiedenen Altersgruppen gibt es einen signifikanten Unterschied bei den 17-18 jährigen SchülerInnen. Hier haben die österreichischen SchülerInnen mit einem Mittelwert von 12.50 einen wesentlich besseren Wert als die Berliner 17-18 jährigen SchülerInnen (10.03). In den beiden andern Altersgruppen gibt es keine signifikanten Unterschiede.

Ähnlich verhält es sich bei den unterschiedlichen Schultypen: hier haben die österreichischen AHS-/ BHS- SchülerInnen mit einem Mittelwert von 11.77 ebenfalls einen signifikant besseren Wert als die Berliner GesamtschülerInnen mit einem Mittelwert von 9.58.

Die detaillierte Darstellung der Ergebnisse der Varianzanalyse ist dem Anhang beigelegt.

11. Normierung

Lienert beschreibt den Begriff der Normierung folgendermaßen: „Unter den Gütekriterien der Normierung versteht man, dass über einen Test Angaben vorliegen sollen, die als Bezugssystem für die Einordnung des individuellen Testergebnisses dienen können.“ (Lienert & Raatz, 1998, S. 11)

Da der WMT-2 auch international eingesetzt werden soll, erschien es sinnvoll, folgende Normtabellen zu erstellen:

- GesamtschülerInnen 14-16 Jahre
- GesamtschülerInnen 17-18 Jahre
- BerufsschülerInnen 19-25 Jahre

Leider konnten aufgrund der zu kleinen Anzahl an Testpersonen keine weiteren Normtabellen erstellt werden.

Bei der Erstellung der Normen wurde dabei folgendermaßen vorgegangen:

Zuerst wurde in allen Gruppen die Verteilung der Rohscores auf Normalverteilung geprüft. Da sich eine signifikante Abweichung der Rohscores in allen Eichgruppen zeigte, sind in den Normtabellen die Prozentränge (PR) angegeben. Die Prozentränge können nämlich auch aus nicht-normalverteilten Testwerten ermittelt werden, indem sie aus der Häufigkeitsfunktion der Eichstichprobe empirisch bestimmt werden.

Mittels der Prozentränge kann der relative Anteil von Personen in der Referenzpopulation bestimmt werden, die denselben oder einen niedrigeren Testwert erreichen.

Weiters werden aufgrund der nicht-normalverteilten Rohscores nur die Stanine-Werte angegeben. Der Wertebereich liegt dabei zwischen 1 und 9. Ein Wert von 5 beutet eine durchschnittliche Leistung.

Die genaue Darstellung der Normtabellen ist im Anhang dieser Arbeit ersichtlich.

12. Testgütekriterien

Ein weiteres Ziel meiner Arbeit war es, die Hauptgütekriterien Objektivität, Reliabilität und Validität zu überprüfen. Dazu werden die wichtigsten Testkennwerte dargestellt und mit den Werten im Testmanual des WMT-2 verglichen.

12.1 Objektivität

Da die Instruktion des WMT-2 im Testmanual genau vorgegeben ist, kann Durchführungsobjektivität angenommen werden. Ebenso erfolgt die Auswertung mittels des im Testmanual vorgegebenen Lösungsvektors. Somit liegt Auswertungsobjektivität vor. Auch die Interpretation der Ergebnisse ist im Testmanual festgelegt, womit Interpretationsobjektivität angenommen werden kann.

12.2 Reliabilität

Um zu überprüfen, ob der WMT-2 „innere Konsistenz“ zeigt, wurde Cronbachs Alpha berechnet.

Das Alpha beträgt hier 0.759 und unterscheidet sich somit kaum vom Ergebnis von Formann, Waldherr & Piswanger (in Druck), welches mit 0.819 als zufriedenstellend anzusehen ist. Die Tabelle der Item-Skala-Statistiken zeigt, dass das Weglassen keines der 18 Items den Reliabilitätskoeffizienten wesentlich verbessern oder verschlechtern würde. Die Itemtrennschärfen liegen zwischen 0.128 und 0.465.

Tabelle 13: **Reliabilitätsstatistiken**

	WMT-2
Splithalf-Reliabilität	0.714 (Guttman`s Koeffizient)
Cronbachs Alpha	0.759

Tabelle 14: **Item-Skala-Statistiken**

	Skalenmittel wert, wenn Item weggelassen	Skalenvarianz, wenn Item weggelassen	Korrigierte Item-Skala- Korrelation	Cronbachs Alpha, wenn Item weggelassen
Item01	8,95	12,169	,281	,752
Item02	8,87	12,266	,323	,750
Item03	8,93	12,056	,336	,749
Item04	9,11	11,627	,379	,745
Item05	8,98	12,189	,251	,754
Item06	9,11	11,362	,465	,737
Item07	9,08	11,668	,378	,745
Item08	9,24	11,502	,395	,743
Item09	9,17	11,609	,370	,745
Item10	9,13	11,416	,442	,739
Item11	9,38	12,106	,223	,758
Item12	9,47	12,082	,253	,755
Item13	9,32	11,665	,348	,747
Item14	9,30	11,326	,453	,738
Item15	9,40	11,937	,280	,753
Item16	9,32	11,352	,446	,739
Item17	9,55	12,555	,128	,763
Item18	9,57	12,282	,240	,755

12.3 Validität

Bereits in früheren Arbeiten wurde gezeigt, dass der WMT hohe Korrelationen mit allgemeinen Intelligenztests erzielt. Der Produkt-Moment-Korrelationskoeffizient bei der Berechnung der Korrelation des WMT mit dem Matrizenstest von Raven (1938) betrug 0.92. Einen ebenfalls hohen Wert erzielte die Berechnung der Korrelation von WMT und Intelligenz-Struktur-Test (IST, Amthauer, 1973). Der Produkt-Moment-Korrelationskoeffizient betrug hier 0.85.

Beide Ergebnisse zeigen, dass der WMT ähnlich wie der Matrizenstest von Raven (1938) analytische, ordnende Denkopoperationen erfasst und ähnlich wie der IST „Schlussfolgerndes Denken“ misst.

Weiters wurde in früheren Untersuchungen ein möglicher Zusammenhang zwischen WMT und Dreidimensionalem Würfelstest (3DW, Gittler, 1990) ermittelt. Der Korrelationskoeffizient von $r(\text{WMT}/3\text{DW}) = 0.69$ bestätigt einen Zusammenhang

zwischen WMT und 3DW und kann dahingehend interpretiert werden, dass Matrizentests auch mit dem dritten Sekundärfaktor „Visualisierung“ nach Cattell (1966) korrelieren.

Zusammenfassend kann somit gesagt werden, dass der WMT zur Messung von grundlegenden Teilbereichen der Intelligenz geeignet ist.

Zur Überprüfung der „konvergenten Validität“ des WMT-2 wurde im Rahmen dieser Arbeit die Korrelation mit dem Zahlenverbindungstest (ZVT) von Oswald & Roth (1987) berechnet.

Der ZVT ist laut Manual „ein spezifischer Intelligenztest zur Messung der kognitiven Leistungsgeschwindigkeit“ (Oswald & Roth, 1987) und korreliert hoch mit allgemeiner Intelligenz. Im Gegensatz zum WMT-2 ist der ZVT ein reiner „Speed-Test“ und misst in Anlehnung an Thurstone „Perceptual Speed“, also die „Informations-Verarbeitungsgeschwindigkeit“ (Oswald & Roth, 1987).

Der Grund, warum dieser Test ausgewählt wurde, ist die Testökonomie. Lienert (Lienert & Ratz, 1998, S. 12) beschreibt Testökonomie folgendermaßen: „Ein Test ist dann ökonomisch, wenn er: eine kurze Durchführungszeit beansprucht, wenig Material verbraucht, einfach zu handhaben, als Gruppentest durchführbar, schnell und bequem auszuwerten ist“. Da der ZVT all diese Kriterien erfüllt, war es möglich, beide Tests in einer Schulstunde (45 Minuten) vorzugeben.

Sowohl der Korrelationskoeffizient nach Spearman mit $r(\text{WMT-2/ZVT}) = 0.286$, als auch der Korrelationskoeffizient nach Pearson mit $r(\text{WMT-2/ZVT}) = 0.293$ zeigt, dass der WMT-2 nur schwach mit dem ZVT korreliert.

Der Grund dafür liegt daran, dass wie bereits erwähnt zwar beide den Intelligenztests zuzuordnen sind, jedoch andere Bereiche der Intelligenz erfassen. So misst der WMT-2 nach Thurstones Intelligenzfaktoren den Faktor „Reasoning“ und der ZVT den Faktor „Perceptual Speed“.

Eine weitere Untersuchung von Formann (1973) und Piswanger (1975) zeigte, dass die Mathematiknote einen Einfluss auf die Leistung im WMT hat. Der Korrelationskoeffizient betrug in ihren Arbeiten -0.294 und -0.423, womit ein schwacher bis mittlerer Zusammenhang zwischen hoher oder niedriger Leistung im WMT und der Mathematiknote besteht. Das negative Vorzeichen sagt aus, dass eine hohe Testleistung mit einer niedrigen Mathematiknote korreliert und umgekehrt.

Auch spätere Untersuchungen zeigten, dass die Mathematiknote einen Einfluss auf die Testleistung im WMT hat (Gittler, 1990). SchülerInnen mit „sehr gut“ in Mathematik haben signifikant bessere Testleistungen im WMT als SchülerInnen mit schlechteren Noten.

Um diese früheren Untersuchungen zu bestätigen, wurde in dieser Arbeit ebenfalls ein möglicher Zusammenhang zwischen Mathematiknote und der Testleistung im WMT-2 berechnet. Dazu wurde die Stichprobe in gute MathematikschülerInnen (Note 1-3) und schlechte MathematikschülerInnen (Note 4-6) geteilt.

Die Ergebnisse zeigen mit einem Korrelationskoeffizienten nach Pearson von $r(\text{WMT}/\text{Mathenote}) = -0.29$ und einem Korrelationskoeffizienten nach Spearman von $r(\text{WMT}/\text{Mathenote}) = -0.273$, dass es einen Zusammenhang gibt. Auch hier gilt: ein hoher Rohscore im WMT korreliert (jedoch nur schwach) mit einer guten Mathematiknote.

Frühere Untersuchungen konnten somit bestätigt werden, wonach eine gute Mathematiknote (1-3) mit einer hohen Leistung im WMT-2 korreliert und umgekehrt. Da jedoch nur ein schwacher Zusammenhang nachgewiesen wurde, lässt sich nicht sagen, dass eine gute Mathematiknote auch zwingend zu einer guten Testleistung im WMT führt.

13. Zusammenfassung und Ausblick

Aufgrund der Tatsache, dass die Veröffentlichung des WMT (Formann & Piswanger, 1979) nun beinahe 30 Jahre zurückliegt, erschien es notwendig, eine Neunormierung des WMT zu machen. Weiters war es interessant, die Eindimensionalitätshypothese der Items mittels nun zur Verfügung stehender Modelltests zu überprüfen. Auch die sozialen und bildungspolitischen Veränderungen der letzten 30 Jahre waren ein Anstoß, um eine Neunormierung zu veranlassen. (Formann, Waldherr & Piswanger, in Druck)

Nach umfangreichen Modelltests wurden sechs Items ausgeschlossen. Somit besteht die revidierte Fassung des WMT aus 18 Items, die geeignet sind, um „einen Eindruck vom allgemeinen intellektuellen Leistungsniveau einer Person zu erhalten“. (Formann, Waldherr & Piswanger, in Druck, S. 3)

Die österreichische Normstichprobe umfasste 2494 Personen und wurde im Gegensatz zum WMT um Erwachsene im Alter zwischen 19 und 64 Jahren erweitert.

Mit der revidierten Fassung des WMT steht somit ein sehr ökonomischer, reliabler und valider Kurztest zum sprachfreien Screening der intellektuellen Leistungsfähigkeit für den Altersbereich von 14-64 Jahren zur Verfügung. (Formann, Waldherr, Piswanger, in Druck, S. 3)

In dieser Arbeit sollte überprüft werden, ob auch beim WMT-2 in einer zu Österreich sehr ähnlichen Kultur die Eindimensionalitätshypothese der Items gilt bzw. ob der Test Rasch-homogen ist. Dazu wurden in einem Zeitraum von Februar bis April 2008 insgesamt 728 Berliner SchülerInnen an Gesamt- und Berufsschulen getestet. Ursprünglich war geplant die Testungen auch an Gymnasien durchzuführen, jedoch war die Kooperation von Seiten der Schulleiter nicht gegeben. Daher beschränkte sich die Datenerhebung auf Gesamt- und Berufsschulen. Aufgrund mangelnder Motivation beim Bearbeiten der Testaufgaben mussten 68 Testhefte ausgeschlossen werden, sodass eine Gesamtstichprobe von 660 SchülerInnen verblieb.

Um die Rasch-Homogenität zu überprüfen, wurde das Programm LPCM-Win (Fischer & Ponocny-Seliger, 1998) verwendet. Insgesamt wurden neun Modelltests

anhand vorher festgelegter Teilungskriterien durchgeführt. Anschließend wurde eine Alpha-Adjustierung nach Bonferroni durchgeführt, um das Alpha-Fehler Risiko zu reduzieren.

Auch nach Durchführung der Alpha-Adjustierung erwies sich der Modelltest Rohscore als signifikant. Im Anschluss an die Modelltests wurde eine Grafische Modellkontrolle durchgeführt. Hier zeigten sich keine nennenswerten Auffälligkeiten, sodass auch hier Rasch-Homogenität angenommen werden kann. Somit kann bestätigt werden, dass der WMT-2 auch in einer zu Österreich sehr ähnlichen Kultur Rasch-homogen misst.

Weiters wurden insgesamt fünf Modelltests zum Vergleich Österreich – Berlin durchgeführt. Hier erwiesen sich die Modelltests Alter 17-18 Jahre und Schultyp Gesamtschule als signifikant. Die Grafische Modellkontrolle zeigte jedoch auch hier tolerierbare Abweichungen in allen fünf Modelltests.

Um die Homogenität von potentiell dimensionsfremden Itemteilmengen zu überprüfen, wurden Martin – Löff Tests durchgeführt. Dabei wurde die Stichprobe anhand der unterschiedlichen Konstruktionskriterien des WMT in jeweils zwei Teilmengen getrennt. Untersucht wurde, ob bei der Richtung der Regelgeltung, bei der Aufbauregel Überlagern, bei der Aufbauregel Variieren und bei leichteren vs. schwereren Items dieselbe latente Dimension erfasst wird. Auch hier wurde eine Alpha-Adjustierung nach Bonferroni durchgeführt. Dabei erwiesen sich alle vier Modelltests als nicht signifikant. Es kann somit angenommen werden, dass in allen Itemteilmengen dieselbe Dimension der Intelligenz erfasst wird. Die Eindimensionalitätshypothese konnte somit angenommen werden.

Ebenso sollten in meiner Arbeit eventuelle Unterschiede in den Testleistungen der beiden Stichproben Österreich und Berlin untersucht werden. Die Ergebnisse der Varianzanalyse zeigen, dass Berliner 17-18 jährige SchülerInnen und Berliner GesamtschülerInnen signifikant schlechtere Testleistungen als österreichische 17-18 jährige SchülerInnen bzw. österreichische AHS / BHS SchülerInnen haben. Hierbei muss nochmals erwähnt werden, dass es mir leider nicht möglich war an Berliner Gymnasien zu testen, sodass der Vergleich der Berliner Gesamtschulen mit einer österreichischen AHS / BHS im Hinblick auf die Unterschiede in den Schulsystemen nur eingeschränkt interpretiert werden kann.

Ein weiteres Ziel dieser Arbeit war es, die Testgütekriterien, insbesondere die Reliabilität und Validität zu überprüfen. Es zeigten sich, zufriedenstellende Werte bei Reliabilität und Validität. Die Validität wurde dabei mittels Korrelationsberechnung mit dem ZVT ermittelt.

Im letzten Punkt dieser Arbeit wurden für die Berliner Stichprobe Normen erstellt, da sich bei den Mittelwertsvergleichen signifikante Unterschiede zur österreichischen Stichprobe zeigten und um bei Vorgabe des WMT-2 in Deutschland Vergleichswerte zu haben.

Wie bereits bisherige Untersuchungen und auch diese Arbeit zeigen, kann zusammenfassend gesagt werden, dass der WMT-2 als ein rasch-homogener Intelligenztest anzusehen ist, der auch in anderen Kulturen eingesetzt werden kann. Im Hinblick auf weitere Untersuchungen zum WMT-2 wäre es interessant, die Vergleichsstichproben zu erweitern, also Testungen auch in anderen Ländern durchzuführen. Weiters wäre es sinnvoll, die Normen, die in dieser Arbeit erstellt wurden, durch Vergleichsstichproben aus anderen Teilen Deutschlands zu ergänzen.

14. Literaturverzeichnis

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

Bortz, J. & Döring, N. (2002): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 3. überarbeitete Auflage. Heidelberg: Springer.

Broer, M. (1996): *Rasch-homogene Leistungstests im Kulturvergleich: Chile – Österreich*. Dissertation. Wien: Universität Wien.

Cattell, R.B.: *Die empirische Erforschung der Persönlichkeit*. Beltz. Weinheim-Basel 1973.

Edjeh, D. (1980): *Transkultureller Vergleich des WMT von Formann und Raschskalierung des CFT-3 und WMT für den Iran*. Dissertation. Wien: Universität Wien.

Fischer, G.H. & Ponocny-Seliger, E. (1998): *Structural Rasch Modeling: Handbook of the Useage of LPCM-Win 1.0*. Groningen, NL: Programma.

Formann, A.K. (1973): *Die Konstruktion eines neuen Matrizentests und die Untersuchung des Lösungsverhaltens mit Hilfe des linearen logistischen Testmodells*. Dissertation. Wien: Universität Wien.

Formann, A.K. & Piswanger, K. (1979): *WMT. Wiener Matrizen-Test. Ein Rasch-skaliertes sprachfreier Intelligenztest*. Weinheim: Beltz-Test.

Formann, A.K., Waldherr, K. & Piswanger, K. (2008): *WMT-2. Wiener Matrizen-Test – 2. Ein Rasch-skaliertes sprechfreier Intelligenztest*. Unveröffentlichte Ausgabe.

Hahne, J. (1999): *Lerneffekte innerhalb von Leistungstests*. Diplomarbeit. Wien: Universität Wien.

Helbig, G. & Grieser Ch. (2005): *Schulen in Berlin. Ein Wegweiser durch die Berliner Schullandschaft*. Online unter URL:

<http://www.kulturbuch-verlag.de/Shop/Leseproben/453010.pdf>

Lienert, G.A. (1994): *Testaufbau und Testanalyse*. 6. Auflage. Weinheim: Beltz.

Martin-Löf, P. (1973). *Statistiska Modeller. Anteckningar från seminarier lasaret 1969-1970, utarbetade av Rolf Sundberg*. Obetydligt ändrat nytryck, Oktober 1973. Stockholm: Institutet för Försäkringsmatematik och Matematisk Statistisk vid Stockholms Universitet.

Mohammadzadeh Koucheri, F. (1993): *Interkultureller Vergleich mit einer variierten Form des Matrizentests von Formann*. Diplomarbeit. Wien: Universität Wien.

Müller-Phillipp, S. & Tarnai, Ch. (1989): *Signifikanz und Relevanz von Modellabweichungen beim Rasch-Modell*. In Kubinger, K. (Hrsg.), *Moderne Testtheorie: ein Abriss samt neuesten Beiträgen*. Weinheim: Beltz.

Nepita, B. (1991): *Dimensionalitätsanalyse aller Rasch-homogenen Intelligenztests. Zur Frage der Dimensionalität der Intelligenz*. Diplomarbeit. Wien: Universität Wien.

Piswanger, K. (1975): *Interkulturelle Vergleiche mit dem Matrizentest von Formann*. Dissertation. Wien: Universität Wien.

Rost, J. (1996): *Testtheorie, Testkonstruktion*. Göttingen, Toronto, Seattle: Huber.

Senatsverwaltung für Bildung, Wissenschaft und Forschung (2006): *Aufbau des Berliner Schulsystems*. Online unter URL:

http://www.berlin.de/imperia/md/content/sen-bildung/bildungswege/aufbau_berliner_schulsystem.pdf

Senatsverwaltung für Bildung, Wissenschaft und Forschung (2004): *Schulgesetz für das Land Berlin (Schulgesetz - SchulG)*. Online unter URL:

<http://www.berlin.de/imperia/md/content/sen-bildung/rechtsvorschriften/schulgesetz.pdf>

Senatsverwaltung für Gesundheit, Soziales und Verbraucherschutz (2004): *Sozialstrukturatlas Berlin 2003*. Online unter URL:

http://www.efb-berlin.de/download/berliner_sozialatlas_2003_kurz.pdf

Spearman, Ch. (1927): *The abilities of man*. London: Macmillan.

Stephenson, T. (1989): *Raumvorstellung und Berufsausbildung*. Dissertation. Wien: Universität Wien.

Vogt, B. (2007): *Reanalyse und Neunormierung des Wiener Matrizen-tests*. Diplomarbeit. Wien: Universität Wien.

15. Anhang

15.1 Schulenverzeichnis

15.2 Antwortblatt

15.3 Normtabellen

15.4 Ergebnisse der Rasch-Modelltests

15.4.1 Ergebnisse des Rasch-Modelltests Geschlecht

15.4.2 Ergebnisse des Rasch-Modelltests Alter

15.4.3 Ergebnisse des Rasch-Modelltests Muttersprache

15.4.4 Ergebnisse des Rasch-Modelltests Schultyp

15.4.5 Ergebnisse des Rasch-Modelltests Mathematiknote

15.4.6 Ergebnisse des Rasch-Modelltests Darstellende Geometrie

15.4.7 Ergebnisse des Rasch-Modelltests Geometrisch Zeichnen

15.4.8 Ergebnisse des Rasch-Modelltests Rohscore

15.4.9 Ergebnisse des Rasch-Modelltests Item 1

15.4.10 Ergebnisse des Rasch-Modelltests Österreich / Berlin Alter 14-16 Jahre

15.4.11 Ergebnisse des Rasch-Modelltests Österreich / Berlin Alter 17-18 Jahre

15.4.12 Ergebnisse des Rasch-Modelltests Österreich / Berlin Alter 19-25 Jahre

15.4.13 Ergebnisse des Rasch-Modelltests Österreich AHS, BHS / Berlin Gesamtschule

15.4.14 Ergebnisse des Rasch-Modelltests Österreich / Berlin Berufsschule

15.5 Ergebnisse der Modelltests nach Martin-Löf

15.6 Ergebnisse der Varianzanalyse

Lebenslauf

15.1 Schulenverzeichnis

Berufsschulen

- Brillat Savarin Berufsschule
- Hans Böckler Berufsschule

Gesamtschulen

- Friedensburg Oberschule
- Kurt Schwitters Gesamtschule
- Thüringen Oberschule
- Wilma Rudolph Oberschule

15.2 Antwortblatt

Antwortblatt

Wiener Matrizen-Test WMT

Schultyp:
Schulstufe:
Codename:
Letzte Mathematiknote:

Alter in Jahren:
Geschlecht:
Muttersprache:

Hatten Sie eine Ausbildung in Geometrie?
oder in Geometrisches Zeichnen?

Ja ☐ Nein ☐
Ja ☐ Nein ☐

Testaufgaben

Antworten

1		a	b	c	d	e	f	g	h
2		a	b	c	d	e	f	g	h
3		a	b	c	d	e	f	g	h
4		a	b	c	d	e	f	g	h
5		a	b	c	d	e	f	g	h
6		a	b	c	d	e	f	g	h
7		a	b	c	d	e	f	g	h
8		a	b	c	d	e	f	g	h
9		a	b	c	d	e	f	g	h
10		a	b	c	d	e	f	g	h
11		a	b	c	d	e	f	g	h
12		a	b	c	d	e	f	g	h
13		a	b	c	d	e	f	g	h
14		a	b	c	d	e	f	g	h
15		a	b	c	d	e	f	g	h
16		a	b	c	d	e	f	g	h
17		a	b	c	d	e	f	g	h
18		a	b	c	d	e	f	g	h

15.3 Normtabellen

Normtabelle für 14-16 jährige GesamtschülerInnen, N = 302

Roh-score	Frequency	Prozent %	Kumulierte Prozente Kum. %	Prozentrang PR	Stanine-Wert
0	0	0.0	0.0	0	1
1	2	0.7	0.7	1	1
2	3	1.0	1.7	2	1
3	6	2.0	3.7	4	1
4	10	3.3	7.0	7	2
5	13	4.3	11.3	11	2
6	22	7.3	18.6	19	3
7	28	9.3	27.9	28	4
8	37	12.3	40.2	40	4
9	33	10.9	51.1	51	5
10	32	10.6	61.7	62	6
11	25	8.3	70.0	70	6
12	27	8.9	78.9	79	7
13	26	8.6	87.5	88	7
14	22	7.3	94.8	95	8
15	8	2.6	97.4	97	9
16	1	0.3	97.7	98	9
17	6	2.0	99.7	100	9
18	1	0.3	100.0	100	9

Mittelwert = 9.51, Median = 9, Streuung = 3.34

Normtabelle für 17-18 jährige GesamtschülerInnen, N = 158

Roh-score	Frequency	Prozent %	Kumulierte Prozente Kum. %	Prozentrang PR	Stanine-Wert
0	0	0.0	0.0	0	1
1	1	0.6	0.6	1	1
2	1	0.6	1.2	1	1
3	3	1.9	3.1	3	1
4	6	3.8	6.9	7	2
5	4	2.5	9.4	9	2
6	9	5.7	15.1	15	3
7	14	8.9	24.0	24	4
8	17	10.8	34.8	35	4
9	12	7.6	42.4	42	5
10	19	12.0	54.4	54	5
11	13	8.2	62.6	63	6
12	15	9.5	72.1	72	6
13	15	9.5	81.6	82	7
14	17	10.8	92.4	92	8
15	7	4.4	96.8	97	9
16	3	1.9	98.7	99	9
17	2	1.3	100.0	100	9
18	0	0.0	100.0	100	9

Mittelwert = 10.03, Median = 10, Streuung = 3.40

Normtabelle für 19-25 jährige BerufsschülerInnen, N = 129

Roh-score	Frequency	Prozent %	Kumulierte Prozente Kum. %	Prozentrang PR	Stanine-Wert
0	0	0.0	0.0	0	1
1	1	0.8	0.8	1	1
2	4	3.1	3.9	4	1
3	2	1.6	5.5	6	2
4	3	2.3	7.8	8	2
5	5	3.9	11.7	12	3
6	7	5.4	17.1	17	3
7	4	3.1	20.2	20	3
8	12	9.3	29.5	30	4
9	9	7.0	36.5	37	4
10	11	8.5	45.0	45	5
11	11	8.5	53.5	54	5
12	13	10.1	63.6	64	6
13	11	8.5	72.1	72	6
14	8	6.2	78.3	78	7
15	10	7.8	86.1	86	7
16	11	8.5	94.6	95	8
17	5	3.9	98.5	99	9
18	2	1.6	100.1	100	9

Mittelwert = 10.76, Median = 11, Streuung = 4.058

15.4 Ergebnisse der Rasch-Modelltests

15.4.1 Ergebnisse des Rasch-Modelltests Geschlecht

Conditional Likelihood-Ratio-Test
(External Criterion = Geschlecht)

Andersen chi_2 = 22.0089
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-1.1375	n.s.
2	-2.1041	sign.(5%)
3	-0.6092	n.s.
4	1.8506	n.s.
5	-1.4049	n.s.
6	0.3035	n.s.
7	-1.4773	n.s.
8	1.4335	n.s.
9	-0.9165	n.s.
10	0.2610	n.s.
11	2.2238	sign.(5%)
12	-1.1399	n.s.
13	0.2711	n.s.
14	1.4681	n.s.
15	3.3004	sign.(5% and 1%)
16	3.8440	sign.(5% and 1%)
17	0.6828	n.s.
18	-0.3762	n.s.

Pearson Correlation between parameters = 0.9873

15.4.2 Ergebnisse des Rasch-Modelltests Alter

Conditional Likelihood-Ratio-Test
(External Criterion = Alter)

Andersen chi_2 = 22.7097
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	0.2088	n.s.
2	1.2073	n.s.
3	-0.4514	n.s.
4	-0.7880	n.s.
5	0.3670	n.s.
6	1.7198	n.s.
7	-2.6997	sign.(5% and 1%)
8	-2.8487	sign.(5% and 1%)
9	-2.4259	sign.(5%)
10	1.5078	n.s.
11	-1.0636	n.s.
12	1.1017	n.s.
13	-2.7308	sign.(5% and 1%)
14	0.6584	n.s.
15	2.7042	sign.(5% and 1%)
16	-2.5482	sign.(5%)
17	1.9942	sign.(5%)
18	0.7625	n.s.

Pearson Correlation between parameters = 0.9834

15.4.3 Ergebnisse des Rasch-Modelltests Muttersprache

Conditional Likelihood-Ratio-Test
(External Criterion = Muttersprache)

Andersen chi_2 = 30.1547
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	0.9454	n.s.
2	2.0185	sign.(5%)
3	0.7308	n.s.
4	3.0941	sign.(5% and 1%)
5	-1.4167	n.s.
6	1.6926	n.s.
7	3.3387	sign.(5% and 1%)
8	-0.2362	n.s.
9	-1.9864	sign.(5%)
10	-1.3536	n.s.
11	-2.2784	sign.(5%)
12	-3.2350	sign.(5% and 1%)
13	1.3546	n.s.
14	0.6387	n.s.
15	-2.1901	sign.(5%)
16	3.8316	sign.(5% and 1%)
17	-2.8199	sign.(5% and 1%)
18	-0.9232	n.s.

Pearson Correlation between parameters = 0.9758

15.4.4 Ergebnisse des Rasch-Modelltests Schultyp

Conditional Likelihood-Ratio-Test
(External Criterion = Schultyp)

Andersen chi_2 = 17.9438
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-0.9680	n.s.
2	-1.1145	n.s.
3	-0.5202	n.s.
4	2.3415	sign.(5%)
5	1.2109	n.s.
6	-0.6770	n.s.
7	2.1549	sign.(5%)
8	4.1808	sign.(5% and 1%)
9	0.3884	n.s.
10	-0.9212	n.s.
11	-1.0023	n.s.
12	-0.8771	n.s.
13	2.1730	sign.(5%)
14	0.3906	n.s.
15	-2.3283	sign.(5%)
16	0.3363	n.s.
17	-0.5289	n.s.
18	-1.3047	n.s.

Pearson Correlation between parameters = 0.9821

15.4.5 Ergebnisse des Rasch-Modelltests Mathematiknote

Conditional Likelihood-Ratio-Test
(External Criterion = Mathematiknote)

Andersen chi_2 = 29.4404
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-1.9512	n.s.
2	1.2431	n.s.
3	-0.8287	n.s.
4	2.7998	sign.(5% and 1%)
5	-0.4159	n.s.
6	2.6747	sign.(5% and 1%)
7	2.8249	sign.(5% and 1%)
8	-0.4375	n.s.
9	-3.5282	sign.(5% and 1%)
10	3.3154	sign.(5% and 1%)
11	0.2312	n.s.
12	-1.6232	n.s.
13	-1.7331	n.s.
14	2.8381	sign.(5% and 1%)
15	-2.3862	sign.(5%)
16	0.7124	n.s.
17	-1.3874	n.s.
18	0.4610	n.s.

Pearson Correlation between parameters = 0.9778

15.4.6 Ergebnisse des Rasch-Modelltests Darstellende Geometrie

Conditional Likelihood-Ratio-Test
(External Criterion = DG)

Andersen chi_2 = 25.2742
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-2.0639	sign.(5%)
2	-0.9362	n.s.
3	2.0018	sign.(5%)
4	0.7966	n.s.
5	1.1848	n.s.
6	-1.6450	n.s.
7	1.4860	n.s.
8	1.9317	n.s.
9	-1.8872	n.s.
10	-0.6080	n.s.
11	1.7858	n.s.
12	1.0285	n.s.
13	-0.6715	n.s.
14	1.8677	n.s.
15	-1.8494	n.s.
16	-0.7417	n.s.
17	-2.6302	sign.(5% and 1%)
18	0.7646	n.s.

Pearson Correlation between parameters = 0.9577

15.4.7 Ergebnisse des Rasch-Modelltests Geometrisch Zeichnen

Conditional Likelihood-Ratio-Test
(External Criterion = GZ)

Andersen chi_2 = 20.8536
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-0.5005	n.s.
2	0.6143	n.s.
3	0.0663	n.s.
4	0.7647	n.s.
5	1.3299	n.s.
6	0.9703	n.s.
7	1.7645	n.s.
8	-0.3865	n.s.
9	-1.8134	n.s.
10	1.3134	n.s.
11	-2.2774	sign.(5%)
12	-0.9044	n.s.
13	2.4043	sign.(5%)
14	3.2445	sign.(5% and 1%)
15	1.1416	n.s.
16	-3.2253	sign.(5% and 1%)
17	-2.9434	sign.(5% and 1%)
18	-0.0264	n.s.

Pearson Correlation between parameters = 0.9855

15.4.8 Ergebnisse des Rasch-Modelltests Rohscore

Conditional Likelihood-Ratio-Test
(External Criterion=Rohscore)

Andersen chi_2 = 91.6228
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-0.3703	n.s.
2	-0.6687	n.s.
3	-0.5718	n.s.
4	-1.2368	n.s.
5	2.7180	sign.(5% and 1%)
6	-3.6241	sign.(5% and 1%)
7	-2.5295	sign.(5%)
8	-0.8520	n.s.
9	-0.9327	n.s.
10	-3.2106	sign.(5% and 1%)
11	5.6010	sign.(5% and 1%)
12	2.8541	sign.(5% and 1%)
13	-1.0143	n.s.
14	-2.6843	sign.(5% and 1%)
15	3.6321	sign.(5% and 1%)
16	-5.0318	sign.(5% and 1%)
17	6.0771	sign.(5% and 1%)
18	2.8795	sign.(5% and 1%)

Pearson Correlation between parameters = 0.9481

15.4.9 Ergebnisse des Rasch-Modelltests Item 1

Conditional Likelihood-Ratio-Test
(External Criterion = Item 1)

Andersen chi_2 = 21.3329
Degrees of Freedom = 16
chi_2 at alpha(5 percent) = 26.2861 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 32.0265 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-2.4837	sign.(5%)
2	-2.2293	sign.(5%)
3	-1.1727	n.s.
4	0.9219	n.s.
5	-0.5494	n.s.
6	-0.7132	n.s.
7	-1.8665	n.s.
8	-0.4797	n.s.
9	-0.5354	n.s.
10	2.1860	sign.(5%)
11	2.4700	sign.(5%)
12	0.0159	n.s.
13	-1.1013	n.s.
14	2.4642	sign.(5%)
15	0.9861	n.s.
16	0.5413	n.s.
17	1.5074	n.s.

Pearson Correlation between parameters = 0.9824

15.4.10 Ergebnisse des Rasch-Modelltests Österreich / Berlin Alter 14-16 Jahre

Conditional Likelihood-Ratio-Test
(External Criterion = Ö / B (14-16 Jahre))

Andersen chi_2 = 40.4492
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	1.7571	n.s.
2	0.8854	n.s.
3	4.0488	sign.(5% and 1%)
4	-0.8006	n.s.
5	-1.6893	n.s.
6	1.2899	n.s.
7	0.8580	n.s.
8	1.0504	n.s.
9	-3.4461	sign.(5% and 1%)
10	-1.4334	n.s.
11	0.1796	n.s.
12	-2.3438	sign.(5%)
13	-2.8582	sign.(5% and 1%)
14	-3.4143	sign.(5% and 1%)
15	1.3486	n.s.
16	4.6448	sign.(5% and 1%)
17	-0.1891	n.s.
18	-1.0090	n.s.

Pearson Correlation between parameters = 0.9783

15.4.11 Ergebnisse des Rasch-Modelltests Österreich / Berlin Alter 17-18 Jahre

Conditional Likelihood-Ratio-Test
(External Criterion = Ö / B (17-18 Jahre))

Andersen chi_2 = 49.7401
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	0.7576	n.s.
2	1.2793	n.s.
3	1.0398	n.s.
4	0.2240	n.s.
5	-2.9407	sign.(5% and 1%)
6	-3.7885	sign.(5% and 1%)
7	4.1830	sign.(5% and 1%)
8	4.0014	sign.(5% and 1%)
9	1.9457	n.s.
10	-2.7180	sign.(5% and 1%)
11	1.7341	n.s.
12	-2.9037	sign.(5% and 1%)
13	0.5595	n.s.
14	-3.4692	sign.(5% and 1%)
15	-0.9435	n.s.
16	4.3395	sign.(5% and 1%)
17	-0.9496	n.s.
18	-1.4221	n.s.

Pearson Correlation between parameters = 0.9570

15.4.12 Ergebnisse des Rasch-Modelltests Österreich / Berlin Alter 19-25 Jahre

Conditional Likelihood-Ratio-Test
(External Criterion = Ö / B (19-25 Jahre))

Andersen chi_2 = 26.0327
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-0.6434	n.s.
2	-1.5719	n.s.
3	0.4625	n.s.
4	0.4996	n.s.
5	2.3815	sign.(5%)
6	-0.2261	n.s.
7	3.9971	sign.(5% and 1%)
8	1.6757	n.s.
9	2.2827	sign.(5%)
10	1.8397	n.s.
11	2.2422	sign.(5%)
12	-1.9515	n.s.
13	0.1606	n.s.
14	-2.4392	sign.(5%)
15	-1.3783	n.s.
16	-0.2570	n.s.
17	-2.5160	sign.(5%)
18	-1.5784	n.s.

Pearson Correlation between parameters = 0.9271

15.4.13 Ergebnisse des Rasch-Modelltests Österreich / Berlin Berufsschule

Conditional Likelihood-Ratio-Test
(External Criterion = Ö / B (Berufsschule))

Andersen chi_2 = 20.1115
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	-0.1160	n.s.
2	0.2786	n.s.
3	0.7078	n.s.
4	0.5158	n.s.
5	-0.0922	n.s.
6	-4.1466	sign.(5% and 1%)
7	2.3300	sign.(5%)
8	2.6302	sign.(5% and 1%)
9	-0.3795	n.s.
10	-2.7792	sign.(5% and 1%)
11	-1.6034	n.s.
12	-0.6714	n.s.
13	3.0919	sign.(5% and 1%)
14	-1.5381	n.s.
15	-0.9644	n.s.
16	2.0659	sign.(5%)
17	-0.2931	n.s.
18	0.3372	n.s.

Pearson Correlation between parameters = 0.9484

15.4.14 Ergebnisse des Rasch-Modelltests Österreich AHS, BHS / Berlin Gesamtschule

Conditional Likelihood-Ratio-Test
(External Criterion = Ö AHS,BHS / B Gesamtschule)

Andersen chi_2 = 77.6382
Degrees of Freedom = 17
chi_2 at alpha(5 percent) = 27.5775 (Wilson-Hilferty approximation)
chi_2 at alpha(1 percent) = 33.4347 (Wilson-Hilferty approximation)

Wald z-Values
(External Criterion)

Item No.	z-Value	Significance
1	2.4530	sign.(5%)
2	2.0740	sign.(5%)
3	3.8459	sign.(5% and 1%)
4	0.8936	n.s.
5	-3.3253	sign.(5% and 1%)
6	-0.9723	n.s.
7	2.6757	sign.(5% and 1%)
8	2.9589	sign.(5% and 1%)
9	-1.2937	n.s.
10	-1.2401	n.s.
11	2.5166	sign.(5%)
12	-3.7026	sign.(5% and 1%)
13	-4.1405	sign.(5% and 1%)
14	-4.6555	sign.(5% and 1%)
15	0.6125	n.s.
16	4.1223	sign.(5% and 1%)
17	-1.0911	n.s.
18	-3.6320	sign.(5% and 1%)

Pearson Correlation between parameters = 0.9603

15.5 Ergebnisse der Modelltests nach Martin – Löf

Richtung der Regelgeltung

Martin-Löf Test

Group 1: 2, 3, 6, 7, 8, 9, 12, 14, 15, 18,
Group 2: 1, 4, 5, 10, 11, 13, 16, 17,
LogLikelihood (Group 1): -2213.627855
LogLikelihood (Group 2): -1803.354100
Martin-Löf-Statistic: 75.3000 n.s.(5% und 1%)
(chi_2(5%)= 100.7463, chi_2(1%)= 111.1581, and df=79)

Aufbauregel Überlagern

Martin-Löf Test

Group 1: 1, 2, 3, 4, 5, 7, 11, 12, 17,
Group 2: 6, 8, 9, 10, 13, 14, 15, 16, 18,
LogLikelihood (Group 1): -1790.625456
LogLikelihood (Group 2): -2105.594114
Martin-Löf-Statistic: 119.8110 sign.
(chi_2(5%)= 101.8772, chi_2(1%)= 112.3428, and df=80)

Aufbauregel Variieren

Martin-Löf Test

Group 1: 2, 6, 9, 10, 15, 16, 17,
Group 2: 1, 3, 4, 5, 7, 8, 11, 12, 13, 14, 18,
LogLikelihood (Group 1): -1370.915581
LogLikelihood (Group 2): -2647.999617
Martin-Löf-Statistic: 60.7381 n.s.(5% und 1%)
(chi_2(5%)= 97.3486, chi_2(1%)= 107.5969, and df=76)

Leichtere vs. schwerere Items

Martin-Löf Test

Group 1: 10, 11, 12, 13, 14, 15, 16, 17, 18,
Group 2: 1, 2, 3, 4, 5, 6, 7, 8, 9,
LogLikelihood (Group 1): -2186.084514
LogLikelihood (Group 2): -1829.019924
Martin-Löf-Statistic: 74.7168 n.s.(5% und 1%)
(chi_2(5%)= 101.8772, chi_2(1%)= 112.3428, and df=80)

15.6 Ergebnisse der Varianzanalyse

Mittelwertsvergleiche anhand der unabhängigen Variable Schultyp

(I) Alter_neu	(J) Alter_neu	Mittlere Differenz (I-J)	Standardf ehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
B<=16	B17-18	-,428	,324	,884	-1,51	,65
	B>18	-1,316*	,367	,025	-2,54	-,09
	Ö<=16	-,686	,238	,142	-1,48	,11
	Ö17-18	-2,113*	,254	,000	-2,96	-1,27
	Ö19-25	-2,772*	,427	,000	-4,19	-1,35
B17-18	B<=16	,428	,324	,884	-,65	1,51
	B>18	-,889	,401	,426	-2,22	,45
	Ö<=16	-,258	,287	,976	-1,21	,70
	Ö17-18	-1,686*	,300	,000	-2,68	-,69
	Ö19-25	-2,344*	,456	,000	-3,86	-,83
B>18	B<=16	1,316*	,367	,025	,09	2,54
	B17-18	,889	,401	,426	-,45	2,22
	Ö<=16	,630	,335	,617	-,48	1,75
	Ö17-18	-,797	,346	,380	-1,95	,35
	Ö19-25	-1,455	,487	,113	-3,08	,17
Ö<=16	B<=16	,686	,238	,142	-,11	1,48
	B17-18	,258	,287	,976	-,70	1,21
	B>18	-,630	,335	,617	-1,75	,48
	Ö17-18	-1,427*	,204	,000	-2,11	-,75
	Ö19-25	-2,086*	,399	,000	-3,41	-,76
Ö17-18	B<=16	2,113*	,254	,000	1,27	2,96
	B17-18	1,686*	,300	,000	,69	2,68
	B>18	,797	,346	,380	-,35	1,95
	Ö<=16	1,427*	,204	,000	,75	2,11
	Ö19-25	-,658	,408	,761	-2,02	,70
Ö19-25	B<=16	2,772*	,427	,000	1,35	4,19
	B17-18	2,344*	,456	,000	,83	3,86
	B>18	1,455	,487	,113	-,17	3,08
	Ö<=16	2,086*	,399	,000	,76	3,41
	Ö17-18	,658	,408	,761	-,70	2,02

*. Die Differenz der Mittelwerte ist auf dem Niveau .05 signifikant.

Mittelwertsvergleiche anhand der unabhängigen Variable Schultyp

(I) Besucher Schultyp der Testperson	(J) Besucher Schultyp der Testperson	Mittlere Differenz (I-J)	Standardf ehler	Signifikanz
B_Berufsschule	B_Gesamtschule	-,593	,558	,770
	Ö_Berufsschule	-,570	,552	,785
	Ö_AHS/BHS	-2,796*	,553	,000
B_Gesamtschule	B_Berufsschule	,593	,558	,770
	Ö_Berufsschule	,024	,208	1,000
	Ö_AHS/BHS	-2,203*	,210	,000
Ö_Berufsschule	B_Berufsschule	,570	,552	,785
	B_Gesamtschule	-,024	,208	1,000
	Ö_AHS/BHS	-2,226*	,193	,000
Ö_AHS/BHS	B_Berufsschule	2,796*	,553	,000
	B_Gesamtschule	2,203*	,210	,000
	Ö_Berufsschule	2,226*	,193	,000

Lebenslauf

Persönliche Daten:

Name: Natascha Marakovits
Anschrift: Gumpendorferstraße 83-85/1/3
1060 Wien
E-Mail: natascha_ma@hotmail.com
Telefon: 0664/92 33 275
Geburtsdatum: 28.09.1981
Geburtsort: Güssing
Familienstand: ledig

Ausbildung:

1988-1999: Volksschule, Güssing
1992-1996: Hauptschule, Güssing
1996-2000: Bundesoberstufenrealgymnasium, Güssing
Seit SS 2001: Psychologiestudium, Universität Wien

Berufliche Tätigkeiten:

Sommer 2001: Verkaufstätigkeit, Schlecker
Sommer 2002,03,04,05: Feriertätigkeit, Post AG
Sept. 2005 – Dez. 2005: geringfügige Beschäftigung, ENWC Rechtsanwälte GmbH
Juli 2006: Firmendatenrecherche, Compass Verlag
Neben dem Studium: Mitarbeit in Marktforschungsinstituten
Seit Februar 2007: Honorartrainerin im BBRZ (Berufliches Bildungs- und Rehabilitationszentrum)

Zusätzliche Qualifikationen:

31.08.2005 – 23.11.2005: Praktikum bei HILL INTERNATIONAL
Okt. 2007 – Nov. 2007: Praktikum in der „Praxis für Neurorehabilitation“
Febr. 2008 – April 2008: Forschungsaufenthalt in Berlin